The Measurement of Environmental and Resource Values

THEORY AND METHODS

THIRD EDITION

A. Myrick Freeman III,
Joseph A. Herriges,
and Catherine L. Kling

RFF PRESS
RESOURCES FOR THE FUTURE

ROUTLEDGE

The Measurement of Environmental and Resource Values

The first edition of this important work was the winner of the 2002 Publication of Enduring Quality award by the Association of Environmental and Resource Economists. The continuing premise for the book is that estimates of the economic values of environmental and natural resource services are essential for effective policy-making. Like previous editions, the third edition, which includes two additional co-authors, presents a comprehensive treatment of the theory and methods involved in estimating environmental benefits.

Researchers, policy-makers, and practitioners will welcome the work as an up-to-date reference on recent developments. Students will gain a better understanding of the contribution that economics as a discipline can make to decisions concerning pollution control and human health, recreation, environmental amenities, and other critical issues concerning the way we use and interact with environmental and natural resource systems. To reflect recent progress in both the theory and practice of non-market valuation, this third edition includes more details on empirical approaches to measurement, expanded discussion of the reasons for divergence between “willingness to pay” and “willingness to accept compensation,” and increased coverage of econometric issues encountered in estimation. In keeping with its cutting-edge orientation, it also includes more discussion of survey design, equilibrium sorting models, and the implications of behavioral economics for welfare measurements and benefit cost analysis.

A. Myrick Freeman III is William D. Shipman Professor of Economics Emeritus at Bowdoin College in Maine, USA, and a former Senior Fellow at Resources for the Future (RFF), Washington DC, USA.

Joseph A. Herriges is Professor of Economics at Iowa State University, USA. He was Managing Editor of the *Journal of Environmental Economics and Management* from 2001 to 2006.

Catherine L. Kling is Charles F. Curtiss Distinguished Professor at Iowa State University, USA, and the Director of the Center for Agricultural and Rural Development. She is a former President of the Association of Environmental and Resource Economists.

About Resources for the Future *and* RFF Press

Resources for the Future (RFF) improves environmental and natural resource policymaking worldwide through independent social science research of the highest caliber. Founded in 1952, RFF pioneered the application of economics as a tool for developing more effective policy about the use and conservation of natural resources. Its scholars continue to employ social science methods to analyze critical issues concerning pollution control, energy policy, land and water use, hazardous waste, climate change, biodiversity, and the environmental challenges of developing countries.

RFF Press supports the mission of RFF by publishing book-length works that present a broad range of approaches to the study of natural resources and the environment. Its authors and editors include RFF staff, researchers from the larger academic and policy communities, and journalists. Audiences for publications by RFF Press include all of the participants in the policymaking process—scholars, the media, advocacy groups, NGOs, professionals in business and government, and the public.

"This new edition provides the complete package in environmental valuation – adding more detail on empirical methods, surveys, behavioral economics, and frontier issues to the previous classic edition. A particularly appealing aspect of the book is the rigorous linkage between economic theory and the practice of valuation. It is truly an outstanding contribution to the field."

– W.L. (Vic) Adamowicz, Professor, Department of Resource Economics and Environmental Sociology, University of Alberta, Canada

"This superb book sets a new standard for guidelines regarding the valuation of environmental resources. Embracing the full range of available methods, it provides a comprehensive and readily accessible guide to valuation exercises that will prove a key reference for both researchers and research users."

– Ian Bateman, Professor of Environmental Economics, University of East Anglia, UK

"This book, co-authored by three of the world's leading experts in the field of nonmarket valuation, should be on the shelf of anyone interested in an authoritative and up-to-date source of information about the theory and methods used in this field."

– Kathleen Segerson, Alumni Association Distinguished Professor, Department of Economics, University of Connecticut, USA

The Measurement of Environmental and Resource Values

Theory and Methods

Third Edition

A. Myrick Freeman III,
Joseph A. Herriges, and
Catherine L. Kling

RFF PRESS
RESOURCES FOR THE FUTURE

First published 2014
by RFF Press
Taylor & Francis, 2 Park Square, Milton Park, Abingdon, Oxon
OX14 4RN

And by RFF Press
Routledge, 711 Third Avenue, New York, NY 10017

RFF Press is an imprint of the Taylor & Francis Group, an informa business

First edition published 1993

Second edition published 2003

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Freeman, A. Myrick, 1936-
The measurement of environmental and resource values : theory and methods / Myrick Freeman III, Joseph A. Herriges, Catherine L. Kling. -- Third edition.
pages cm
Summary: "The first edition of this important work was the winner of the 2002 Publication of Enduring Quality award by the Association of Environmental and Resource Economists. The continuing premise for the book is that estimates of the economic values of environmental and natural resource services are essential for effective policy-making. As previous editions, the third edition, which includes two additional co-authors, presents a comprehensive treatment of the theory and methods involved in estimating environmental benefits"-- Provided by publisher.
Includes bibliographical references and index.
1. Environmental policy--Cost effectiveness. 2. Environmental policy--Evaluation. I. Herriges, Joseph A., 1956- II. Kling, Catherine L. III. Title.
HC79.E5F713 2014
333.7--dc23 2013042862

ISBN: 978-0-415-50157-6 (hbk)
ISBN: 978-0-415-50158-3 (pbk)
ISBN: 978-1-315-78091-7 (ebk)

Typeset in Baskerville
by HWA Text and Data Management, London

Dedicated to our spouses:

Claire M. Darrow
Mary J. Herriges
Terence J. Alexander

Contents

Illustrations

Figures

Tables

Foreword

I appreciate the opportunity to write the foreword for this new edition of *The Measurement of Environmental and Resource Values* because it allows me to be the first to comment on this revision to a classic. Over 35 years ago, when Freeman wrote the book that started his incredible legacy to environmental economists, the information available to economists about how people were affected by and responded to changes in environmental quality was limited. The primary data available were scarce. Surveys of outdoor recreation use (such as those conducted by the Bureau of Outdoor Recreation) and data describing average housing "prices" from the Census were the best we had. The norm for information allowing people's choices to be linked to environmental conditions was "bleak." Most data were from secondary sources. They were often aggregated in ways that compromised their ability to inform researchers about the importance of the differences in spatially delineated environmental services that people experienced as part of their everyday lives.

Today there appear to be few limits to what creative young environmental economists have been able to construct. We now have the ability to observe individuals and their households with considerable detail. Indeed, confidentiality concerns seem to be the single most important limitation to the granularity in the records this new generation has developed. We observe a host of market and non-market behaviors including how people use their time; and we can connect the environmental conditions that affect them in many diverse locations on the Earth. These records include variations between individuals at different locations at a point in time as well as over time. Often we can observe the same person (or household) responding to changes in environmental conditions over time. Economic analyses of what we can observe have also changed. Today there is greater integration of econometric methods and the economic theory describing how heterogeneous agents respond to changes in environmental services.

In the second edition of *Measurement*, Freeman noted that his book was intended to serve a complementary role to the treatment of econometric methods in other books. This new edition represents a significant departure from this strategy. In my view it is the first book in applied welfare economics to take seriously the need to integrate economics and micro-econometrics so that both contribute to informing the decisions analysts must make in using the rich detail in this new information landscape to understand people's choices.

The result of this integration is important. We have the same clear explanations and attention to the role of assumptions in understanding the economic behaviors we wish to describe as Rick offered in his earlier editions. What is new is the direct link to the ways econometric models represent them. Some examples may help to illustrate my point. The discussion of the definitions and measures for changes in individual welfare in response to an exogenous change in the circumstances governing a person's choices now blends the careful treatment of how price, quantity, and quality changes are evaluated in Marshallian and Hicksian terms with an expansion in the discussion of how these distinctions are represented within discrete choice models. This blending of theory and empirical insights can also be seen in the discussion of how theory, experimental results, and new behavioral hypotheses all contribute to interpretations of the willingness to pay / willingness to accept disparity. There are many other examples of this integration throughout the new edition. The econometric treatment of data problems so often encountered in recreation demand modeling is recognized to be equally important to the measurement of consumer surplus with the models we estimate. This discussion is now able to highlight the modeling tradeoffs—simplicity in measuring tradeoffs versus strong restrictions to the data generating process. The discussion of modeling choices for describing multiple-site models is equally nuanced. The bottom line to these examples is that in each area where theory must be adapted to meet the complexities of real choice processes, the authors bring readers to the frontiers of our understanding. This is true for travel cost recreation demand, hedonic models, including locational equilibrium approaches, and stated preference surveys.

Ten years ago when I prepared the forward for Rick's second edition, I wondered (to myself) will this be the last revision? I wished there was a way to assure the Freeman legacy of providing a platform for clear access to what has been learned about measuring the economic values for changes in environmental services could be made sustainable. Rick and Resources for the Future answered my question. They recognized that many generations of environmental economists have been using those early editions of Freeman to learn and, as a result, appreciate the importance of maintaining it. This new edition adds two of the leaders in our field, Joe Herriges and Cathy Kling, to the team so that his legacy is sustained! By recruiting the best of his early "students" to help, RFF has assured the Freeman, Herriges, and Kling edition of *Measurement* will continue to help new generations of environmental economists understand what has been accomplished and build on it. In the process, I believe the addition of Joe and Cathy will enhance Rick's legacy by encouraging greater integration of economic theory and econometric methods in the ways we evaluate strategies for measuring environmental and resource values.

V. Kerry Smith, Cave Creek, AZ

Preface

Freeman has been very gratified by the reception that the second edition of this book received. He appreciates the many requests that he undertake another revision of the book to reflect recent developments in the field of nonmarket valuation. However, he recognizes that the field has been rapidly developing since his retirement from teaching some 13 years ago. He is immensely grateful that Joseph Herriges and Catherine Kling agreed to join him as coauthors of this, the third edition. Herriges and Kling, for their part, feel privileged to have been asked. As a graduate student, Kling used the first edition of Freeman's book and both Herriges and Kling taught for many years using the second edition. To be invited by the scholar that literally "wrote the book" on measuring environmental and natural resource values to participate in a revision of that book is truly an honor. Our goal was for the third edition to read as if written by a single author; consequently, we all contributed to all chapters, working and re-working the text together. We share equal responsibility for errors and oversights.

The objectives of this edition are essentially the same as those of the first two editions. These objectives are, first, to provide an introduction and overview of the principal methods and techniques of resource valuation to professional economists and graduate students who are not directly engaged in the field and, second, to give practitioners in the field an up-to-date reference on recent developments in the theory and methods underlying the practice of resource valuation. While we have tried to be comprehensive in our coverage of topics, this book is not a "how-to" manual. That kind of book would have to deal in much more detail with a host of econometric, data, and related technical issues and several excellent volumes are available. We do hope however that this book provides the necessary background in theory, basic models, data needs, and econometric overviews so that the reader will have a strong basis for diving into the task of undertaking a nonmarket valuation study.

What Is New

As the field of applied welfare economics and nonmarket valuation has become more mature, researchers have adapted and developed increasingly sophisticated econometric tools. New theory and an increasing number of theoretical puzzles have also become apparent. A primary goal in this revision has been to update the text to reflect these advances and to freshen the examples with current empirical work on contemporary environmental issues. In addition to updates throughout the text, the chapter on stated preference methods has been completely rewritten and the recreation demand chapter significantly updated. In the chapter on property value models we have added a section on equilibrium sorting models. Also we have added discussions and references to recent work on behavioral economics and its implications for nonmarket valuation methods. Given the extraordinary creativity of environmental economists, we acknowledge that the text will quickly begin sliding out of date and we look forward to continuing to read and learn from our colleagues as this field continues to progress.

Acknowledgments

We would like to thank both Nathan Cook and Becky Olsen for their patience and expertise in editing and preparing the manuscript for publication. We also want to thank Jiaqi Ge, Mainul Hoque, Hocheol Jeon, Yongjie Ji, and DongGyu Yi for their willingness to read through and comment on early drafts of this edition. Finally, we want to thank the publishers of the *Southern Economic Journal* and the *Journal of Risk and Uncertainty* for permission to adapt materials that first appeared in their journals.

A. Myrick Freeman III, Georgetown, ME
Catherine L. Kling, Ames, IA
Joseph A. Herriges, Ames, IA

Chapter 1

Resource Evaluation and Public Policy

The premise underlying this book is that estimates of the economic values of environmental and resource services can be a valuable part of the information base supporting resource and environmental management decisions. The importance of this premise is illustrated by a number of current environmental and resource policy issues, all of which involve in one way or another questions of economic values and tradeoffs. Consider these issues:

- Achieving the air and water pollution control objectives established by Congress requires massive expenditures on the part of both the public and private sectors. Is this diversion of resources from the production of other goods and services making us better off?
- Economists since A.C. Pigou (1929) have advocated placing taxes on emissions of air and water pollutants based on the damages they cause. What tax rate should be placed on these damages? How much do these rates vary across locations and time? An important related question is, are the gains from moving to pollution taxes greater than the costs of estimating the relevant marginal damages?
- The development of new reserves of petroleum and minerals is increasingly impinging on wild and natural areas that provide other environmental and resource services. Areas that might be affected include the Arctic National Wildlife Refuge, with its fragile habitat for caribou and other species, and the outer continental shelf, where commercial and recreational fisheries may be threatened by petroleum exploration and production. The 2010 oil spill in the Gulf of Mexico, and associated pictures on the nightly news of impacted wildlife, provided a cogent reminder that these same tradeoffs can occur in areas of existing production as well. Are restrictions on development in ecologically sensitive areas worth the costs they impose on society in the form of reduced availability of, and higher prices for, energy and minerals?
- The development and management of large river systems such as the Columbia River basin involves choosing among alternative combinations of hydroelectric power, water supply, and commercial and recreational fishing. There are also proposals to remove existing dams from many rivers. Are

the ecological and recreational benefits of removing a dam greater than the costs in the form of reduced power generation and water storage? Is it worthwhile to curb water withdrawals for irrigation or reduce discharges for power production in order to protect populations of salmon and other migratory fish?

- The commercial exploitation of some natural resource systems may be proceeding at unsustainable rates. Examples include some tropical forests and many of the world's fisheries. Shifting to sustainable rates of harvest may involve substantial short-term costs in the form of forgone incomes in order to achieve long-term increases in the flows of other ecological services. Are the long-term gains from achieving sustainable rates of harvest greater or less than the short-term costs?
- The scientific consensus is that substantial reductions in the emissions of greenhouse gasses will be required to slow or reverse the warming of the global climate. What degree of emissions reduction can be justified by the benefits of slowing or preventing global warming?
- Many people are now advocating that countries expand their system of national income accounts to include measures of the values of nonmarket environmental services, and deductions for the costs of environmental degradation and resource depletion. See, for example, Nordhaus and Kokkelenberg (1999).[1] How are these values and costs to be measured?

This book is about how, by providing measures of the economic values of the services of environmental and natural resource systems, economics as a discipline can contribute to answering questions such as these. We begin by introducing the idea of the natural environment as a set of assets or a kind of natural capital (Kareiva et al. 2011; Barbier 2011).

The Assets of Nature

Natural resources, such as forests and commercially exploitable fisheries, and environmental attributes, such as air quality, are valuable assets in that they yield flows of services to people. Public policies and the actions of individuals and firms can lead to changes in these service flows, thereby creating benefits and costs. Because of externalities and the common property and public good characteristics of at least some of these services, market forces can be relied on neither to guide them to their most highly valued uses nor to reveal prices that reflect their true social values. Externalities arise when a real variable (not a price) chosen by one economic agent enters the utility or production function of other economic agents. Inefficiencies can occur when there is no requirement to, or incentive for,

1 For information and references on green accounting, see the web page of the United Nations Environmental Program Green Accounting Resource Center at: www.unep.ch/etb/areas/VRC_index.php.

the first agent to take the effect on others into account when making choices. An example is the level of emissions of smoke chosen by an electric generating plant when that smoke causes ill health to people downwind of the plant. A public good is nonexcludable and nondepletable—that is, once the good has been provided to one individual, others cannot be prevented from making use of the good, and one person's use does not diminish the use that others can make of the good. It is the externalities and public good character of many environmental services that are responsible for the failure of the market system to allocate and price resource and environmental services correctly, and that create the need for economic measures of values to guide policymaking.

Benefit-cost analysis as the basis for making decisions about water resources investments came into its own more than 50 years ago. However, since the 1950s when the techniques of conventional benefit-cost analysis were being developed and refined, there have been significant changes in the nature of the problems being dealt with and the analytical tools that have become available. V. Kerry Smith called attention to these changes in his keynote lecture at Resources for the Future's 35th anniversary celebration in 1987. He went on to say that

> This expansion of applications has far-reaching implications for the techniques used and for the treatment of measures of the benefits and costs. Consequently, it has led me to argue for the use of a broader term, *resource evaluation*, to describe more adequately the amendments and expansions to benefit-cost methods in evaluating today's environmental and natural resource issues.
>
> (Smith 1988, 2)

One of the changes noted by Smith is the expanding range of resource and environmental management problems being subjected to economic analysis. As Smith pointed out, benefit-cost analysis was first developed to assess the net economic values of public works projects, especially water resource developments, that withdrew productive factor inputs (land, labor, capital, and materials) from the economy to produce tangible outputs (for example, hydroelectric power and transportation). Many of the outputs had market counterparts, so estimation of monetary values was relatively straightforward. For example, the savings in the monetary costs of repairing flood damages was taken to be a measure of the benefits of controlling floods. In contrast, today the effects of many public actions are much more subtle and wide-ranging. This is true for both the favorable effects (benefits) and unfavorable effects (costs and damages). What were once considered unquantifiable and perhaps relatively unimportant intangibles, such as improved recreation and visual amenities, are now recognized as significant sources of value. Also, consequences that were once unrecognized (for example, small changes in the risk of cancer) or were thought to lie outside the realm of economic analysis (say, loss of biodiversity and the preservation of endangered species and unique ecological systems), are often central issues in the analysis of policy choices today.

Another change is that the distinction between natural resources and the environment that has prevailed in the economics discipline for so long is often no longer meaningful. The objects of analysis for natural resource economists have typically been such resources as the forest, the ore body, and the fish species that produced a flow of commodities to the economy such as wood, metal, and fish sticks. The environment has been viewed as the medium through which the externalities associated with air, noise, and water pollution have flowed and, sometimes, as the source of amenities. Increasingly, this distinction appears to be artificial as we recognize both the variety of service flows provided by natural resources and the importance of a variety of forms of externalities. This recognition is apparently what Smith had in mind when he suggested the need to "model both natural and environmental resources as assets" (1988, 3) that yield a variety of valuable services. Freeman, Haveman, and Kneese had earlier suggested that we "view the environment as an asset or a kind of nonreproducible capital good that produces a stream of various services for man. Services are tangible (such as flows of water or minerals), or functional (such as the removal, dispersion, storage, and degradation of wastes or residuals), or intangible (such as a scenic view)" (1973, 20). Ecologists are now also adopting this perspective as they refer to "natural capital" and the values of ecosystem services (Prugh 1999; Daily et al. 1997; Daily et al. 2000; Daily et al. 2011).

As this change in perspective is adopted, it will be necessary to take a more expansive view of natural and environmental resources as complex systems with multiple outputs and joint products. The natural resource-environmental complex can be viewed as producing five kinds of service flows to the economy. First, as in the conventional view of resource economics, the resource-environmental system serves as a source of material inputs to the economy such as fossil fuels, wood products, minerals, water, and fish. Second, some components of the resource-environmental system provide life support services for people in the form of a breathable atmosphere, clean water, and a livable climatic regime. Changes in the flows of some of these life support services can be measured in terms of changes in the health status and life expectancies of affected populations. Third, the resource-environmental system provides a wide variety of amenity services, including opportunities for recreation, wildlife observation, the pleasures of scenic views, and perhaps even services that are not related to any direct use of the environment (sometimes called nonuse or existence values). Fourth, this system disperses, transforms, and stores the residuals that are generated as by-products of economic activity. This is usually referred to as the waste receptor service of the environment (Kneese, Ayres, and d'Arge 1970; Freeman, Haveman, and Kneese 1973). Finally, the resource-environmental system serves as a repository of genetic information that helps to determine the stability and resilience of the system in the face of anthropogenic and other shocks. Many of the services provided by natural resource-environmental systems can be characterized as direct services since their benefits accrue directly to people, for example, materials flows and life support services. Other environmental services could be better described as indirect services in the sense that they support other biological and ecological production processes that yield value to people. Examples include recycling of nutrients, decomposition of organic materials, generation and renewal of soil fertility, pollination of crops and natural vegetation, and biological control of agricultural and other pests.

A forest, such as a unit in the U.S. National Forest system, is an example of a resource-environmental system that provides a wide range of services, from materials such as wood and fiber to amenities like scenic vistas, hiking, and wildlife observation, and from the regulation of stream flow and control of erosion to the absorption of atmospheric carbon dioxide. In addition, since trees are known to emit nonmethane hydrocarbons, at least in some circumstances forests may contribute to the impairment of the life support services (Chameides et al. 1988). In the list of service flows there are examples of joint products—that is, pairs of services that can be increased or decreased together. However, often an increase in the flow of one type of service must be accompanied by a decrease in the flow of some other service, all things being equal. In other words, this system is characterized by scarcity and tradeoffs and requires a multipurpose approach to its management (Bowes and Krutilla 1989).

The economic value of a resource-environmental system as an asset is the sum of the discounted present values of the flows of all of the services. Since many of these service flows are not bought or sold in markets and therefore do not have market prices, the economic value of a natural asset may be quite different from its market value. For example, an acre of wetland might trade in the market for land on the basis of its value for commercial or residential development; but this value could be quite different from the value of its services as wildlife habitat and as means of controlling floods and recharging groundwater aquifers. It is important to emphasize, particularly to noneconomists, that in such a case the true economic value of this wetland includes both the marketed value as well as the nonmarketed ecosystem services.

The benefit of any public policy that increases the flow of one type of service is the increase in the present value of that service. However, the policy may have costs in the form of decreases in the flows of other services. Similarly, what is termed as damage due to pollution, or some other human intervention, is the reduction in the value of the flow of services it causes. All of these changes in resource flows, whether benefits, costs, or damages, have their counterparts in changes in the value of the resource-environmental system as an asset. Some attention must therefore be devoted to the theories of asset pricing, and the role of time and discounting in calculating changes in environmental and resource values. These topics will be taken up in Chapter 6.

Some of the service flows of resource-environmental systems are linked directly or indirectly to markets, and hence are responsive to market forces. Many service flows however, are not properly regulated by markets because of externalities, their public goods characteristics of nonexcludability and nondepletability, and other factors. As is well known, this means that a decentralized market system is unlikely to lead to the optimal pattern of service flows. Hence, there is a potential role for public policy in the management of resource-environmental systems and a need for information on the values of the service flows.

The Economic Concept of Value

The term "value" can have several different meanings. For example, economists and ecologists often use the term in two different ways in discussions of environmental services and ecosystems. One common use of the term is to mean "that which is desirable or worthy of esteem for its own sake; thing or quality having intrinsic worth" (*Webster's New World Dictionary*). In contrast, economists use the term in a sense more akin to a different definition, "a fair or proper equivalent in money, commodities, etc." (*Webster's* again), where "equivalent in money" represents the sum of money that would have an equivalent effect on the welfare or utilities of individuals.

These two different uses of the word correspond to a distinction made by philosophers between *intrinsic* value and *instrumental* value. According to philosophers, something has *intrinsic* value "if it is valuable *in* and *for* itself—if its value is not derived from its utility, but is independent of any use or function it may have in relation to something or someone else. …an intrinsically valuable entity is said to be an 'end-in-itself,' not just a 'means' to another's ends" [emphasis in original] (Callicott 1989, 131). In contrast, something has instrumental value if it is valued as a means to some other end or purpose. In this view, the value of something lies in its contribution to some other goal (Costanza and Folke 1997, 49).

Some people have argued that nature has intrinsic value for various reasons, including its "harmony" or its natural balance. However, from the perspective of the "new ecology" which emphasizes disturbance and change in ecosystems (for example, Botkin 1990), this explanation of why nature has intrinsic value is problematic. A conservation biologist might argue that the part of nature consisting of the variety of organisms and their interactions, and especially their genetic diversity, has intrinsic value. However, this view does not endow any particular manifestation of nature with more or less intrinsic value than some alternative manifestation. Nature's value is preserved as long as diversity in the broad sense is preserved. Although the concept of intrinsic value as applied to the environment is attractive in many respects, it does not provide a basis for dealing with the kinds of environmental management questions that were identified in the first section of this chapter. In contrast, the concept of instrumental value, and in particular the economic form of instrumental value, is well suited to helping answer these questions.

In order to assess the instrumental value of nature, it is necessary to specify a goal and to identify the contributions that specific components of nature make toward the furtherance of that goal. Economics is the study of how societies organize themselves to provide for the sustenance and well-being of their members. Thus in economics, the goal is increased human well-being. The economic theory of value is based on the ability of things to satisfy human needs and wants, or to increase the well-being or utility of individuals. Sometimes this view is referred to as the preference-based account of welfare or well-being. See for example Adler and Posner (2006, 29, 33–35). Under this view of welfare, the economic value of something is a measure of its contribution to human well-being. The economic value of resource-environmental systems, then, resides in the contributions that the variety of ecosystem functions and services make to human well-being.

The economic concept of value employed here has its foundation in neoclassical welfare economics. The basic premises of welfare economics are that the purpose of economic activity is to increase the well-being of the individuals who make up the society, and that each individual is the best judge of how well off he or she is in a given situation. Each individual's welfare depends not only on that individual's consumption of private goods and of goods and services produced by the government, but also on the quantities and qualities each receives of nonmarket goods and service flows from the resource-environmental system—for example, health, visual amenities, and opportunities for outdoor recreation. It follows that the basis for deriving measures of the economic value of changes in resource-environmental systems is their effects on human welfare.

The anthropocentric focus of economic valuation does not preclude a concern for the survival and well-being of other species. Individuals can value the survival of other species not only because of the uses people make of them (for food and recreation, for example), but also because of an altruistic or ethical concern. Indeed, numerous studies strongly suggest that people do significantly value the well-being of other species and the preservation of ecosystems to their own end. The latter can be the source of existence or nonuse values, a form of economic value discussed in Chapter 4.

If society wishes to make the most (in terms of individuals' well-being) of its endowment of all resources, it should compare the values of what its members receive from any environmental change or use of a resource (that is, the benefits) with the values of what its members give up by taking resources and factor inputs from other uses (that is, the costs). A society that is concerned with the economic well-being of its citizens should make changes in environmental and resource allocations only if what is gained by the change is worth more in terms of individuals' welfare than what is given up by diverting resources and inputs from other uses.

The standard economic theory for measuring changes in individuals' well-being was developed for the purpose of interpreting changes in the prices and quantities of goods purchased in markets. This theory has been extended in the past 40 years or so to public goods and other nonmarket services such as environmental quality and health. The theory is based on the assumption that people have well-defined preferences among alternative bundles of goods, where bundles consist of various quantities of both market and nonmarket goods. The theory also assumes that people know their preferences, and that these preferences have the property of substitutability among the market and nonmarket goods making up the bundles. By substitutability, economists mean that if the quantity of one element in an individual's bundle is reduced, it is possible to increase the quantity of some other element so as to leave the individual no worse off because of the change. In other words, the increase in the quantity of the second element substitutes for the decrease in the first element. The property of substitutability is at the core of the economist's concept of value because substitutability establishes tradeoff ratios between pairs of goods that matter to people.

Given the central role of the substitutability property in the definition and measurement of economic values, it is important to consider the evidence supporting the assumption of substitutability. This assumption is the basis of most of the models of individual choice that are used to analyze and predict a wide variety of economic behavior both in and outside of markets. These models include those of consumer demand and response to changes in prices, savings, and of the supply of labor. They also include models of a variety of individuals' behaviors related to environmental and health considerations, including participation in outdoor recreation activities, choices among jobs with varying degrees of risk of fatal accident, and choices of where to live and work when houses and urban centers offer different packages of amenities and environmental pollution. The successful development and application of these models would not be possible if substitutability was not a common feature of individuals' preferences. However, some researchers have found evidence of lexicographic preferences—preferences where substitutability is not evident. For a review and discussion of some of this evidence, see works by Common, Reid, and Blamey (1997), and Veisten, Navrud, and Valen (2006).

The tradeoffs that people make as they choose less of one good and substitute more of some other good reveal something about the values people place on those goods. If the monetary value of one of the goods is known, the revealed values can also be expressed in monetary units. The money price of a market good is just a special case of a tradeoff ratio because the money given up to purchase one unit of one element of the bundle is a proxy for the quantities of one or more of the other elements in the bundle that had to be reduced in order to make the purchase. However, even when money prices are not available, the tradeoff ratios can be interpreted as expressions of economic values. In fact, there is a growing literature exploring such tradeoff ratios. For example, Viscusi et al. (1991) asked respondents about their willingness to trade off the risk of contracting chronic bronchitis against the risk of death in an automobile accident.

Value measures based on substitutability can be expressed in terms of either willingness to pay (WTP), or willingness to accept compensation (WTA). WTP and WTA measures can be defined in terms of any good that the individual is willing to substitute for the good being valued. In the following discussion, money is used as the numeraire in which tradeoff ratios are expressed, but WTP and WTA could be measured in terms of any other good that mattered to the individual. The choice of a numeraire for measuring WTP or WTA is irrelevant in terms of its effect on how any one individual ranks alternative outcomes. However, as Brekke

(1997) has shown, the choice of a numeraire can affect the rankings of outcomes based on aggregation of welfare measures across individuals.

WTP is the maximum sum of money the individual would be willing to pay rather than do without an increase in some good such as an environmental amenity. This sum is the amount of money that would make the individual indifferent between paying for and having the improvement and forgoing the improvement while keeping the money to spend on other things. WTA is the minimum sum of money the individual would require to voluntarily forgo an improvement that otherwise would be experienced—it is the amount that would make a person indifferent between having the improvement and forgoing the improvement while getting extra money. Both value measures are based on the assumption of substitutability in preferences, but they adopt different reference points for levels of well-being. WTP takes as its reference point the absence of the improvement, while WTA takes the presence of the improvement as the base level of welfare or utility. In principle, WTP and WTA need not be exactly equal. WTP is constrained by the individual's income; but there is no upper limit on what that person could require as compensation for forgoing the improvement. Differences between WTP and WTA measures and the question of which measure is appropriate under various circumstances are discussed in more detail in Chapter 3.

Finally, we offer some words of caution based on our working in interdisciplinary research groups and serving on multidisciplinary advisory committees and panels. While semantic issues alone do contribute sizably to the misunderstandings that often arise when noneconomists are first exposed to an economist's notions of valuation, it is important to recognize that researchers from the academic disciplines that must be relied upon in valuing environmental and resource system changes do not always share the economists' commitment to the anthropocentric and preference-based perspectives on value. Nor do they all accept the assumptions of well-formed and stable preference orderings, substitutability, and rational choice that underlie our methods of economic valuation. The same thing is likely to be true of politically responsible decision-makers. Economists must be sensitive to this fact and be willing to engage in broad-ranging discussions of alternative value concepts, approaches to making choices about environmental policy, and alternative valuation methods. Economists must argue for care and rigor in applying the basic tenets of economic value in contexts in which it is important, such as benefit-cost analysis, so as to be sure that apples and oranges are not being mixed. However, there may be policy settings where other concepts may aid decision making (U.S. Environmental Protection Agency 2009). Economists who are not sensitive to these facts run the risk of being ignored in the policymaking process and undermining the value of economic analysis in the process.

Economic Values in Public Policy

There may be potential for substantial gains in economic welfare through better resource management and the judicious use of the principles of resource valuation in some cases, such as those involving the issues described at the beginning of this chapter. If the objective of management is to maximize the net economic values associated with the use of environmental and natural resources, then benefit-cost analysis becomes, in effect, a set of rules for optimum management and a set of definitions and procedures for measuring benefits and costs. Once the objective of maximum net economic value or economic efficiency has been accepted, policy becomes an almost mechanical (but not necessarily easy) process of working out estimates of marginal benefit and marginal cost curves and seeking their point of intersection.

Most current resource and environmental policy, however, is not based solely, or even primarily, on the efficiency criterion. One reason, of course, is that at the time that many of the basic policy objectives were established, it was not within the capability of analysts to provide the kind of information about values that would be required to implement the efficiency objective. But, it is also true that decision-makers may have other objectives besides economic efficiency. For example, decision-makers may be concerned with equity considerations, intergenerational effects, the sustainability of resource systems, or social risk aversion. Thus, it is not particularly useful to advocate benefit-cost analysis as a routine or simple decision rule. Rather, as Arrow et al. (1996) argued, it should be considered as a framework and a set of procedures to help organize available information. Viewed in this light, benefit-cost analysis does not dictate choices, nor does it replace the ultimate authority and responsibility of decision-makers and open public input processes. It is simply a tool for organizing and expressing certain kinds of information on the range of alternative courses of action. It is in the context of this framework for arraying information that the usefulness of value estimates must be assessed.

Some people may be distrustful of economists' efforts to extend economic measurements to such things as human health and safety, ecology, and aesthetics, and to reduce as many variables as possible to commensurate monetary measures. Some skepticism about the economist's penchant for monetary measurement is no doubt healthy, but it should not be overdone. It is sometimes argued that there are some things like human health and safety or the preservation of endangered species that cannot be valued in terms of dollars or some other numeraire. However, the real world often creates situations where tradeoffs between such things as reducing risks of death and some other things of value cannot be avoided. Where individuals can make choices for themselves about these tradeoffs, their values can be inferred from these choices; and where government policies affecting health and safety are involved, these policy choices imply values. The questions really are how the problem of making choices about such tradeoffs is to be approached and what information can be gathered to help in the problem of choice.

Consider a hypothetical and highly simplified case of an air pollutant. Assume that the following information is known with certainty. At present levels of emissions, the pollutant causes excess mortality of 10 deaths per year in the population at risk. Reducing emissions by 30 percent would cost $5 million and would reduce the excess mortality to 5 deaths per year. Reducing emissions by 60

Table 1.1 Hypothetical example of trade-offs

Level of control (percentage)	*Costs of control (millions of dollars)*	*Excess mortality (deaths per year)*
0	0	10
30	5	5
60	15	3

percent would reduce excess mortality to 3 deaths per year but would cost $15 million. This information can be displayed as in Table 1.1.

The problem is clearly one of tradeoffs between lives and the value of resources used up in the process of controlling emissions. If the monetary value of saving lives was known, the right-hand column of the table could be converted to dollar measures of benefits, and the appropriate benefit-cost rules could be applied to determine the optimum level of emissions control. But, in the absence of some agreed-upon basis for making lifesaving and control costs commensurable in dollar terms, no simple decision rule can be applied to determine the correct choice.

Choices of this sort are made in the political realm by decision-makers such as the administrator of the Environmental Protection Agency (EPA). Whatever the choice, there is an implicit value of lifesaving that is consistent with that choice and can be said to have been revealed by that choice. In this example, if the decision-maker chooses the 30 percent control level, the value of lifesaving is revealed to be at least $1 million per death avoided. The 30 percent control level "buys" 5 lives saved at a cost of $5 million. The choice further reveals that the value of lifesaving to the decision-maker is less than $5 million, since the decision-maker declined the opportunity to "purchase" the additional 2 lives saved that the additional $10 million of control costs would make possible. If the 60 percent control level had been chosen, this would have revealed a value of lifesaving of at least $5 million.

In this example, with only three data points the implicit value can only be determined within some range. If control costs and mortality as functions of the level of control were continuous relationships and known with certainty, the choice of a control level would imply a precise value of life. If it is assumed that the control level was established so as to equate marginal benefits and marginal costs, and marginal costs are known, the marginal benefit or value can be inferred.

A number of studies have shed light on the implied value of life saving by estimating the costs per death avoided for various regulatory policies. See, for example, Cropper et al. (1992), Van Houtven and Cropper (1996), Hamilton and Viscusi (1999), and Tengs et al. (1995).

In the previous example choice revealed value, rather than value determining the choice. Either way, the problem of valuation cannot be avoided, but it can be hidden. However, in a democratic society, the more open decision-makers are about the problems of making choices and the values involved, and the more information they have about the implications of their choices, the better their choices are likely to be. Estimates of values in monetary terms are one such source of information.

Most resource and environmental management problems have structures similar to the one just discussed. For example, in Table 1.1 the first column could be the rate of harvest in a forest, the second column the net economic value of the harvest, and the third column the probability of survival of an endangered species requiring old-growth forest habitat. For an example, see Montgomery, Brown, and Adams (1994). Reducing harvest levels increases the probability of survival of the species, but at a cost. The rate of harvest actually chosen implies something about the value to the decision-makers of increasing the probability of survival. Of course, the second column need not be in dollars. The tradeoff could be between having more timber wolves and fewer deer. The fact that there is no monetary measure in that example does not make it at its root any less an economic problem. Whatever choice is made about a population level for wolves implies something about the relative values placed on wolves and deer by the person making the decision. This example is another manifestation of the fundamental economic fact of scarcity—that is, that more of one thing means less of something else that people value.

Because policy choices about resources and environmental quality are made in a political context and are likely to involve comparisons and tradeoffs among variables for which there is no agreement about commensurate values, monetary benefit and cost data will not always be the determining factors in decision making. Benefit and cost estimates, however, are an important form of information. Their usefulness lies in the fact that they use easily understood and accepted rules to reduce complex clusters of effects and phenomena to single-valued commensurate magnitudes (that is, to dollars). The value of the benefit-cost framework lies in its ability to organize and simplify certain forms of information into commensurate measures (Arrow et al. 1996).

Classifications of Values

In this section, we describe some ways of classifying the types of environmental and resource service flows for which value measures might be desired. Of course, any classification system contains a certain element of arbitrariness. The usefulness of any particular classification depends on how well it illuminates important similarities and differences among types of service flows. Just which similarities and differences are important depends upon the particular questions being examined.

One basis for classification is the type of resource or environmental media. Environmental effects are often classified according to whether they stem from changes in air quality, water quality, land quality, and so forth. The current legal and administrative division of responsibilities for environmental management and pollution control is consistent with this basis for classification. However, it is becoming less and less relevant as cross-media effects are becoming better understood. For example, controlling emissions of nitrogen oxides from coal-burning power plants might be part of a cost-effective strategy for improving water quality in estuaries because of the impact of nitrate deposition on nutrient levels in these waters. In addition, controls on land use are a key part of the strategy for reducing nonpoint-source water pollution.

A second type of classification is based on the economic channel through which human well-being is affected. Environmental and resource service flows can be classified according to whether they convey their effects through the market system in the form of changes in incomes to producers and changes in the availability of, and prices for, marketed goods and services to consumers, or through changes in the availability of goods and services not normally purchased through markets. Goods and services not normally purchased could be, for example, health, environmental amenities such as visibility, and opportunities for outdoor recreation. The subject matter of this book is the methods and techniques for measuring the values of these latter nonmarket services. However, many of the policies for managing environmental and resource systems will affect the flows of both market and nonmarket goods and services—so policy assessments will often need to make use of market as well as nonmarket valuation methods.

A third way of classifying environmental and resource service flows is according to whether they impinge directly on humans, indirectly on humans through their impact on other living organisms, or indirectly through inanimate systems. Direct impacts on humans include the morbidity and mortality effects of associated air and water pollutants, hazardous wastes, pesticide residues and the like, and the nonhealth effects of pollutants manifesting themselves as odors, reduced visibility, and reduced visual attractiveness of outdoor settings.

Impacts on humans involving biological mechanisms and other organisms include those on the economic productivity of both managed and natural ecosystems, such as agricultural croplands, commercial forests, and commercial fisheries. Market valuation methods would be used to value these effects. There are also impacts on nonmarket direct service flows to people such as recreational uses of ecosystems for hunting, fishing, and nature observation; and there are impacts on indirect or intermediate ecosystem services, such as pollination, decomposition, biological pest control, and nutrient recycling.

Impacts acting through nonliving systems include: damages to materials and structures and increases in cleaning and repair costs at commercial activities, which would be measured by market valuation techniques; damages to materials and structures and increases in cleaning and repair costs for households, which would be measured by nonmarket valuation techniques; and impacts on weather and climate which would be measured by either market or nonmarket valuation techniques, depending on the nature of the activity affected.

Finally, we can distinguish between those services that an individual values because they make use of them in some way (use values), and those which are valued independent of any kind of observable use. These have been called nonuse values, or more recently passive use values. Questions of defining use and use vs. nonuse values will be taken up in Chapter 4.

Dealing with Uncertainties in Policymaking

One problem in carrying out an analysis on environmental benefits or costs is that the values for some physical, technical, or economic parameters of the model may not be known with certainty. The state of the art in measurement is not sufficiently far advanced to produce exact measures of value for many kinds of environmental and resource changes. This leads to the question: must policymakers wait for further research to produce exact measures before they can use value and benefit information to guide decision making? If not, how should they interpret the ranges of values that current research has produced?

To counsel waiting for exact measures is equivalent to saying that in many cases value measures should never be used. The state of the art cannot be expected to advance to the point of producing exact values for all kinds of environmental change. This is because of the inherent uncertainty and imprecision in measurement techniques based on statistical inference, and because of the fact that the true values held by individuals will vary with their circumstances (age, income, and so forth) and with the description of the specific changes being valued. So how are policymakers to proceed in the face of continued and inherent uncertainty about values?

The simplest approach is to base the calculations of benefits and costs on the expected values of the uncertain parameters and to base decisions on these expected values. However, decision-makers will often want to know more about the magnitude of the uncertainties in the estimates of benefits and costs. They could be provided with the upper and lower bounds of the ranges of values along with the expected values. Clearly, if the benefits of a policy calculated with the upper end of the range are less than the lower end of the range of estimated costs, the policy is unlikely to be justifiable on economic grounds; and if the benefits calculated with the lower end of the range exceed the upper end of the range of costs, the economic case for the policy is quite strong.

This simple-minded approach is a step in the right direction, but it can be criticized because it does not make use of all of the relevant information contained in the estimates making up the range of values. Formally, the range reflects only the information contained in the two estimates yielding the highest and lowest values. It ignores information on the quality of these two estimates, and it ignores all of the information contained in the other estimates that yielded values within the range. There is a way to make use of the results of all of the available estimates and to incorporate judgments about the quality of each of these estimates. This formal approach is based on viewing probabilities as statements about the degree of confidence held about the occurrence of some possible event. The approach involves assigning probabilities to all of the values produced by the available estimates, where a higher probability reflects a greater degree of confidence in that estimate. For example, the assignment of a probability of one to a particular estimate means we can be certain that this study has produced the correct value. Once the probabilities have been assigned, various statistical manipulations can be performed. For example, the expected value of the parameter in question (the mean of its probability distribution) can be calculated and used for benefit-cost calculations. The variance of the distribution can be used to determine confidence intervals on the value to be used, thus preserving for policymakers information on the uncertainty about values; or, when there are multiple uncertainties, Monte Carlo methods can be employed to draw from the assumed distributions to generate a probability distribution of outcomes.

Ex Post and Ex Ante Analysis of Values

The decision-maker who is trying to allocate scarce resources and is faced with a number of competing goals needs ex ante analyses of the effects of alternative policy actions to guide decision making. Ex ante analysis involves the prediction of the physical and economic consequences of policies on the basis of a model of the physical and economic processes involved. It involves visualizing two alternative states of the world, one with the policy in question and one without, and then comparing these alternative futures in terms of some established criterion such as net economic efficiency. Ex post analysis of a policy involves measuring the actual consequences of the policy by comparing the observed state with a hypothetical alternative—the state of the world without the policy. Ex post analysis, in effect, treats the policy as a controlled laboratory experiment except that the control is hypothetical rather than real. Natural resource damage assessment is an example of ex post evaluation, in that the damaged state is observed and must be compared with a hypothetical or counterfactual alternative in which there was no pollution event, but other relevant factors are assumed to have remained unchanged.

Ex post and ex ante analyses are not competitive alternatives. Rather, they should be viewed as complementary techniques for improving our knowledge. An ex post analysis of a policy can be viewed as a check on the validity of the ex ante analysis. The ex ante analysis is a prediction of what will happen; the ex post analysis is a check of what actually did happen.

It is particularly important that the economic analysis of environmental and resource policies includes ex post analysis. Our knowledge of the physical and economic systems on which present ex ante analyses are based is extremely limited. It is necessary not only to develop more comprehensive models of the physical, biological, and economic aspects of the system, but also to devote more effort to verifying these models through ex post comparisons of the predictions with observed results.

Although ex ante analyses of environmental, health, and safety regulations have become quite common in the United States (Hahn 2000), ex post analyses of environmental and resource policies are quite rare. An early study of the realized benefits and costs of U.S. Army Corps of Engineers' water resource development projects was done by Haveman (1972). A notable and controversial study was U.S. EPA's retrospective analysis of the benefits and costs of the Clean Air Act (U.S. Environmental Protection Agency 1997).[2]

It must be emphasized that the ex post verification of the analytical models used in resource valuation is not simply a comparison of actual results with predictions. Ex ante models are based upon some view of the future with projections of economic magnitudes such as population levels, real income, and price levels. Care must be taken in ex post analysis to sort out the effects of unforeseen developments, such as war or uncontrolled inflation, on the variables in question. For example, if the failure of income levels to rise on the projected path results in a shortfall of recreation benefits at a particular site, this is not a failure of the analytical model so much as a reflection on our inability to perceive the future. The real benefit of ex post analysis is in making the most of the opportunity to improve on the analytical models used.

Preview

The major task of this book is to review and summarize the basic theory of economic welfare measurement, and to present resource valuation and benefit measurement techniques that are consistent with this underlying theory. The next chapter provides an overview of valuation and welfare measurement methods and discusses the relationship between the economic methods of valuation and the physical and biological relationships that define the resource and environmental systems being valued.

Chapters 3 and 4 constitute the theoretical core of the book. Chapter 3 lays out the basic premises and value judgments that underlie the economic concept of benefits and presents the basic theory of the measurement of economic welfare changes. Chapter 4 introduces the basic methods and models for deriving welfare and value measures from the revealed choices of individuals, from observed changes in market prices, and from individuals' responses to hypothetical questions. These so-called stated preference methods, including willingness-to-pay surveys and bidding games, direct referendum questions, and questions about how individuals would rank alternative bundles of environmental and private goods are described in more detail in Chapter 12.

Chapter 5 extends the theory of value and welfare change to a situation of risk where people are uncertain about what the actual state of the world will be. Public policies toward the management of environmental resources can affect either the probabilities of alternative outcomes or the magnitudes of environmental services in alternative states of the world. Thus, this chapter describes the application of the basic theory of welfare change to evaluating environmentally induced changes

2 For some of the controversy generated by this report, see Crandall (1997), Lutter and Belzer (2000), and Brown et al. (2001).

in risks. Chapter 6 takes up the question of valuation across time and the role of discounting in welfare measurement.

The remaining chapters describe the application of the various models and methods for welfare measurement to specific situations such as the values of reducing the risk of premature mortality and improving human health (Chapter 7), the values of environmental changes affecting producers' costs and productivity (Chapter 8), the valuation of resources that support recreation activities (Chapter 9), applications of the hedonic price model to housing prices and wage rates (Chapters 10 and 11), and a more detailed treatment of stated preference methods (Chapter 12). Chapter 13 provides an overview of some recent developments in the literature, including benefits transfer methods, combining revealed and stated preference data, some of the implications of the relatively new field of behavioral economics, and the valuation of ecosystem services. Chapter 14 offers some conclusions and identifies areas where additional research is needed.

Mathematical Notation

In this book, we use the following conventions regarding mathematical notation, with exceptions noted where they occur:

- Vectors are represented by boldface uppercase letters; lowercase letters with or without subscripts represent values for individual variables in these vectors, for example, $\boldsymbol{X} = x_1, \ldots, x_i, \ldots, x_N$.
- The subscript letters i, j, k, m, and n and subscripted numbers index elements of vectors.
- The meanings of other subscripted letters are specified when they are first used.
- Superscripted letters are used to index such things as utility functions and production functions to specific individuals and firms. For example, $u^i(\boldsymbol{X})$ gives individual i's utility as a function of that individual's consumption of goods $x_1, \ldots, x_j, \ldots, x_J$.
- Primes and superscripted numerals represent specific values for variables. For example, M^0 represents the initial value of the variable M. Similarly, $\Delta x = x'' - x'$ means the change in x from x' to x''.
- Uppercase letters represent variables expressed as quantities of money. For example, M represents income, and CV is the compensating variation measure of welfare change.

References

Adler, Matthew D., and Eric A. Posner. 2006. *New Foundations of Cost-Benefit Analysis*. Cambridge: Harvard University Press.

Arrow, Kenneth J., Maureen L. Cropper, George C. Eads, Robert W. Hahn, Lester B. Lave, Roger G. Noll, Paul R. Portney, Milton Russell, Richard Schmalensee, V. Kerry

Smith, and Robert N. Stavins. 1996. "Is There a Role for Benefit-Cost Analysis in Environmental, Health, and Safety Regulation?" *Science* 272: 221–222.

Barbier, Edward B. 2011. *Capitalizing on Nature: Ecosystems as Natural Assets*. Cambridge: Cambridge University Press.

Botkin, Daniel B. 1990. *Discordant Harmonies: A New Ecology for the Twenty-First Century*. New York: Oxford University Press.

Bowes, Michael D., and John V. Krutilla. 1989. *Multiple-Use Management: The Economics of Public Forestlands*. Washington, D.C.: Resources for the Future.

Brekke, Kjell Arne. 1997. "The *Numéraire* Matters in Cost-Benefit Analysis." *Journal of Public Economics* 64(1): 117–123.

Brown, Gardiner M., Trudy A. Cameron, Maureen L. Cropper, A. Myrick Freeman III, Don Fullerton, Jane V. Hall, Paul Lioy, Richard Schmalensee, and Tom Tietenberg. 2001. "Misrepresenting ACCACA." *Regulation* 24(3): 4–5.

Callicott, J. Baird. 1989. *In Defense of the Land Ethic: Essays in Environmental Philosophy*. Albany: State University of New York Press.

Chameides, W.L., R.W. Lindsay, J. Richardson, and C.S. Kiang. 1988. "The Role of Biogenic Hydrocarbons in Urban Photochemical Smog: Atlanta as a Case Study." *Science* 241(4872): 1473–1475.

Common, M., I. Reid, and R. Blamey. 1997. "Do Existence Values for Cost-Benefit Analysis Exist?" *Environmental and Resource Economics* 9(2): 225–238.

Costanza, Robert, and Carl Folke. 1997. "Valuing Ecosystem Services with Efficiency, Fairness and Sustainability as Goals." In *Nature's Services: Societal Dependence on Natural Ecosystems*, edited by Gretchen C. Daily. Washington, D.C.: Island Press.

Crandall, Robert W. 1997. "The Costly Pursuit of the Impossible." *The Brookings Review* 15(3): 41–47.

Cropper, Maureen L., William N. Evans, Stephen J. Beradi, Maria M. Ducla-Soares, and Paul R. Portney. 1992. "The Determinants of Pesticide Regulation: A Statistical Analysis of EPA Decision Making." *Journal of Political Economy* 100(1): 175–189.

Daily, Gretchen C., Susan Alexander, Paul Ehrlich, Larry Goulder, Jane Lubchenco, et al. 1997. "Ecosystem Services: Benefits Supplied to Human Societies by Natural Ecosystems." *Issues in Ecology* no. 2 (Spring). Available at: www.esa.org/esa/wp-content/uploads/2013/03/issue2.pdf.

Daily, Gretchen C., Tore Söderqvist, Sara Aniyar, Kenneth Arrow, Partha Dasgupta, Paul R. Ehrlich, Carl Folke, AnnMari Jansson, Bengt-Owe Jansson, Nils Kautsky, Simon Levin, Jane Lubchenco, Karl-Göran Mäler, David Simpson, David Starrett, David Tilman, and Brian Walker. 2000. "The Value of Nature and the Nature of Value." *Science* 289: 395–396.

Daily, Gretchen C., Peter Kareiva, Stephen Polasky, Taylor Ricketts, and Heather Tallis. 2011. "Mainstreaming Natural Capital into Decisions." In *Natural Capital: Theory and Practice of Mapping Ecosystem Services*, edited by Peter Kareiva, Heather Tallis, Taylor H. Rickets, Gretchen C. Daily, and Stephen Polasky. Oxford: Oxford University Press.

Freeman, A. Myrick, III, Robert H. Haveman, and Allen V. Kneese. 1973. *The Economics of Environmental Policy*. New York: John Wiley.

Hahn, Robert W. 2000. *Reviving Regulatory Reform: A Global Perspective*. Washington, D.C.: AEI-Brookings Joint Center for Regulatory Studies.

Hamilton, James T., and W. Kip Viscusi. 1999. "How Costly is 'Clean?' An Analysis of the Benefits and Costs of Superfund Remediations." *Journal of Policy Analysis and Management* 18(1): 2–27.

Haveman, Robert H. 1972. *The Economic Performance of Public Investments*. Baltimore: The Johns Hopkins University Press for Resources for the Future.

Kareiva, Peter, Heather Tallis, Taylor H. Ricketts, Gretchen C. Daily, and Stephen Polasky. 2011. *Natural Capital: Theory and Practice of Mapping Ecosystem Services*. Oxford: Oxford University Press.

Kneese, Allen V., Robert U. Ayres, and Ralph C. d'Arge. 1970. *Economics and the Environment: A Materials Balance Approach*. Baltimore: The Johns Hopkins University Press for Resources for the Future.

Lutter, Randall, and Richard B. Belzer. 2000. "EPA Pats Itself on the Back." *Regulation* 23(3): 23–28.

Montgomery, Claire A., Gardiner M. Brown Jr., and Darius M. Adams. 1994. "The Marginal Cost of Species Preservation: The Northern Spotted Owl." *Journal of Environmental Economics and Management* 26(2): 111–128.

Nordhaus, William D., and Edward C. Kokkelenberg, eds. 1999. *Nature's Numbers: Expanding the National Economic Accounts to Include the Environment*. Washington, D.C.: National Academy Press.

Pigou, Arthur C. 1929. *The Economics of Welfare*. London: Macmillan.

Prugh, Thomas. 1999. *Natural Capital and Human Economic Survival*. 2d ed. Boca Raton: Lewis Publishers.

Smith, V. Kerry. 1988. "Resource Evaluation at the Crossroads." *Resources* 90: 2–6.

Tengs, Tammy O., Miriam E. Adams, Joseph S. Pliskin, Dana Gelb Safran, Joanna E. Siegel, Milton C. Weinstein, and John D. Graham. 1995. "Five-Hundred Life-Saving Interventions and Their Cost- Effectiveness." *Risk Analysis* 15(3): 369–390.

U.S. Environmental Protection Agency. 1997. *The Benefits and Costs of the Clean Air Act – 1970 to 1990*. Washington, D.C.: U.S. EPA.

U.S. Environmental Protection Agency. 2009. *Valuing the Protection of Ecological Systems and Services, a Report of the EPA Science Advisory Board*. Washington, D.C.: U.S. EPA.

Van Houtven, George, and Maureen L. Cropper. 1996. "When is a Life Too Costly to Save? The Evidence from U.S. Environmental Regulations." *Journal of Environmental Economics and Management* 30(3): 344–368.

Veisten, Knut, Stale Navrud, and Johnny S.Y. Valen. 2006. "Lexicographic Preference in Biodiversity Valuation: Tests of Inconsistencies and Willingness to Pay." *Journal of Environmental Planning and Management* 49(2): 167–180.

Viscusi, W. Kip, Wesley A. Magat, and Joel Huber. 1991. "Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-offs for Chronic Bronchitis." *Journal of Environmental Economics and Management* 21(1): 322–351.

Chapter 2

Measuring Values, Benefits, and Costs

An Overview

In the first section of this chapter, a simple general equilibrium model of an economy with nonmarket environmental and resource service flows is presented to show how the values of these flows emerge as shadow prices from the solution to a welfare maximization problem. The major types of methods for estimating values and benefits are then presented in the second section. The third section discusses some theoretical frameworks used in revealed preference models. The fourth section presents a simple model to illustrate the relationships between physical and biological changes in environmental and resource systems and the changes in well-being and values realized by the people affected by these changes. These models help to make clear the kinds of data required to carry out resource evaluations, and to show the roles of economists and physical and biological scientists in the evaluation process. The fifth section describes some issues about the noneconomic foundations of resource valuation. The chapter concludes with a discussion of the nature of cost in welfare theory.

Resource Values as Shadow Prices

Economic values can only be defined in terms of some underlying criterion that identifies what is to be considered good. As discussed in Chapter 1, in neoclassical welfare economics good is defined in terms of the well-being of individuals. Here, it is assumed that an individual's well-being can be represented by an ordinal utility function. Of course, this assumption is only a first step in defining what is good because it does not deal with interpersonal comparisons. Specifically, it does not answer the question of whether it is good when one individual's utility increases while another individual's utility decreases.

The concept of *Pareto optimality* is the route economists often choose for dealing with this set of problems. An allocation of resources, goods, and services in an economy is Pareto optimal if there is no feasible reallocation that can increase any one person's utility without decreasing someone else's utility. Of course, there are an infinite number of Pareto optimum allocations for an economy, each with a different distribution of utilities across individuals. In order to rank the allocations it is necessary to have a social welfare function that aggregates the utilities of the individuals, perhaps by assigning social welfare weights. If such a social welfare function exists, Pareto optimality is a necessary, but not sufficient, condition for maximizing that function. Hence, despite its limitations, Pareto optimality has usefulness in evaluating economic outcomes.

Pareto optimality is the solution to a constrained maximization problem in which some of the constraints are the exogenously determined environmental and resource service flows. The shadow prices on these constraints are the economic values of these service flows. In the general equilibrium model used here, it is clear that the values assigned to the environmental and resource services flows are not fixed parameters, but are determined by their roles in enhancing individuals' well-being, and arise from their scarcity or limited availability. This model is a variation of the general equilibrium externalities model presented in Baumol and Oates (1988).

In this model the subscript j indexes a vector of J commodities ($j = 1, \ldots, J$), i indexes the N individuals in the economy, and k indexes the K firms. Each of the commodities is divisible with well-defined and enforceable property rights. The variable names for these commodities reflect their role in economic activity. $\boldsymbol{X}$ represents the vector of commodities used as consumer goods by individuals. Specifically, let x_{ji} = the amount of commodity j consumed by individual i, ($j = 1, \ldots, J$) ($i = 1, \ldots, N$). $\boldsymbol{Y}$ represents the vector of commodities being produced by firms, where negative values indicate commodities being used as resource inputs to production. Specifically, y_{jk} = the amount of commodity j produced (used) by firm k, ($j = 1, \ldots, J$) ($k = 1, \ldots, K$). There is also an endowment of commodities represented by $\boldsymbol{S}$: s_j = the total of endowment of commodity j available to the economy ($j = 1, \ldots, J$).

For each commodity, the initial endowment plus the quantities produced by firms must just equal the sum of the quantities used as inputs by firms and consumed by individuals. This requirement is expressed in a set of production-consumption constraints:

$$s_j + \sum_{k=1}^{K} y_{jk} - \sum_{i=1}^{N} x_{ji} = 0 \qquad j = 1, \ldots, J. \tag{2.1}$$

In addition, let there be a resource service flow, r. The endowment of r is determined exogenously and is both nonrival and nonexcludable; hence, it is a public good. For simplicity, assume that r does not vary across space. This means that r takes the same value for all individuals. Finally, let d be some environmental quality parameter, say the concentration of a pollutant at a specific point in space. The level of d will depend on the discharges of pollutants by firms and will determine the quantities of pollution experienced by each individual. These relationships will be specified below.

Individuals' preferences can be represented by utility functions:

$$u^i = u^i(\boldsymbol{X}_i, r, c_i) \quad i = 1, \ldots, N \tag{2.2}$$

Let $\partial u^i / \partial x_{ji} \geq 0$, $\partial u^i / \partial r \geq 0$, and $\partial u^i / \partial c_i < 0$, where c_i is the concentration of the pollutant to which the ith individual is exposed and $\boldsymbol{X}_i \equiv (x_{1i}, \ldots, X_{Ji})$.

The production side of the economy can be represented by a set of production functions for multiproduct firms:

$$f^k(y_{1k}, \ldots, y_{jk}, \ldots, y_{Jk}, r, d_k) = 0 \quad k = 1, \ldots, K, \tag{2.3}$$

where y_{jk} is the production of good j by firm k, r is the level of the nonmarket resource service flow, and d_k is the firm's contribution to the environmental quality parameter d as represented by, for example, emissions of a pollutant. Let $\partial f^k/\partial y_{jk} \geq 0$, $\partial f^k/\partial r \leq 0$, and $\partial f^k/\partial d_k \leq 0$, which implies that both r and d_k can be viewed as inputs into the production process.

The variable d is some function of the emissions of all firms. For simplicity, let $d = \sum_{k=1}^{K} d_k$. Also, assume that $c_i = \alpha_i \times d$, where α_i relates each individual's exposure to the pollutant to the aggregate level of pollution, d. For a perfectly mixed unavoidable pollutant, $\alpha_i = 1$ for all i. It would also be possible to represent averting behavior by making α_i a choice variable, perhaps a function of some x_{ji}.

The Pareto optimum conditions for this economy can be found by determining the conditions that maximize each individual's utility, subject to the production function and resource constraints and the constraint that all other individuals' utilities are held constant at some level u^{i*} representing the status quo. This procedure makes it clear that any Pareto optimum allocation is only as good as our judgment about the associated distribution of utilities across individuals. For individual one, the problem is to choose the values for $\boldsymbol{X}_{j1}$ and d_1 so that

$$\max\ L_1 = u^1(\cdot) + \sum_{i=2}^{N} \lambda_i \left[u^i(\cdot) - u^{i*}\right] - \sum_{k=1}^{K} \mu_k f^k(\cdot) + \sum_{j=1}^{J} \rho_j \left(s_j + \sum_{k=1}^{K} y_{jk} - \sum_{i=1}^{N} x_{ji}\right) + \gamma\left(r^* - r\right) \tag{2.4}$$

where r^* is the exogenously determined quantity of r. Assuming an interior solution and letting $\lambda_1 = 1$, the first-order conditions for a Pareto optimum are:

$$\lambda_i \left(\frac{\partial u^i}{\partial x_{ji}}\right) - \rho_j = 0 \Rightarrow \lambda_i \left(\frac{\partial u^i}{\partial x_{ji}}\right) = \rho_j \qquad i = 1, \ldots, N; j = 1, \ldots, J \tag{2.5}$$

$$-\mu_k \left(\frac{\partial f^k}{\partial y_{jk}}\right) + \rho_j = 0 \Rightarrow \mu_k \left(\frac{\partial f^k}{\partial y_{jk}}\right) = \rho_j \qquad k = 1, \ldots, K; j = 1, \ldots, J \tag{2.6}$$

$$\sum_{i=1}^{N} \lambda_i \left(\frac{\partial u^i}{\partial r}\right) - \sum_{k=1}^{K} \mu_k \left(\frac{\partial f^k}{\partial r}\right) - \gamma = 0 \tag{2.7}$$

$$\sum_{i=1}^{N} \lambda_i \left(\alpha^i \frac{\partial u^i}{\partial d}\right) - \sum_{k=1}^{K} \mu_k \left(\frac{\partial f^k}{\partial d_k}\right) = 0 \tag{2.8}$$

where $\lambda_i > 0$, $\mu_k > 0$, $\rho_j > 0$, and $\gamma > 0$.

Conditions (2.5) and (2.6) generate the marginal rates of substitution and marginal rates of transformation for the marketed goods. They can be interpreted as saying either that each individual's marginal valuation of each good must equal the marginal cost to firms of its production, or that each individual's marginal rate of substitution between any pair of goods must equal each firm's marginal rate of transformation between that pair of goods.

In condition (2.7), the Lagrangian term γ is the shadow value for r^* (that is, it gives the increase in the objective function for a marginal increase in the constrained resource service r). Condition (2.7) also shows that the shadow price depends on the utility and production functions. Specifically, the shadow price is equal to the sum of the marginal values attached to r by all individuals and producers in the constrained solution.

The first term in (2.8) is the value to individuals of reducing firm k's emissions by one unit. Condition (2.8) says, in effect, that Pareto optimality requires that the aggregate value of reducing the emissions of firm k must be just equal to the marginal cost of that reduction (the second term in the equation). The second term in (2.8) can also be interpreted as the marginal value to the kth firm of being able to emit one more unit to the environment—that is, it is the marginal value of the waste receptor services of the environment.

Conditions (2.7) and (2.8) give the marginal values for changes in r and d_k. They also imply the existence of aggregate marginal willingness-to-pay curves that give the marginal willingness to pay for the service flows as functions of the quantities of the flows being supplied. As is shown in Chapter 3, willingness to pay can be taken to be equal to the area under such a marginal willingness-to-pay curve. Value estimation, then, involves determining directly or indirectly the shapes of these marginal willingness-to-pay curves for environmental services.

If the services of the environment could be purchased in a perfectly functioning market, there would be observable demand functions for them, making estimation of the marginal willingness-to-pay curves a standard econometric problem. Also, then, environmental and resource management would not be an important public policy matter. However, environmental and resource service flows typically have characteristics such as nonexcludability and nonrivalry in consumption, which make it difficult or impossible for markets for these services to function well. Often, individuals are not free to vary independently the level of the services that they consume. The public good character of environmental services then leads to market failure; and without a market, there are no price and quantity data from which the demand relationships can be estimated.

To sum up, the economic values of r and of reducing d emerge as part of the solution to the welfare maximization problem. These values are context dependent in the sense that changes in preferences (equation 2.2), the production technology (equation 2.3), or the resource endowments and constraints will affect these values. This point is worth emphasizing: just as the values of private goods vary across individuals, so too will the values of nonmarket goods. The value of an improvement in water quality or the reduced risk of illness may be quite different across individuals due to differences in preferences and/or source constraints. There will rarely be a single value associated with a nonmarket good.

This context dependence also means that if the economy is not at a Pareto optimum allocation as defined by equations (2.5) through (2.8), for example because of an additional constraint, then the shadow prices or values attached to r and d will be different.

Methods for Measuring Values

One principle distinction among methods for valuing changes in environmental goods is based on the source of the data (Mitchell and Carson 1989, 74–87). The data can come either from observations of people acting in real-world settings where people must live with the consequences of their choices or from people's responses to hypothetical questions of the form "what would you do if … ?" or "how much would you be willing to pay for … ?" It is common in the literature to refer to these as *revealed preference* and *stated preference* methods, respectively.

Revealed preference methods are based on actual behavior reflecting utility maximization subject to constraints. One type of revealed preference method is based on observed choices in a "take-it-or-leave-it" setting. For example, a survey might collect information from a family about whether it visited a particular environmental attraction on the previous weekend. Essentially, the family faced a take-it-or-leave-it decision with respect to that visit—either they took the visit and enjoyed the site, or they stayed home and did something else. If the family chose to take the trip, this information "reveals" that the value of the trip exceeded the costs that the family undertook to experience the attraction. In this case, the reported behavior reveals only whether the value of the offered good to the individual was greater or less than the offering price (the cost of admission and travel). Because of the information limitations provided by such data, it is typically necessary for analysts to make assumptions about preferences so that models can be estimated with the data. For this reason (and others discussed below), values from revealed preference methods are subject to limitations.

In many instances, the environmental service does not have a direct offering price, but sometimes its quantity does affect the choices people make about other things such as quantities of market goods. In these cases, the value of the environmental service can be inferred through the application of some model of the relationship between market goods and the environmental service. Most such models are based on the assumption of some kind of substitute or complementary relationship between the environmental service and marketed goods and services. Examples of these models include the household production model (which includes models of household spending on cleaning and on repair of materials damaged by air pollution), the travel cost demand model for visits to a recreation site, and the hedonic property value and hedonic wage models. Revealed preference methods involve a kind of detective work in which clues about the values individuals place on environmental services are pieced together from the evidence that people leave behind as they respond to prices and other economic signals. The basic properties of these models are discussed in Chapter 4.

The principal difference between revealed preference and stated preference methods is that the latter draw their data from people's responses to hypothetical questions rather than from observations of real-world choices. The earliest stated preference techniques involved asking people directly about the values they place on environmental services by creating, in effect, a hypothetical market. As the responses are contingent upon the specific conditions laid out in the hypothetical market, this form of stated preference methods is broadly referred to as *contingent valuation*. A variety of elicitation formats are possible. For example, people could be directly asked what value they place on a specified change in environmental services. The responses, if truthful, are direct expressions of value, and would be interpreted as measures of willingness to pay. Because the format allows the respondent to provide any possible measure of value, the term *open-ended contingent valuation method* is conventionally used to refer to approaches based on this form of questioning.

While a variety of contingent valuation elicitation formats have been considered, the most popular today is the *dichotomous choice referendum format*, which asks for a yes or no answer to the question: "Given a cost to you of $\$X$, would you vote in favor of a referendum that would achieve the following changes … ?" Responses to such questions reveal only an upper bound (for a no) or a lower bound (for a yes) on the relevant welfare measure. Discrete choice methods applied to a large sample of individual responses can be used to estimate willingness-to-pay functions or indirect utility functions from data on responses and on the characteristics of the individuals in the sample.

In a second type of stated preference question, known as *contingent behavior* questions, individuals are asked how they would change the level of some activity in response to a change in an environmental amenity. If the activity can be interpreted in the context of some behavioral model, such as an averting behavior model or a recreation demand model, the appropriate revealed preference valuation method can be used to obtain a measure of willingness to pay, as if the reported behavioral intentions were actual behaviors. McConnell (1986), for example, applied a recreation demand model to questions of the form "how often would you visit these beaches if they were free of PCBs?" in order to estimate the damages (resource value lost) from the pollution of the waters of New Bedford Harbor, Massachusetts, with polychlorinated biphenyls.

In a third form of stated preference question, respondents are given a set of hypothetical alternatives, each depicting a different situation with respect to the available environmental amenities and other characteristics, and are asked to rate or rank the alternatives in order of preference, or to simply pick the most preferred alternative. Several names have been applied to variations of this approach including *attribute-based stated choice* and *choice experiments*. The rankings or choices can then be analyzed to determine, in effect, the marginal rate of substitution between any other characteristic and the level of the environmental amenity. If one of the other characteristics is a monetary price, then it is possible to compute the respondent's willingness to pay for the good on the basis of the ranking of alternatives. For a more complete discussion, see Holmes and Adamowicz (2003).

Some issues and problems in stated preference methods are specific to the particular form of the question being asked. For example, when people are asked how much they would be willing to pay for something, they might say "zero" because they reject the idea of having to pay for something they consider to be rightfully theirs. Other problems are generic to all methods based on hypothetical questions, for example, problems in scenario specification, sampling, and item nonresponse. The major questions regarding all stated preference methods concern the validity and reliability of the data; that is, whether the hypothetical nature of the questions asked inevitably leads to some kind of bias or results in so much "noise" that the data are not useful for drawing inferences. Further discussion of these questions is left to Chapter 12.

The Methodology of Revealed Preference Models

Because the revealed preference methods for measuring values use data on observed behavior, some theoretical framework must be developed to model this behavior, and to relate the behavior to the desired monetary measures of value and welfare change. A key element in the theoretical framework is the model of the optimizing behavior of an economic agent (individual or firm) that relates the agent's choices to the relevant prices and constraints, including the level of environmental or resource quality q. If a behavioral relationship between observable choice variables and q can be specified and estimated, this relationship can be used to calculate the marginal rate of substitution between q and some observed-choice variable with a money price tag, thereby revealing the marginal value of changes in q.

Welfare measurement in the case where changes in q affect individuals involves three steps. The first is to derive the expression for willingness to pay as a function of the environmental variable, usually either from the indirect utility function or the expenditure function. This expression gives the compensating change in income that holds utility constant for the change in the environmental parameter. The second step is to develop a model of individual utility maximizing behavior that relates the individual's choices to the relevant prices and constraints, including the level of environmental quality. The first-order conditions for optimization can then be derived. These first-order conditions involve equating measures of marginal value to price, or equating some marginal rate of substitution or marginal rate of transformation to some price ratio. The third step is to examine the model to see whether the first-order conditions include a relationship between the desired marginal value for the environmental change and some observable variable. If they do, then the observable variable can be taken as a measure of the marginal change in welfare. Several types of models using this methodology are presented in Chapter 4.

Drawing inferences about the marginal values of environmental changes can be challenging. However, policymakers often need even more information than this. Often, they will need estimates of values for nonmarginal changes. Deriving these estimates is considerably more difficult since what are needed are not the marginal values but the marginal value *functions* (see for example Kuminoff, Smith, and Timmins 2013 and Phaneuf, Carbone, and Herriges 2009). More will be said about how these might be obtained in subsequent chapters.

The measures of value and welfare change derived from optimization models often produce results that are quite different from those of the naive models employed in the early literature on value and benefit measurement. The early models were often based on what has come to be known as the *damage function approach*. The damage function approach involves estimating some physical relationship between a measure of environmental quality (or its converse, pollution) and some physical measure of damage or loss (such as number of workdays lost to sickness in the case of health, or percentage of crop lost in the case of effects of air pollution on agricultural productivity). Then some unit price is applied to the physical measure to convert it to monetary terms. In some studies, for example, lost wages and medical costs were used to determine the value of avoiding one day of illness induced by air pollution. Similarly, the market price of a crop was often used to determine the value of lost productivity. The benefit of a pollution control program would be estimated as the reduction in the damages calculated according to this damage function approach.

The damage function approach can be considered naive for at least two reasons. First, this approach implies a model of the world in which behavioral and market responses to changes in q are implicitly ruled out. Farmers, in fact, can adjust to changes in air pollution by changing cultivation practices, shifting to more resistant cultivars of the same crop, or even changing to entirely different crops that are less sensitive to pollution. Furthermore, the prices of agricultural crops may change because of changes in crop supplies. It may be that the changes in prices are of greater significance to human welfare than the changes in physical yields of crops. Similarly, people can choose defensive or mitigating activities in response to air pollution that affects health. These behavioral changes result in welfare consequences that have a monetary dimension, which should be taken into account in calculating values.

In addition to behavioral changes, a second problem with some of the early approaches is that they sometimes used the costs of treating an externality as the value of avoiding the externality. This is still a common practice in some health studies where lost wages and/or costs associated with medical care are used to estimate the value of avoiding the morbidity or mortality. This correspondence is true only in some cases (Bockstael and McConnell 2007, ch. 8) and will be discussed in greater length in Chapter 7.

A Model of Environmental and Resource Values

Although measuring values involves the use of economic theory and technique, value measures must be based on other types of knowledge. For example, estimates of the value of a salt marsh in sustaining a marine fishery must be based on knowledge of the biological and ecological links between the marsh and the exploited fish species. Estimates of the health benefits from air pollution control must be based on scientific knowledge of the relationship between pollutant concentrations and human health; and estimates of the recreational fishing benefits stemming from water pollution control require knowledge of the relationships among pollutant levels, biological productivity, and anglers' activities (Freeman 1995). Lack of knowledge of these relationships may be, in some instances, a major barrier to empirical measurement of values. This section lays out a very simple model for examining the relationship between the economic concept of value and the physical and biological dimensions of the resource system being valued. The model helps to make clear that economic valuation requires some knowledge of the underlying physical and biological relationships that determine the quantity and quality of environmental and resource service flows.

The economic values of the service flows from a resource-environmental system can be viewed as the product of three sets of functional relationships. The first relates some measure of environmental or resource quality to the human interventions that affect it. Let q represent a qualitative or quantitative measure of some environmental or resource attribute. Examples include the biomass of some species of fish of commercial or recreational interest, the stock of standing timber in a forest, or the concentration of some pollutant in the atmosphere. Two kinds of human intervention need to be specified. One involves the unregulated activities of the market economy (for example, the commercial exploitation of a fishery or the discharge of pollutants into the air), and these will be left implicit in the relationships presented here. The other kind of intervention is government actions taken to prevent or ameliorate the adverse impacts of unregulated market activities, or to protect or enhance the value of market and nonmarket services provided by the environment. Let $\boldsymbol{G}$ represent the set of government interventions. For example, if q represents the population of waterfowl, $\boldsymbol{G}$ could be the stock of protected habitat and breeding grounds. Alternatively, $\boldsymbol{G}$ could represent a set of regulations designed to attain a stated ambient air quality standard, or it could be a management plan for a national forest. Let us represent the relationship between q and $\boldsymbol{G}$ as

$$q = q(\boldsymbol{G}). \tag{2.9}$$

As discussed below, this relationship could be quite complex in its spatial and temporal dimensions.

Where the government regulates private activities that influence q, the effect of a change in $\boldsymbol{G}$ on q can depend in complex ways on the responses of private decision-makers to the public regulation. The most obvious example of this dependence is the question of compliance with pollution control regulations. For any given $\boldsymbol{G}$, q increases as the degree of compliance with the regulation increases. Some public regulations by their very nature only indirectly link to the relevant environmental quality measure. An example is the automotive emissions standards set under the Clean Air Act, which regulate emissions of certain pollutants in grams per mile traveled by the automobile. In this case, the effect of a change in $\boldsymbol{G}$ on q depends also on how automobile use is affected. To take account of these complexities, it might be more appropriate to write q as a function of both $\boldsymbol{G}$ and some measure of private responses to the government regulation, $\boldsymbol{R}(\boldsymbol{G})$:

$$q = q[\boldsymbol{G}, \boldsymbol{R}(\boldsymbol{G})]. \tag{2.9′}$$

The second set of functional relationships involves the human uses of the environment or resource and their dependence on q. Let $\boldsymbol{X}$ represent the levels of some set of activities involving use of the environment or resource. For example, $\boldsymbol{X}$ could be days of recreational activity on some water body, tons of fish caught from some commercially exploited fishery, and, where human health depends on the level of environmental quality, some measure of health status. Typically, the level of $\boldsymbol{X}$ will depend not only on q, but also on the inputs of labor, capital, and other materials and resources including time, and these will also depend on q. For example, if $\boldsymbol{X}$ is agricultural output and q is air pollution, farmers might adjust to changes in pollution by changing inputs of water, fertilizer, or labor. Let $\boldsymbol{Y}$ represent these other inputs into the production of environmental services or activities based on the resource. The second functional relationship can be written as

$$\boldsymbol{X} = \boldsymbol{X}\,[q, \boldsymbol{Y}(q)]. \tag{2.10}$$

This expression ignores possible feedbacks from $\boldsymbol{X}$ to q such as when agricultural output is associated with increases in water pollution from pesticide or fertilizer runoff.

The third set of functional relationships gives the economic value of the uses of the environment. Let V represent the money value of the flow of services or activities based on the environment or resource. The relationship

$$V = V(\boldsymbol{X}) \tag{2.11}$$

embodies whatever value judgments society has adopted for economic welfare purposes. Here it is assumed that the value function is a simple aggregation of individuals' values; but $V(\boldsymbol{X})$ could also incorporate social welfare weights that reflect some social equity goals. Alternatively, $V(\boldsymbol{X})$ could incorporate concepts of environmental ethics or social norms. Also, this expression could incorporate nonuse values as in $V = V(\boldsymbol{X}, q)$.

By substitution of (2.9′) and (2.10) into (2.11), we have

$$V = V(\boldsymbol{X}\{q[\boldsymbol{G}, \boldsymbol{R}(\boldsymbol{G})], \boldsymbol{Y}[\boldsymbol{G}, \boldsymbol{R}(\boldsymbol{G})]\}). \tag{2.12}$$

The marginal value of the change in $\boldsymbol{G}$ can be found by taking the total derivative of equation (2.12) to reflect the private adjustments of $\boldsymbol{R}$ and $\boldsymbol{Y}$ to the public intervention. The benefits of a nonmarginal change in policy intervention that increases $\boldsymbol{Y}$ are given by

$$\begin{aligned} B \equiv \Delta V = {} & V\left(\boldsymbol{X}\left\{g\left[\boldsymbol{G}^{1},\boldsymbol{R}\left(\boldsymbol{G}^{1}\right)\right],\boldsymbol{Y}\left[\boldsymbol{G}^{1},\boldsymbol{R}\left(\boldsymbol{G}^{1}\right)\right]\right\}\right) \\ & -V\left(\boldsymbol{X}\left\{g\left[\boldsymbol{G}^{0},\boldsymbol{R}\left(\boldsymbol{G}^{0}\right)\right],\boldsymbol{Y}\left[\boldsymbol{G}^{0},\boldsymbol{R}\left(\boldsymbol{G}^{0}\right)\right]\right\}\right) \end{aligned} \tag{2.13}$$

where $\boldsymbol{G}^0$ and $\boldsymbol{G}^1$ are the pre- and post-policy levels of intervention and $\boldsymbol{R}$ and $\boldsymbol{Y}$ are optimally adjusted to the change in $\boldsymbol{G}$.

The set of relationships represented by equation (2.9′) is largely noneconomic in nature because it involves a variety of physical and biological processes. However, there is a social science or behavioral component to the private responses to $\boldsymbol{G}$. The relationship reflected in equation (2.11) is wholly within the realm of economics because it involves the theory of economic welfare and the use of economic data. The set of relationships reflected in equation (2.10) represents the interface between the natural science and social science disciplines. Some aspects of these relationships, for example how recreation use varies with changes in water quality, are primarily behavioral or social. Other aspects are almost wholly physical or biological, as in the effects of air pollution on human health and mortality. However, even here, to the extent that people can “defend themselves” against the adverse effects of air pollution (say, by purchasing home air purifiers), or mitigate the symptoms of illness induced by air pollution, behavioral relationships are embedded in equation (2.10). The effect of an air pollutant on a particular type of vegetation is also a biological question; but if farmers alter crop patterns as a way of adapting to changes in air pollution, then the behavioral and biological aspects of the relationship must be considered together.

Figure 2.1 illustrates the three sets of relationships for the case of the benefits associated with an improvement in ambient water quality. Varieties of substances are discharged into water bodies. Reductions in the discharges can affect physical, chemical, and biological indicators of water quality, such as dissolved oxygen, temperature, algae levels, and fish populations (stage 1). Changes in the indicators can be predicted with water quality models. The resulting water quality, as measured by the indicators, can in turn affect human uses of the water body (stage 2). These could include either withdrawal uses (for example, for industrial or municipal water supply or irrigation) or instream uses (for example, for fishery production or recreation). The major difficulty at stage 2 arises from the fact that only rarely is the level of use a simple function of a single water quality indicator like dissolved oxygen. Rather, some uses (as in the case of commercial fisheries and recreation) depend in complex ways on the whole range of physical, chemical, and biological water quality indicators. The feedback loops from stage 2 to stage 1 reflect possible impacts of changes in human uses on measures of water quality,

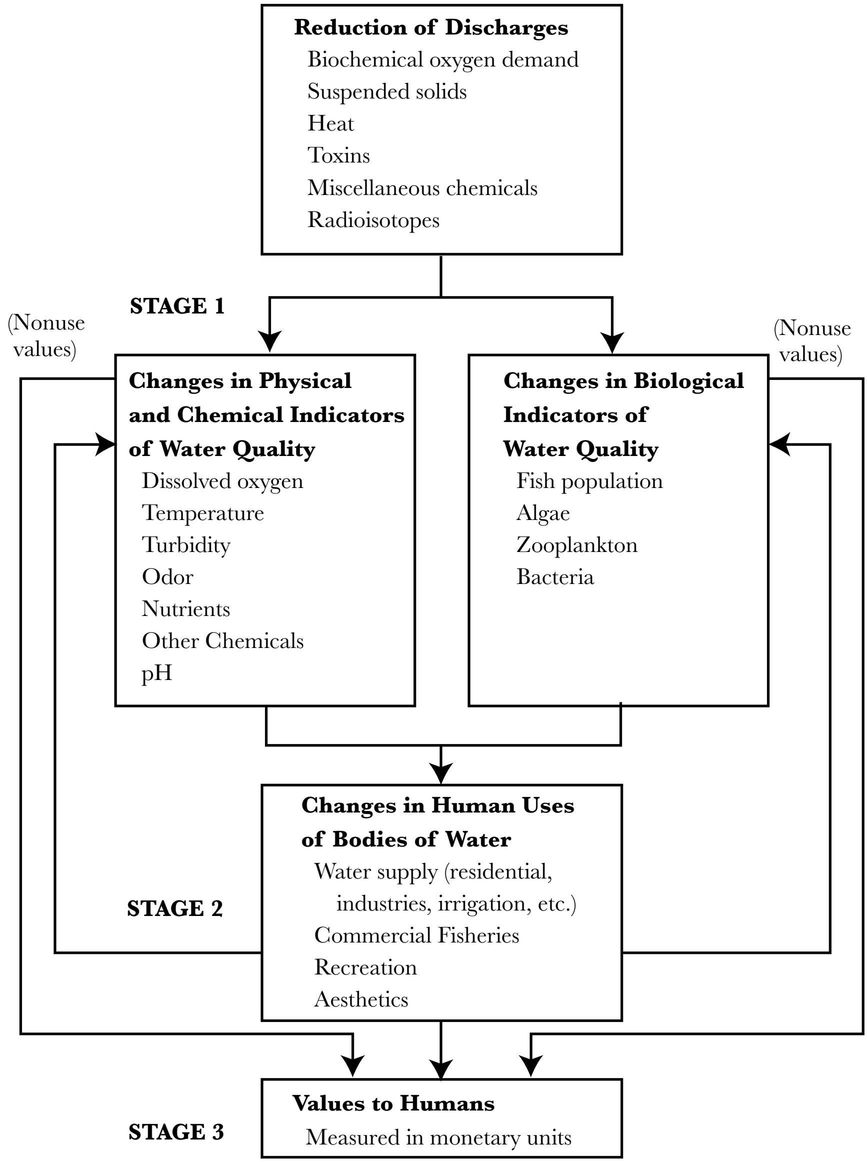

Figure 2.1 The production of benefits from improved ambient water quality

for example reductions in fish stocks due to overfishing. Figure 2.1 also shows nonuse values that are independent of stage 2.

In our illustration, estimating water quality benefits involves determining the monetary values that people place on such things as improved recreational opportunities, increases in fish production, and the availability of particular species of fish (stage 3). Regarding the analysis of this stage, there is a well-developed theory of economic value. As discussed earlier, there are a number of approaches for estimating these values under different circumstances.

The Noneconomic Foundations of Resource Valuation

In a book that is about determining the economic values attached to the services affected by environmental change, the discussion cannot proceed far without acknowledging the importance of the relationship between environmental and resource quality and the uses of the environment. For example, the value of a recreation user-day at a lake is affected by fish populations and species distribution, algae levels, the number and type of bacteria present, temperature, smell, turbidity, and concentration of toxic substances. Further complicating matters, an increase in the magnitude of one characteristic may affect one use favorably while affecting an alternative use in a negative way. Higher water temperatures, for example, may make for better swimming while adversely affecting trout and salmon populations. Industrial discharges of acids may adversely affect recreation and fisheries while improving the value of water for industrial uses because of retarded algae growth.

The difficulties in tracing out the effects of the discharge of a pollutant on the many parameters of environmental quality and, in turn, their effects on human uses of the environment substantially limit our ability to do careful benefit-cost analyses of environmental quality and improvements. This has not always been fully appreciated by many advocates of greater use of benefit-cost analysis in this field.

In 1968 Allen Kneese wrote,

> I believe that our limited ability to evaluate the recreational losses associated with poor quality water, or conversely, the benefits of water improvement, is an extremely important barrier to rational water quality management … The first [complexity] is the relationship between the level of various water quality parameters and the recreational attractiveness of the water resource. This relationship can be viewed as being composed of two linkages: a natural one and a human one. I think these are both about equally ill-understood. It is my impression … that the biological sciences are almost never able to tell us specifically what difference a change in measured parameters of water quality will make in those biological characteristics of the water that contribute to its recreational value … Perhaps the undeveloped state of forecasting is a result of the fact that biologists have seldom been confronted with the types of questions we would now like them to answer … There is also a human linkage that is ill-understood. What quality characteristics of water do human beings find attractive for recreation? This is still largely an area of ignorance.
>
> (Kneese 1968, 180–181)

Although substantial progress has been made in the past 45 years or so since Kneese wrote this, there are still substantial gaps in our understanding of some of these linkages. What is true of water quality and recreation is also true of the other uses we make of water bodies; and it is also true of the other dimensions of environmental quality and the uses we make of the environment.

Describing Resource and Environmental Quality

An analysis of the value of a resource, or of the benefits of an environmental or resource policy change, must begin with a description of the resource flow or some measure of environmental quality. This description requires choices about what attributes or characteristics of the resource-environmental service are important. Suppose the question is, "What will be the benefits of achieving the auto emissions standards mandated by federal law?" In order to answer this question it is necessary to determine what things that matter to people are adversely affected by automotive emissions, and then to trace out the links between emissions, and those things that are valued. It is now understood that automobile emissions and the subsequent products of their photochemical reactions adversely affect human health, visibility, and agricultural productivity, among other things. Therefore, in order to estimate values, it is necessary to determine what specific measures of air quality are linked to these effects and how these measures of air quality are affected by the mandated standards. The measures used to characterize effects on visibility may be quite different from those relevant to effects on human health.

Two kinds of problems arise in this stage of the analysis. The first concerns the choice of parameters for describing the resource or environmental quality. The second involves determining the functional relationship between the policy instrument and the resource service flow or environmental quality measure.

Consider the case of water quality. A single effluent discharge can contain many substances that affect water quality—oxygen-demanding organic wastes (biochemical oxygen demand), suspended solids, waste heat, and toxic chemicals are all examples. When these substances enter the waterway, they affect—in sometimes simple and sometimes complex ways—such measurable components of water quality as dissolved oxygen, temperature, and concentrations of chemicals. A nondegradable chemical substance, for example, will simply be diluted, and its concentration in the water body will be a calculable fraction of its concentration in the effluent stream. In contrast, organic wastes affect water quality parameters in a more complicated way. As they are degraded by bacteria, they reduce dissolved oxygen levels to an extent and at a rate dependent on water temperature, wind, rates of river flow, and other physical and biological characteristics of the receiving water.

Some of the physical measures of water quality, such as turbidity and smell, affect human uses of the water directly. In addition, these and other physical parameters affect the stream ecology in complex and not always well-understood ways. The populations and species distributions of fish, algae, zooplankton, and bacteria may also be affected, and not necessarily in the same direction, by changes in the physical and chemical parameters of water quality.

Even providing a descriptive characterization of this first stage is a formidable task. Water quality cannot be represented by a single number on some scale, but rather is an *n*-dimensional vector of the relevant parameters. Which subsets of these parameters are most important in influencing the uses of a water body (commercial fishing, boating, or swimming, for example) is still a major question for research.

Developing predictive models for these parameters is also a major research priority. The most commonly used water quality models relate dissolved oxygen to the biochemical oxygen demand of discharges of organic wastes. Dissolved oxygen levels, however, are only crudely related to the suitability of a water body for fishery production or recreational use.

In the case of air quality, the choice of parameters is somewhat easier, but not without pitfalls. It is only in the past 25 years or so that attention has turned from measurement of sulfur dioxide to its transformation products, sulfate particles. The latter measure of air quality is now known to be a more important air pollutant indicator than sulfur dioxide because of its effects on human health and ecology (U.S. Environmental Protection Agency 1997 and references therein).

Temporal and Spatial Aggregation

Typically, measures of environmental quality vary over time and space. The dissolved oxygen level at any point in a river rises and falls with changes in streamflow, discharge rates, water temperature and the like, and is different at different locations. Air pollution readings vary with the time of day, the day of the week, over the year, and from one place to another. One problem in empirical research on effects of pollution is how to define a variable or set of variables in a way that adequately reflects the temporal and spatial variations in environmental quality while still being manageable.

To put the problem in a concrete setting, consider an attempt to estimate the relationship between the level of an air pollutant and some health effect, say the occurrence of an asthma attack. The air pollution level at any particular point in an urban area is an instantaneous variable that fluctuates over time. The true exposure of an individual located at that point is measured by a trace of the time path of that instantaneous variable over the relevant time period. However, the individual may also move from one point to another in the urban space. The published data on air pollution levels that are used to generate exposure variables for empirical research involve various approaches to summarizing the instantaneous time paths at the locations of air-pollutant measurement devices. One common measure is the annual mean, either arithmetic or geometric. Averages are also struck over shorter time periods—a 24-hour average for particulates, and 8-hour and 1-hour averages for other pollutants, for example. Also, readings taken at one or two points in the space must be used to represent the space as a whole. Inevitably, summarizing involves loss of information. One or two of these temporal and spatial summaries cannot completely represent the true exposure of any individual. Research on health effects is hampered by our inability to characterize accurately the exposure of individuals.

The Welfare Economics of Costs

So far, this chapter has focused attention on valuing the benefits of environmental changes. It is time now to turn attention to the costs of achieving these changes. It is a commonly held view both within and outside the economics profession that the costs of environmental regulations are relatively easy to measure, at least in comparison with the task of measuring environmental benefits. This optimistic view is consistent with a naive theory of cost, which takes the following form. Firms respond to pollution control regulations by purchasing and installing waste treatment equipment and control systems that, in effect, are bolted on to the existing factory. The purchase and installation costs of this equipment, plus the added operating and maintenance costs it would entail are readily identified in the firm's accounts. These expenditures are often taken, at least as a first approximation, to be the social costs of complying with the regulation.

The naive theory fails to recognize the fundamental symmetry between benefits and costs as changes in the utilities of individuals; and it also neglects several important realities concerning the ways in which government regulations can affect people's welfare. The symmetry of benefits and costs stems from the fact that ultimately all costs take the form of utility losses to individuals in their dual roles as receivers of income and consumers of market and nonmarket goods and services. These losses have their monetary counterparts in compensating measures of welfare change—that is, willingness to pay to avoid the cost and willingness to accept compensation for bearing the cost. Because of this fundamental symmetry, proper measurement of costs involves the same kinds of problems as, and is likely to be as difficult as, the measurement of the benefits of environmental improvement. Once the symmetry is acknowledged, the whole economist's tool kit of revealed preference and stated preference methods of measuring welfare changes becomes available for the cost analyst.

The naive view about cost estimation is also implicitly based on the presumption that only firms cause pollution, and that therefore only firms must incur costs in the process of moving to meet environmental quality objectives, but this is not the case. Regulations, for example, may be placed on household activities, and these regulations might not require any identifiable expenditure of money. Suppose that commuters are required to form car pools as part of a program to meet air quality standards for ozone. Household expenditures on commuting may in fact be lower as a consequence of the regulation. The cost of such regulation takes the form of increased commuting time and loss of convenience. These may be difficult to quantify and value. Consider also a regulation that decreases the quality or stream of services from a consumer good. In the 1970s for example, new cars complying with tailpipe emissions standards had lower performance and fuel efficiency than earlier models. The true cost of complying with these standards consisted of both the higher cost of purchasing and operating the car and the decreased value of the services from the car.

When an environmental regulation affects household activities or the availability of nonmarket goods and services, there may be opportunities for employing revealed preference methods to draw inferences about the negative values or costs of these changes. For example, if the quality of a good is reduced as a consequence of a regulation, hedonic price models might be used to estimate the marginal implicit price of the relevant attribute. In the example of imposed car-pooling, an estimate of the shadow price of time could be used to derive the costs of the increased commuting time. Also, stated preference methods such as contingent valuation or choice experiments might be used where a comprehensive set of regulations affects a wide range of activities.

Another factor neglected in the naive theory of costs is that market mechanisms are likely both to shift the burden of firms' expenditures and to change the magnitude of the burden so that the true costs are not accurately measured by summing the expenditures by firms. At least for nonmarginal changes, firms will raise prices and will experience decreases in quantity demanded. Price increases cause losses of consumer surplus, which are part of the social costs; and there may be losses in producers' surpluses and factor incomes as well. Thus, it is not correct simply to equate pollution control expenditures and social costs (Portney 1981).

As an extreme example, consider the case of a tax on the carbon content of fuels as a means of reducing the emissions of carbon dioxide (CO_2). If there were no economically feasible technologies for controlling CO_2 emissions from combustion, then the tax would work entirely through raising the prices of carbon-based fuels and reducing demands for them. Consumers would bear at least part of the costs of the revenues raised by the tax, and the rest would come from decreases in resource rents. However, since these tax revenues represent a transfer to the government, they are not part of the social cost of reducing CO_2 emissions. The social cost of controlling CO_2 by taxation comes entirely in the form of the equivalent of the deadweight losses or welfare triangles associated with the tax.

Where the direct impact of a regulation is on firms, and where its effects are transmitted to individuals through changes in prices and incomes, the process of cost estimation must call on two types of models. The first is a model of the firm's production technology and costs. The second is a market model that can be used to calculate the changes in prices and incomes of the affected individuals. This model, in some circumstances, could be a partial equilibrium model. For broad social regulations, a general equilibrium model of the economy may be required (Hazilla and Kopp 1990).

In modeling the behavior of the firm, a number of questions have to be considered. For example, can the technology be modeled as additively separable so that the total cost is the sum of the costs of producing the marketed outputs and the cost of treating the waste? Or must the technology and costs be modeled as fully joint? The assumption of additive separability may be appropriate if, in fact, firms would respond to a regulation by purchasing add-on control or treatment equipment. However, other kinds of technological responses, such as input substitution and recycling, may have to be modeled in a joint production framework.

Once a model of a firm's cost has been obtained, the next step is to embed that model in a model of the market economy so that the changes in all relevant quantities, prices, and incomes can be predicted. In the simplest case, where the regulation involves a marginal change in costs and where the economy is perfectly competitive, costs can be measured by the predicted change in expenditures on factor inputs. This is because the invariant factor prices are equal to the values of the marginal products of those inputs in other uses; and, in turn, those values of marginal products measure the opportunity losses to consumers associated with the marginal reallocation of inputs.

With nonmarginal changes, it can be expected that there will also be changes in product prices and perhaps factor prices. If a regulation affects only one industry, a partial equilibrium model may be appropriate. Rather than have costs be a function of q, among other things, a parameter reflecting the stringency of the regulation or the degree of pollution control required could be included as a shifter of the cost or production function.

Implicit in this approach are the assumptions that the regulation does not cause any shifts in either the output demand functions or the factor supply functions, either immediately or over time. However, if either of these assumptions are not valid, then a general equilibrium framework would be required. For example, if the regulation affected both the x and y industries, the regulation-induced change in the price of y could shift the demand function for x. Also, if y were an intermediate product and an input in the x industry, the increase in its price or the upward shift in its supply function would have a secondary impact on the cost of producing x. Hazilla and Kopp (1990) have shown that estimating true social costs in a general equilibrium framework can lead to quite different results in comparison with the pollution control expenditure approach utilized by the Environmental Protection Agency (EPA).

More recent work has called attention to another, perhaps more important, general equilibrium effect. If an environmental regulation leads to an increase in the prices of goods and services, then the result is a fall in real wages; and if labor supply elasticities are positive, a reduction in the quantity of labor supplied occurs. Since the labor market is already distorted because of the presence of income and payroll (social security) taxes, the marginal social value of labor exceeds its marginal social cost by a substantial amount. Even a small decrease in the quantity of labor supplied can have a large net welfare cost. This cost is in addition to the direct cost of the environmental regulation. The impact of regulatory costs on the labor market is known as the tax interaction effect. For a clear explanation of this effect and a discussion of its significance for the economic analysis of environmental and other regulatory policies, see Parry and Oates (2000).

Summary

Economists seek measures of values that are based on the preferences of individuals. When value measures are derived using models of behavior, these models should be internally consistent and be based on accepted theories of preferences, choice, and economic interactions. Equally important is the need for a sound understanding of the underlying biological and physical processes by which environmental and resource service flows are generated. However, if empirical observations of individuals' choices are taken without benefit of an underlying theoretical model, researchers may be led to make faulty or erroneous interpretations of the data. An interesting example is the early studies of the land value / air pollution relationship. Researchers discovered that land values and air pollution levels were inversely related in urban areas, other things being equal. They then assumed that changes in welfare associated with reduced pollution would be accurately measured by the associated increases in land values as predicted by the regression equation relating land values at a point in time to air pollution. Subsequent research based on theoretical models of urban land markets has shown that this assumption is not true in general. The relationships among air pollution, land values, and measures of welfare change are discussed in Chapter 10.

References

Baumol, William J., and Wallace E. Oates. 1988. *The Theory of Environmental Policy*. 2d ed. Cambridge: Cambridge University Press.

Bockstael, Nancy E., and Kenneth E. McConnell. 2007. *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models*. Dordrecht: Springer.

Freeman, A. Myrick III. 1995. "The Benefits of Water Quality Improvements for Marine Recreation: A Review of the Empirical Evidence." *Marine Resource Economics* 10(4): 385–406.

Hazilla, Michael, and Raymond J. Kopp. 1990. "The Social Cost of Environmental Quality Regulations: A General Equilibrium Analysis." *Journal of Political Economy* 98(4): 853–873.

Holmes, Thomas P., and Wictor L. Adamowicz. 2003. "Attribute-Based Methods." In *A Primer on Nonmarket Valuation*, edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Dordrecht: Kluwer Academic Publishers.

Kneese, Allen V. 1968. "Economics and the Quality of the Environment: Some Empirical Experiences." In *Social Sciences and the Environment*, edited by Morris Garnsey, and James Gibbs. Boulder: University of Colorado Press.

Kuminoff, Nicholai V., V. Kerry Smith, and Christopher Timmins. 2013. "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets." *Journal of Economic Literature* 51(4): 1007–10562.

McConnell, Kenneth E. 1986. "The Damages to Recreational Activities from PCBs in New Bedford Harbor." Report prepared by Industrial Economics, Inc., Cambridge, MA, for the Ocean Assessment Division, National Oceanic and Atmospheric Administration.

Mitchell, Robert C., and Richard T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, D.C.: Resources for the Future.

Parry, Ian W.H., and Wallace E. Oates. 2000. "Policy Analysis in the Presence of Distorting Taxes." *Journal of Policy Analysis and Management* 19(4): 603–613.

Phaneuf, Daniel J., Jared C. Carbone, and Joseph A. Herriges. 2009. "Non-Price Equilibria for Non-Marketed Goods." *Journal of Environmental Economics and Management* 57(1): 45–64.

Portney, Paul R. 1981. "The Macroeconomic Impacts of Federal Environmental Regulation." In *Environmental Regulation and the U.S. Economy*, edited by Henry M. Peskin, Paul R. Portney, and Allen V. Kneese. Washington, D.C.: Resources for the Future.

U.S. Environmental Protection Agency. 1997. *The Benefits and Costs of the Clean Air Act – 1970 to 1990*. Washington, D.C.: U.S. EPA.

Chapter 3

Welfare Measures

Definitions and Concepts

The theory of the measurement of welfare change has been discussed by others, both at the most rigorous levels of abstraction, and in pragmatic, practical terms of application. See for examples, Johansson (1987), Just, Hueth, and Schmitz (1982, 2004), and Bockstael and McConnell (2007). The earliest work focused on the welfare effects of changes in the prices people pay for the (private) goods they consume, but the literature has expanded broadly into valuing changes in the quantity and quality of both private and public goods. The current chapter provides a systematic development of the definition and measurement of the welfare effects stemming from changes in prices and the quantities and/or qualities of nonmarket environmental and resource service flows.

Changes in environmental quality can affect individuals' welfares through a number of channels: changes in the prices paid for goods bought in markets; changes in the quantities or qualities of nonmarketed goods (for example, public goods such as air quality); changes in the prices received for factors of production; and changes in the risks individuals face. The first two of these channels are the focus of this chapter. After a brief review of the theory of individual preferences and demand, the principles of welfare measurement for price changes are reviewed. These principles are relevant because some forms of environmental change affect people only indirectly through price effects, and because these principles provide a solid foundation for the treatment of quantity and quality changes that follow. Chapter 8 covers the welfare effects of changes in factor prices. The extension of these principles to the valuation of changes in risk—the fourth channel—raises some interesting questions, which will be left to Chapter 5.

The principles and measures developed in this chapter apply equally to decreases and increases in individuals' welfare. It is a basic principle of welfare economics that all costs ultimately take the form of reductions in the utility of individuals. This principle applies equally to the costs of public policies (for example, investment in resource development and the regulation of private activities), and to the costs of private uses of the environment (for example, harvesting from a common property resource and using the waste receptor services of the environment). Hence, the welfare measures developed here provide a foundation for the analysis of both the benefits and the costs of environmental change.

In this chapter, three sets of questions are considered in some detail. The first set concerns how to define an acceptable monetary measure of changes in economic welfare for an individual. The answer to this question hinges partially on what the measure would be used for—that is, a welfare measure should answer the questions posed by policymakers. However, policymakers can ask different kinds of questions. For example, suppose that policymakers wish to evaluate proposed policy changes in terms of an aggregate social welfare function that places different weights on individuals' changes in utility depending on their positions in the income distribution (Bergson 1966). In that case, the welfare measure that answers the policymakers' question must be a money metric of utility changes. Alternatively, if policymakers wish to select policies on the basis of the potential Pareto improvement criterion, they will want measures of required compensation and willingness to make compensating payments. The concluding section of this chapter returns to the question of choosing from among the alternative measures described here.

The second set of questions concerns how changes in welfare would be measured, both in theory and in practice. Theory suggests several alternative ways of calculating either exact or approximate welfare measures using data on observed behavior of individuals—for example, their demand functions for market goods. These alternatives will be described and evaluated, especially from the practical perspective of implementation.

The third set of questions concerns how any measure of welfare changes for individuals might be used to make judgments about social policies affecting many individuals. For example, is it possible to speak of a measure of aggregate welfare for the society as a whole? If so, what significance can be attached to changes in such a measure? Measures of welfare change for an individual can be defined and analyzed without reference to the notions of efficiency and equity. In this sense, the concept is objective—that is, one can define and measure a monetary equivalent of an individual's welfare change without being committed to any particular set of value judgments concerning aggregation across individuals, or the role of such welfare measures in social choice. It is in answering the third set of questions that value judgments about the relative deservingness of individuals, the meaning of efficiency, and the objectives of public policy come into play. Some of these issues are discussed in the Aggregation and Social Welfare section of this chapter.

The following section begins with a review of some of the basic terminology and theory involving individual preferences and demand.[1] Next, the standard case where utility depends only on the consumption of market goods is considered. In this simple context, the theory of measuring the welfare value of changes in

---

1 Throughout most of this text, the presentation is based upon neoclassical theory and the assumption of a rational consumer. However, in recent years the assumption of a rational consumer has been drawn into question by research into behavioral economics (e.g., Camerer, Loewenstein and Rabin 2004; Mullainathan and Thaler 2001; Shogren and Taylor 2008). Discussion of the implications of this line of research for welfare economics is left to Chapter 13.

the prices of these goods is examined, along with the relationships among the Marshallian and Hicksian surplus measures of welfare change for continuous goods (that is, goods for which the main consumer choice problem is to choose how many units to purchase and consume). There are two reasons for choosing this order of presentation. First, it parallels the historical evolution of the theory of welfare change. Second, it makes for an easier exposition of the basic principles. This section concludes with a review of methods for obtaining exact measures of, and approximations to, the desired Hicksian surpluses. The third section examines the case of the welfare effects of changes in the quantities of continuous goods. While potentially applicable to private goods, these welfare measures will most often be relevant for public goods or nonmarket goods since the quantity available for consumption is not a matter of choice for the individual. The next section considers welfare measures for discrete goods: goods for which the key choice is not how many units to consume, but rather which goods, from two or more alternatives, are consumed. In the final section, a review is provided of some of the issues involved in aggregating measures of individual welfare change for public policy decision making and in selecting the appropriate welfare measure.

Individual Preferences and Demand

Before introducing the various possible welfare measures, it will be useful to review briefly the basic theory of individual preferences and the demand for goods as it relates to welfare theory. For alternative treatments of this and related topics, the reader may wish to consult other texts, such as Just, Hueth, and Schmitz (1982, 2004), Boadway and Bruce (1984), Varian (1992), and Johansson (1987). This theory starts with the premise that individuals are their own best judges of their welfares and that inferences about welfare can be drawn for each individual by observing that individual's choices among alternative bundles of goods and services. If an individual prefers bundle *A* to bundle *B*, then bundle *A* must convey a higher level of welfare.

What things are to be included in the bundles (such as *A* and *B*) among which individuals are assumed to have preferences? There is little controversy over the inclusion of all the goods and services that can be bought or sold in markets—consumer goods, the services of household assets such as a house or a car, and consumer durables. Since time can be used in leisure activities, or sold at some wage rate in the labor market, individuals must also have preferences among alternative uses of time, such as reading, outdoor recreation, and working at some wage rate. Since government and the environment both provide a variety of services that enhance the welfares of individuals, these services should also be included in the bundles among which people have preferences. Environmental services include those provided by cleaner air, cleaner water, and scenic amenities. Just as importantly, these environmental services are not limited to direct uses of the environment, such as breathing clean air or observing unspoiled vistas—they can also include services related to the mere presence of environmental goods, such as the knowledge that pristine wilderness areas exist, or that there is a viable breeding population of a particular endangered species.

If we assume that individuals can rank the alternative bundles according to their preferences, what properties will the resulting ordering of bundles display? For our purposes, two properties are important. The first is nonsatiation, or the "more-is-better" property. This means that a bundle with a larger quantity of an element will be preferred to a bundle with a smaller quantity of that element, other things being equal. Formally, if $\boldsymbol{X'}$ consists of $\left(x_1',\ldots,x_j',\ldots,x_J'\right)$ and $\boldsymbol{X''}$ consists of $\left(x_1',\ldots,x_j'',\ldots,x_J'\right)$ and $x_j' > x_j''$ then this individual will prefer $\boldsymbol{X'}$ to $\boldsymbol{X''}$.

The second property is *substitutability* among the components of bundles. This means that if the quantity of one element of a bundle, say x_j, is decreased, it is possible to increase the quantity of another element, say x_k, sufficiently to make the individual indifferent between the two bundles. More formally, suppose that $\boldsymbol{X'}$ consists of $\left(x_1',\ldots,x_j',\ldots,x_k',\ldots,x_J'\right)$; and $\boldsymbol{X''}$ consists of $\left(x_1',\ldots,x_j'',\ldots,x_k',\ldots,x_J'\right)$ with $x_j'' < x_j'$. Substitutability means that there is another bundle $\boldsymbol{X^*}$ consisting of $\left(x_1',\ldots,x_j'',\ldots,x_k^*,\ldots,x_J'\right)$ with $x_k^* > x_{kJ}'$, such that the individual is indifferent as to $\boldsymbol{X'}$ and $\boldsymbol{X^*}$. In other words, $\boldsymbol{X'}$ and $\boldsymbol{X^*}$ lie on the same indifference surface.[2]

The property of substitutability is at the core of the economist's concept of value. This is because substitutability establishes tradeoff ratios between pairs of goods that matter to people. In this formulation, the tradeoff ratio is $\left(x_k^* - x_k'\right)/\left(x_j' - x_j''\right)$ or $|\Delta x_k / \Delta x_j|$. In the limit for infinitesimally small changes, this reduces to $|dx_k / dx_j|$, which is the definition of the marginal rate of substitution between x_j and x_k, or the slope of the two-dimensional indifference curve between these two elements. The money price of a market good is just a special case of a tradeoff ratio, because the money given up to purchase one unit of one element of the bundle is a proxy for the quantities of one or more of the other elements in the bundle that had to be reduced in order to make the purchase.

If the preference ordering has the properties described here, it can be represented by an ordinal preference function, or utility function, that assigns a number to each bundle as a function of the quantities of each element of the bundle. Specifically,

$$u = u(\boldsymbol{X}, \boldsymbol{Q}, \boldsymbol{T}), \tag{3.1}$$

where $\boldsymbol{X}$ is a vector of the quantities of market goods, $\boldsymbol{Q}$ is a vector of public goods and environmental and resource services whose quantities or qualities are fixed for the individual, and $\boldsymbol{T}$ is a vector of the times spent in various activities that yield utility to the individual. This utility function is assumed to be increasing in

2 Two other important properties are transitivity and quasi-concavity. If there are three bundles $\boldsymbol{X'}$, $\boldsymbol{X''}$, and $\boldsymbol{X^*}$, and the individual prefers $\boldsymbol{X'}$ over $\boldsymbol{X''}$ and $\boldsymbol{X''}$ over $\boldsymbol{X^*}$, then transitivity is satisfied if the individual prefers $\boldsymbol{X'}$ over $\boldsymbol{X^*}$. For more on the axiomatic description of these properties of preference ordering, see Boadway and Bruce (1984) or Varian (1992).

all of its arguments, and unique up to a monotonic transformation. For purposes of mathematical modeling and analysis, it is convenient also to assume that this function is continuous, convex, and twice differentiable. This preference function is not the same thing as the cardinal utility function of the classical utilitarians. Since there is no unit of measurement for this ordinal utility, it is not possible to add or otherwise compare the utilities of different individuals.

To simplify the exposition and notation, let us now consider an individual whose utility is a function only of private goods that can be bought and sold in markets. Assume that tastes and preferences (that is, the utility function) are given and do not change. The individual faces a set of given prices for these goods and is assumed to choose the quantities of the goods so as to maximize his utility, given the constraints of prices and a fixed money income M. The maximization problem can be expressed as

maximize $u = u(\boldsymbol{X})$,

subject to

$$\sum_{j=1}^{J} p_j x_j = M \tag{3.2}$$

where $\boldsymbol{X} = \left(x_1, \ldots, x_j, \ldots x_J\right)$ is the vector of quantities. The solution to this problem leads to a set of ordinary, or Marshallian, demand functions

$$x_j = x_j(\boldsymbol{P}, M), \tag{3.3}$$

where $\boldsymbol{P} = \left(p_1, \ldots, p_j, \ldots p_J\right)$ is the vector of prices.

Substituting the expressions for x_j as functions of $\boldsymbol{P}$ and M into the direct utility function gives the indirect utility function—that is, utility as a function of prices and income, assuming optimal choices of goods:

$$v = v(\boldsymbol{P}, M). \tag{3.4}$$

According to Roy's identity, the demand functions can also be expressed in terms of derivatives of the indirect utility function,

$$x_j\left(\boldsymbol{P}, M\right) = -\frac{\partial v / \partial p_j}{\partial v / \partial M} \cdot \tag{3.5}$$

The expenditure function represents a useful perspective on the problem of individual choice. The expenditure function is derived by formulating the dual of the utility maximization problem. The individual is assumed to minimize total expenditure,

$$e = \sum_{j=1}^{J} p_j x_j \ , \tag{3.6}$$

subject to a constraint on the level of utility attained,

$$u\left(\boldsymbol{X}\right) = u^0 \ , \tag{3.7}$$

where u^0 is the maximum utility attained with the solution to the primal problem. Just as the solution to the utility maximization problem yields a set of ordinary demand curves conditional on prices and money income, the solution of the expenditure minimization problem yields a set of functions giving optimal quantities for given prices and utility. These are Hicks-compensated demand functions that show the quantities consumed at various prices, assuming that income is adjusted (compensated), so that utility is held constant at u^0. Substituting these demand functions into the expression for total expenditure yields the expenditure function. This expression gives the minimum dollar expenditure necessary to achieve a specified utility level given market prices. In functional notation:

$$e = e(\boldsymbol{P}, u^0), \tag{3.8}$$

where e is the dollar expenditure and u^0 is the specified utility level. The compensated demand functions can also be found by differentiating the expenditure function with respect to each of the prices:

$$\frac{\partial e}{\partial p_j} = h_j = h_j\left(\boldsymbol{P}, u^0\right), \tag{3.9}$$

where h_j is the compensated demand for x_j.

Now consider the set of ordinary demand functions derived from the utility maximization problem. In order to determine the functional form and parameters of these demand functions, it is necessary to know the underlying utility function, and this may not be directly observable. Suppose instead that we observe an individual's behavior and estimate the demand functions that describe the individual's responses to changes in prices and income. These functions are based on the same information as the underlying preferences. This is assured, provided the demand functions satisfy the so-called *integrability conditions*. These conditions require that the Slutsky matrix of substitution terms,

$$\frac{\partial h\left[\boldsymbol{P}, v\left(\boldsymbol{P}, M\right)\right]}{\partial p_k} = \frac{\partial x_j\left(\boldsymbol{P}, M\right)}{\partial p_k} + \frac{\partial x_j\left(\boldsymbol{P}, M\right)}{\partial M} x_k\left(\boldsymbol{P}, M\right), \tag{3.10}$$

is symmetric and negative semi-definite (Hurwicz and Uzawa 1971; Silberberg 1978; Varian 1992). If these conditions are satisfied, the system of demand functions can be integrated to yield the expenditure function, which in turn can be used to derive the indirect and direct utility functions. If the integrability conditions are not satisfied, the implication is that the observed demand functions are not consistent with the maximization of a well-behaved utility function. As explained below, if the integrability conditions are satisfied, it may be possible to utilize empirically derived descriptions of demand behavior to obtain a complete description of the underlying preferences, as well as exact measures of welfare change for a wide range of postulated changes in economic circumstances.

Welfare Measures for Continuous Goods: Price Changes

An Overview

In order to introduce the alternative welfare measures, consider first the simplest case of only two goods and the welfare gain associated with a nonmarginal decrease in the price of one of these goods. Two types of measures of this welfare change have been identified in the literature. The first is the change in ordinary consumer's surplus, a concept with an origin that can be traced back through Alfred Marshall to Dupuit. Mishan (1960) and Currie, Murphy, and Schmitz (1971) provided useful discussions of the history and evolution of the concept of consumer's surplus. As Marshall explained it,

> [The individual] derives from a purchase a surplus of satisfaction. The excess of the price that he would be willing to pay rather than go without the thing, over that which he actually does pay, is the economic measure of this surplus of satisfaction. It may be called consumer's surplus.
>
> (Marshall 1920, 124)

Ordinary consumer's surplus is measured by the area under a Marshallian ordinary demand curve, but above the horizontal price line. As we will see, the

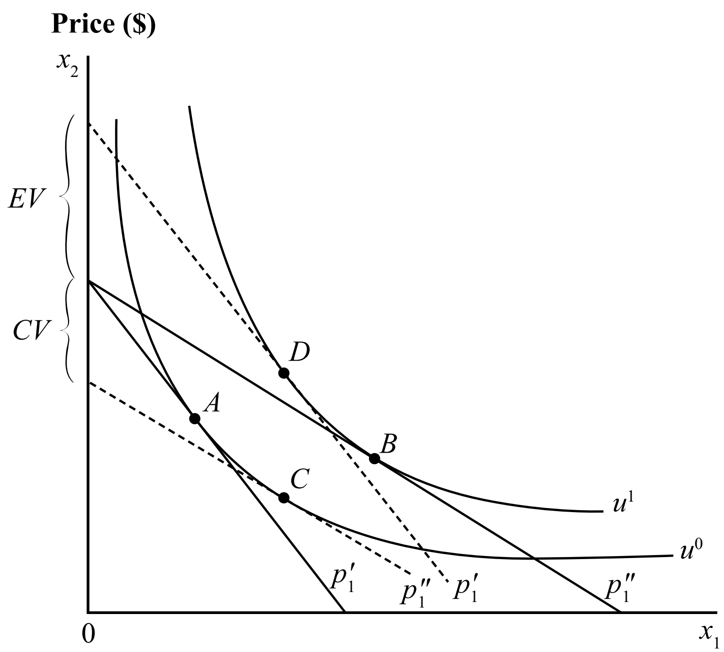

Figure 3.1 Two measures of the welfare gain from a price decrease

consumer surplus measure cannot be defined in terms of the underlying utility function.

The other measures of welfare change are theoretical refinements of the ordinary consumer's surplus (Hicks 1943), and each can be defined in terms of the underlying individual preference mapping. Figure 3.1 is used to illustrate these concepts in the context of two goods (x_1 and x_2), where x_2 is the numeraire good (i.e., $p_2 = 1$). The figure shows two indifference curves for the individual. Assume that an environmental improvement reduces the cost of producing x_1, so that its price drops from p_1' to p_1''. In response to the price reduction, the individual shifts from the consumption bundle marked *A* at utility level u^0 to consumption bundle *B* at utility level u^1. What is the welfare benefit of the price reduction to this individual? Two additional measures of the welfare change can be defined in terms of the numeraire good x_2:

1 Compensating Variation (*CV*). This measure asks what compensating payment (that is, an offsetting change in income) is necessary to make the individual indifferent between the original situation (*A* in Figure 3.1) and the new price set. Given the new price set with consumption point *B*, the individual's income could be reduced by the amount of *CV* and that person would still be as well off at point *C* as at point *A* with the original price set and money income. The measure *CV* is often interpreted as the maximum amount that the individual would be willing to pay for the opportunity to consume at the new price set. However, for a price increase, *CV* measures what must be paid to the individual to make that person indifferent to the price change. For price decreases, the *CV* cannot be greater than the individual's income; but for a price increase, the *CV* could exceed income.
2 Equivalent Variation (*EV*). This measure asks what change in income (given the original prices) would lead to the same utility change as the change in the price of x_1. As shown in Figure 3.1, given the original prices, the individual could reach utility level u^1 at point *D* with an income increase equal to *EV*. *EV* is the income change equivalent to the welfare gain due to the price change. The *EV* measure has also been described as the minimum lump sum payment the individual would have to receive to induce that person to voluntarily forgo the opportunity to purchase at the new price set. For a price increase, *EV* is the maximum amount the individual would be willing to pay to avoid the change in prices.

Note that both the *EV* and *CV* measures allow the individual to adjust the quantities consumed of both goods in response to both changes in relative prices and income levels. Hicks also described two additional measures where the levels of the goods could not be changed. He referred to them as compensating and equivalent surplus. The compensating and equivalent surplus measures for price changes do not answer very useful questions since they both arbitrarily restrict the individual to consuming a specific quantity of the good whose price has changed.

Hence, the original form suggested by Hicks will not be considered further, but we will return to measures of welfare change associated with quantity changes (with no associated price changes) later in this chapter; and note that, following Hicks, such welfare measures are often referred to as compensating and equivalent surplus. The next subsection is devoted to a comparison and evaluation of the compensating and equivalent variations and their relationship to the ordinary consumer surplus.

In the many-good case, x_2 is a composite good that can be treated as an index of the consumption levels of all other goods except x_1. The aggregation of all other goods into a composite good for graphical representation is valid so long as the prices of all of the goods are assumed to move in the same proportion—that is, there are no changes in the relative prices of components of the composite good bundle. This assumption can be maintained, since we are analyzing only the consequences of the change in the price of x_1.

A Closer Look at the Welfare Measures

This section begins with a presentation of the basic welfare measure for a marginal change in one price. Then more rigorous derivations are provided for the consumer surplus, compensating variation, and equivalent variation measures of welfare change for the case of changes in price. For more detailed treatment of these topics see Silberberg (1972), Just, Hueth, and Schmitz (1982, Appendix B), Varian (1992), and Johansson (1987). For a marginal change in, say, p_1, the basic welfare measure is the change in expenditure necessary to hold utility constant. Using equation (3.9) from the previous section, we have

$$w_{p_1} = \frac{\partial e\left(\boldsymbol{P}, u^0\right)}{\partial p_1} = h_1\left(\boldsymbol{P}, u^0\right), \tag{3.11}$$

where w_{p_1}, is the marginal welfare measure. This result also follows from the indirect utility function and Roy's identity:

$$w_{p_1} = x_1 = -\frac{\partial v / \partial p_1}{\partial v / \partial M} \tag{3.12}$$

or

$$\frac{dM}{dp_1} = x_1 . \tag{3.12′}$$

In equation (3.12), the marginal utility of the price change is converted to monetary units by dividing by the marginal utility of income. Equation (3.12′) says that the change in income required to hold utility constant is equal to the change in price multiplied by the quantity of the good being purchased.

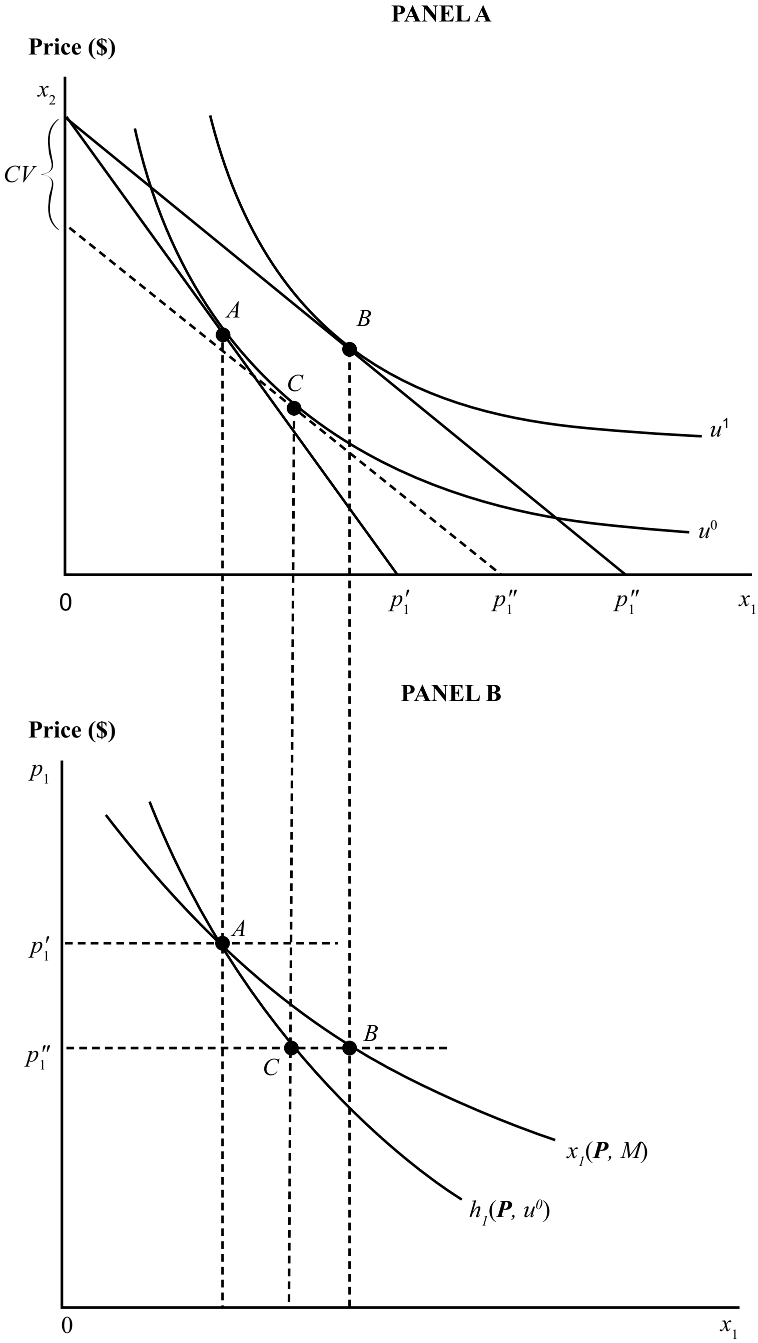

Figure 3.2 The compensating variation and Hicks-compensated demand

Marshallian Consumer Surplus

In Figure 3.2, panel A shows one individual's preference mapping in the simple two-good case. Suppose that the price of good x_1 falls from p_1' to p_1''. The individual responds by moving from the original equilibrium at point A to point B on the new budget line. In panel B of Figure 3.2, these equilibrium positions are plotted in the price and quantity plane. Points A and B are on the ordinary demand curve, holding the price of good x_2 and money income constant. Since the Marshallian surplus associated with the consumption of a good at a given price is the area under the demand curve, the change in surplus for a change in the good's price is the geometric area $p_1'ABp_1''$ in panel B of Figure 3.2. In mathematical form,

$$S = \int_{p_1''}^{p_1'} x_1(\boldsymbol{P}, M)\, dp_1 , \tag{3.13}$$

where S is the change in surplus.

The condition under which S can be interpreted as an indicator of utility change can be seen by employing Roy's identity:

$$x_1(\boldsymbol{P}, M) = -\frac{\partial v(\boldsymbol{P}, M)/\partial p_1}{\partial v(\boldsymbol{P}, M)/\partial M} , \tag{3.14}$$

and substituting this into equation (3.13) to obtain

$$S = -\int_{p_1''}^{p_1'} \left[\frac{\partial v(\boldsymbol{P}, M)/\partial p_1}{\partial v(\boldsymbol{P}, M)/\partial M} \right] dp_1 . \tag{3.15}$$

If the marginal utility of income is constant over the range of the price change, this can be written as

$$S = \frac{v(p_1'', p_2, M) - v(p_1', p_2, M)}{\partial v/\partial M} . \tag{3.16}$$

This expression shows that the Marshallian surplus can be interpreted as the utility change converted to monetary units by a weighting factor—the marginal utility of income. If the marginal utility of income is constant, then S can be said to be proportional to the change in utility for any price change. However, as any one price changes, the constancy of the marginal utility of income is a restrictive condition. The marginal utility of income cannot simultaneously be invariant with respect to income and to changes in all of the prices (Samuelson 1942; Johansson 1987, ch. 4).

Alternatively, as Eugene Silberberg explained (1978, 350–361), the integral of equation (3.15) can be viewed as the sum of a series of small steps from an initial price and income vector of (p_1', p_2', M) to (p_1'', p_2', M), following a path on which p_2 and M are held constant. However, there are other paths over which (3.15) can be integrated involving changes away from the initial values for p_2 and/or M as long as the terminal point is $\left(p_1'', p_2', M\right)$. The other paths, in general, will not lead to the same solution value for the integral. In other words, the integral in general will not be path independent.

A similar problem arises when the Marshallian surplus measure is generalized to simultaneous changes in all prices. In this case, S is defined as a line integral. This integral will be independent of the path of integration (that is, the order in which prices and/or incomes are assumed to change) only if the income elasticities of demand for all goods are equal. The income elasticities of all goods can be equal to each other only if they are all equal to one, in other words, if preferences are homothetic. Finally, if the prices of only a subset of all goods change, a unique S exists if the marginal utility of income is constant with respect to only those prices that are changed. See Just, Hueth, and Schmitz (1982) for more details.

Compensating Variation

Suppose now that as the price of good x_1 is decreased, income is taken away from the individual so that he remains at the initial utility level and indifference curve u^0. Given the price change and the compensating income change, the individual would be in equilibrium at point C in panel A of Figure 3.2. Point C is also plotted in panel B of Figure 3.2. Points A and C are on the Hicks-compensated demand curve, a demand curve that reflects only the substitution effect of the change in relative prices. The device of compensating withdrawals of money income has eliminated the income effect of the price change. Since x_1 is a normal good by assumption—that is, it has an income elasticity greater than zero—the Hicks-compensated demand curve is less price-elastic than the ordinary demand curve. The difference between the Hicks-compensated and the ordinary demand functions is one of the main considerations in the comparison of EV, CV, and consumer surplus measures of welfare change.

Panel A of Figure 3.2 shows the compensating variation measure of the welfare change associated with the price decrease—that is, the reduction in income needed to hold the individual on the original indifference curve. In terms of the indirect utility function, CV is the solution to

$$v\left(\boldsymbol{P}', M\right) = v\left(\boldsymbol{P}'', M - CV\right) = u^0. \tag{3.17}$$

The CV can also be defined in terms of the expenditure function. It is the difference between the expenditures required to sustain utility level u^0, at the two price sets:

$$\begin{aligned} CV &= e\left(p_1', p_2, u^0\right) - e\left(p_1'', p_2, u^0\right) \\ &= M - e\left(p_1'', p_2, u^0\right) > 0. \end{aligned} \tag{3.18}$$

Because CV is defined as the difference between two levels of expenditure, it can also be written as the integral of the marginal welfare measure (equation 3.11) over the relevant range. Specifically,

$$CV = \int_{p_1''}^{p_1'} \frac{\partial e\left(\boldsymbol{P}, u^0\right)}{\partial p_1} dp_1 = \int_{p_1''}^{p_1'} h_1\left(\boldsymbol{P}, u^0\right) dp_1 . \tag{3.19}$$

Since spending M at the new price set yields a higher level of utility, we can also write

$$M = e\left(p_1'', p_2, u^1\right), \tag{3.20}$$

and by substitution

$$CV = e\left(p_1'', p_2, u^1\right) - e\left(p_1'', p_2, u^0\right) > 0 . \tag{3.21}$$

In other words, although the CV is defined in terms of u^0, it also measures the amount of money required to raise utility from u^0 to u^1 at the new set of prices.

The CV is equal to the area to the left of the Hicks-compensated demand curve between the two prices—that is, the area $p_1'ACp_1''$. The partial derivative of the expenditure function with respect to p_1 gives the change in expenditure (income) necessary to keep the individual on u^0 for small changes in p_1. As shown above, this derivative gives the Hicks-compensated demand curve—that is, it gives the optimal quantity for x_1, holding utility constant. For finite changes, the integral of this derivative is the area to the left of the Hicks-compensated demand curve—that is, the CV. In other words,

$$CV = \int_{p_1''}^{p_1'} h_1\left(\boldsymbol{P}, u^0\right) dp_1 . \tag{3.22}$$

Unlike the Marshallian measure of surplus given by equation (3.13), this measure does not rely on any assumption about the constancy of the marginal utility of income. This is because this measure integrates along a constant utility indifference curve at u^0. In the many-good case, when several prices change, the CV of the price changes taken together is the integral of the set of compensated demand functions evaluated by taking each price change successively. The order in which the price changes are evaluated is irrelevant. This follows from the symmetry of the cross price substitution terms—that is, $\partial x_j/\partial p_k = \partial x_k/\partial p_j$.

Equivalent Variation

The equivalent variation can also be derived through the expenditure function. Panel A of Figure 3.3 shows the same preference mapping and price change for an individual. With a price decrease, the EV is defined as the additional expenditure (income) necessary to reach utility level u^1, given the initial set of prices. In terms of the indirect utility function, EV is the solution to

$$v\left(\boldsymbol{P}', M+EV\right) = v\left(\boldsymbol{P}'', M\right) = u^1 . \tag{3.23}$$

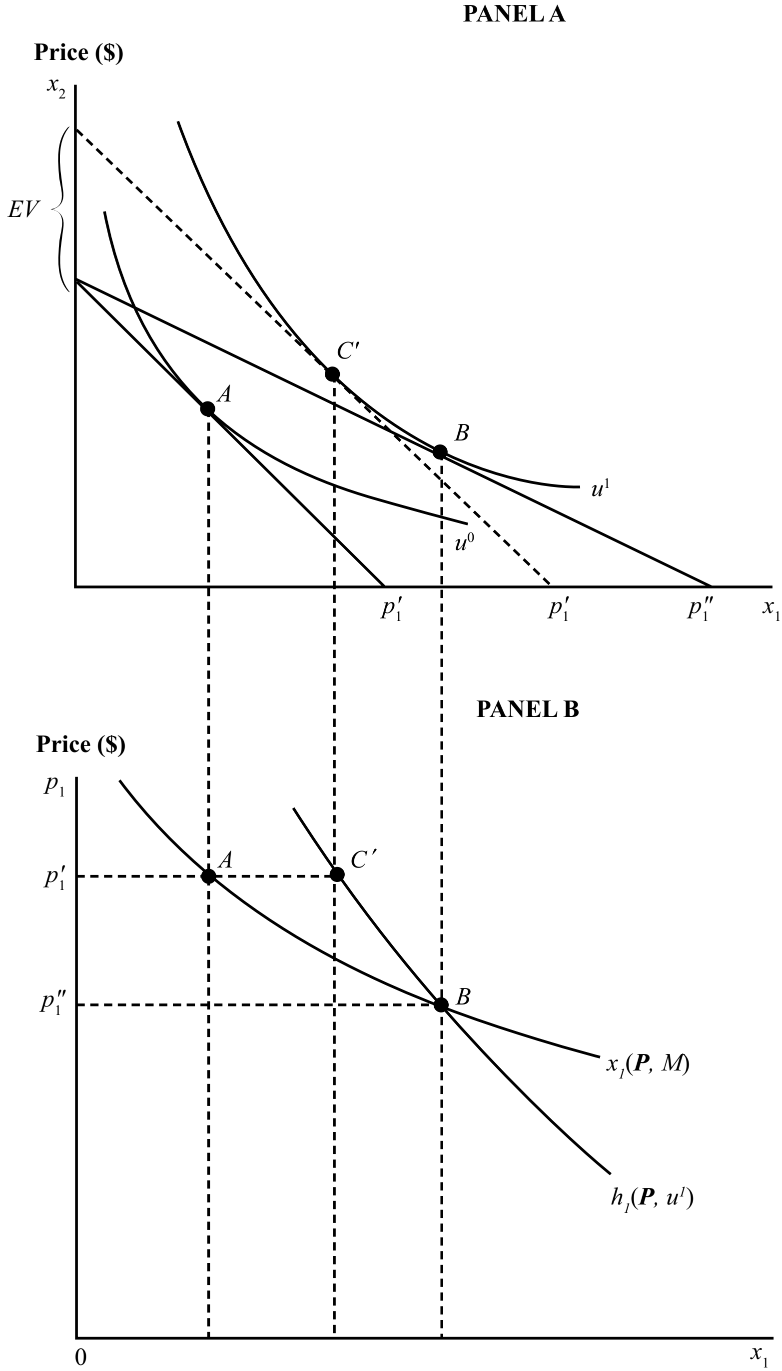

Figure 3.3 The equivalent variation and the Hicks-compensated demand

In Figure 3.3, the EV is the additional expenditure necessary to sustain point C' over point A at the initial prices, or

$$\begin{aligned} EV &= e\left(p_1', p_2, u^1\right) - e\left(p_1', p_2, u^0\right) \\ &= e\left(p_1', p_2, u^1\right) - M > 0. \end{aligned} \tag{3.24}$$

Since the money expenditure levels are the same at point A and point B—that is, $e\left(p_1', p_2, u^0\right) = e\left(p_1'', p_2, u^1\right)$—this can also be written as

$$EV = e\left(p_1', p_2, u^1\right) - e\left(p_1'', p_2, u^1\right). \tag{3.25}$$

In other words, although the EV is defined in terms of the monetary equivalent of a change from u^0 to u^1, it can also be measured by the change in expenditure associated with price changes given utility level u^1.

The EV can also be written as the integral of the marginal value measure (equation 3.11):

$$EV = \int_{p_1''}^{p_1'} \frac{\partial e\left(\boldsymbol{P}, u^1\right)}{\partial p_1} dp_1 . \tag{3.26}$$

The price derivative of the expenditure function (this time holding utility constant at u^1) generates another Hicks-compensated demand curve through point B in panel B of Figure 3.3. The area to the left of this Hicks-compensated demand curve between the two prices (area $p_1'C'Bp_1''$) is the equivalent variation welfare measure. In other words,

$$EV = \int_{p_1''}^{p_1'} h_1\left(\boldsymbol{P}, u^1\right) dp_1 . \tag{3.27}$$

As in the case of the CV, this measure does not require any assumption about the constancy of the marginal utility of income; and the measure for multiple price changes is path independent.

All of this discussion has been in terms of the welfare gain due to a price decrease. The derivation of the welfare cost of a price increase can be worked out in a symmetrical fashion. In general, for any price change, the CV welfare measure is the area to the left of the Hicks-compensated demand curve that passes through the initial position. The EV measure of the welfare change is the area to the left of the Hicks-compensated demand curve that passes through the final position.

A Comparison of the Three Measures

Although the Marshallian consumer surplus has some intuitive appeal as a welfare indicator, it does not measure either of the theoretical definitions of welfare change developed here. In general, it is not a measure of gain or loss that can be employed in a potential compensation test. The Marshallian surplus does lie between the CV and the EV, however, this opens the question of whether it can be a useful approximation to either of these other measures, a question that is taken up below in the subsection Consumer's Surplus Without Apology.

In contrast, the CV and the EV do represent welfare relevant measures. The EV is the monetary equivalent of a price change. It can be interpreted as an index of utility in the sense that it imputes the same monetary value to all changes from an initial position that result in the same final utility level. This is an ordinal utility index (Morey 1984). For example, suppose a change from initial position A to position B has an EV of \$10, while a change from A to C has an EV of \$20. It cannot be inferred that the second change conveys twice as much extra utility as the first change. This is because it evaluates all changes from an initial position at the same set of prices. The CV cannot be interpreted as an index of utility—rather, it measures the offsetting income change necessary to "prevent" a utility change. As Silberberg put it, "the [EV] imputes a dollar evaluation to a change in utility levels for a particular path of price changes, while the [CV] derives dollar values necessary to hold utility constant when prices change" (1972, 948).

The two measures EV and CV will be the same if the income elasticity of demand for good x_1 is zero. In this case, the ordinary and Hicks-compensated demand curves are identical. With positive income elasticity, the EV exceeds the CV for price decreases, but the CV exceeds the EV when price increases are considered. The difference between points C and B in Figure 3.2, and between points A and C' in Figure 3.3, is one of income level. If the income elasticity of demand for x_1 were zero, the income differences would have no effect on the purchase of x_1. The CV and the EV would be exactly equal, and they both could be measured by the area under the ordinary demand curve. The higher the income elasticity of demand for x_1, the larger the difference between the EV and the CV, and the larger the difference between either of the measures and the ordinary consumer surplus.

There is symmetry between the CV and the EV measures that can be seen by comparing Figures 3.2 and 3.3, and by comparing equation (3.21) with equation (3.24), and by comparing equation (3.18) with equation (3.25). For simplicity, let I represent the initial price set (with p_1') and let II represent the second price set (with p_1''). The CV for moving from I to II with u^0 as the reference utility level is exactly equal to the EV of moving from II to I with u^1 as the reference utility level. The CV is a welfare measure for the move from A to B via point C; the EV starts at point B and measures the reduction in income necessary to get to point A, and therefore u^0 via point C'. Similarly, the EV for the move from I to II is just equal to the CV starting at II and u^1, and moving to I.

This symmetry relates to the interpretation of CV and EV as measures of willingness to pay (WTP) and willingness to accept (WTA) compensation. The CV is sometimes described as the maximum willingness to pay for the right to purchase the good at the new price level (i.e., the lump sum payment that the individual would be willing to make that would just exhaust the potential for welfare gain from the new price). This description is accurate only for a price decrease. For a price increase, the CV defines the minimum payment to the individual sufficient to prevent a utility decrease; in other words, it defines a WTA measure. Similarly, the EV defines a WTA measure for a price decrease—that is, the sum of money the individual would require to voluntarily forgo a proposed price decrease. However,

Table 3.1 The implied rights and obligations associated with alternative price sets

Welfare measure	*Price increase*	*Price decrease*
EV – Implied property right in the change	WTP to avoid	WTA to forgo
CV – Implied property right in the status quo	WTA to accept	WTP to obtain

for a proposed price increase, the *EV* is a WTP measure—that is, the maximum sum of money that could be taken away from the individual—yielding a loss of utility equivalent to that caused by the price change. Whatever the direction of the price change, the *CV* takes the initial utility as the reference point.

These two measures can also be interpreted in terms of the implied rights and obligations associated with alternative price sets. The *CV* carries an implicit presumption that the individual has no right to make purchases at a new set of lower prices, but does have a right to the original price set in the case of price increases. In contrast, the *EV* contains the presumption that the individual has a right to (an obligation to accept) the new lower (higher) price set, and must be compensated (make a payment) if the new price set is not to be attained. Based on this interpretation of the two measures, some economists have argued that the choice between them is basically an ethical one—that is, one that depends on a value judgment as to which underlying distribution of property rights is more equitable (Krutilla 1967; Mishan 1976). All of this can be summarized as in Table 3.1.

For two alternative price changes, the welfare measures should be the same if both changes place the individual on the same higher indifference curve. However, if the two price changes place the individual on different indifference curves, the welfare measure should correctly indicate the preference ranking of the two alternatives. The *EV* measure always provides a consistent ranking in this sense, but the *CV* measure does not.

Figure 3.4 illustrates why this is the case. It shows an individual in equilibrium at point *A*, given prices and money income. Suppose that one policy proposal would increase the price of x_1 and decrease the price of x_2 simultaneously. The individual would achieve a new equilibrium at point *B*. The *CV* measure of the welfare change is shown as CV_{AB}. The second policy alternative would decrease the price of x_1 while increasing the price of x_2. This would lead to a new consumer's equilibrium at point *C*. Point *C* has been drawn on the same indifference curve as point *B*. Therefore, the measure of welfare change should be the same for the two policy alternatives. However, as can be seen by inspection, the *CV* for the second policy, CV_{AC}, is larger. The *CV* measure would indicate a preference for the second policy while the individual is in fact indifferent between the two policies. The *EV* gives the same welfare measure for the two policy alternatives. This is because the *EV* measure bases its comparison on a point on the indifference curve passing through the new equilibrium, but with the old prices. If two policies are on the same new indifference curve, the *EV* measure picks the same point for measuring the welfare effects for both policies.

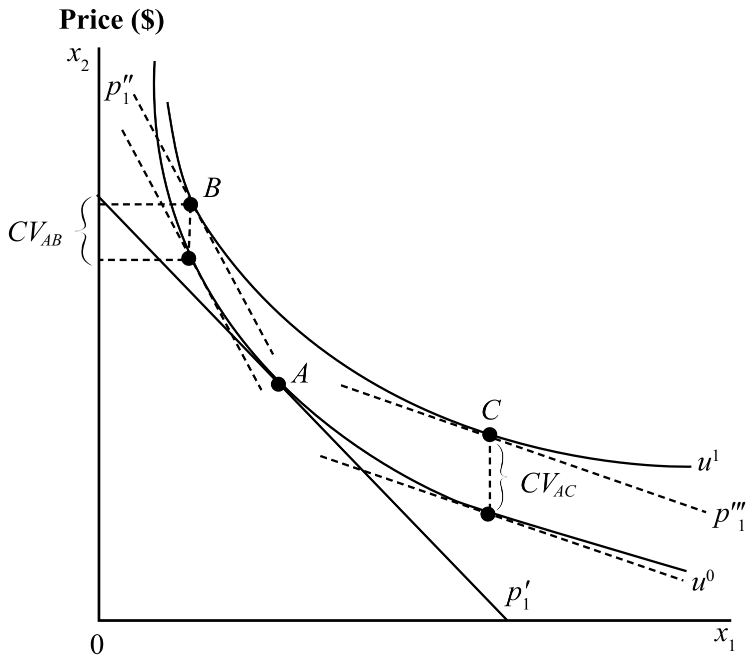

Figure 3.4 The compensating variation incorrectly ranks two alternative policies

If the question being asked by policymakers is, "does the proposed change pass the Kaldor potential compensation test?" then CV is the measure to use. The Kaldor potential compensation test is one form of potential Pareto improvement test that asks whether it is possible for the winners to fully compensate all of the losers from the proposed policy change and still leave someone better off. For each person, the CV gives the compensating income change required to maintain that person at his or her initial utility level. If the sum of what could be collected from all gainers exceeds the sum of the required compensations for losers, the proposal passes this form of the potential Pareto improvement test. The fact that the CV cannot rank consistently two or more policy changes is no obstacle to its use in this manner. This is because the potential Pareto criterion itself provides no basis for ranking two or more proposed policy changes. If two proposed changes both pass the Kaldor potential compensation test, the potential Pareto improvement criterion provides no basis for choosing between them.

On the other hand, if the question being asked by policymakers is, "does the policy pass the Hicks version of the potential compensation test?" then EV is the appropriate measure. The Hicksian test asks whether it is possible for the losers to bribe the gainers to obtain their consent to forgo the proposed policy change. The potential gainers would accept a bribe only if it were large enough to raise their utility by the same amount as the proposed policy would have. The offered bribe would have to be as large as each individual's *EV* measure of welfare gain; and the maximum bribe that would be offered by the potential losers would be their *EV* measure of loss. Thus if the sum of the *EV* of all gainers exceeded the sum of the *EVs* of all losers, the proposal would pass the Hicks form of the potential compensation test. Also, since the Hicks form of the compensation test is based on the *EV* measure, it will consistently rank two or more policy changes, provided that society is indifferent as to the distribution of gains and losses across individuals.

Measurement

Simply put, the problem posed for applied welfare economics is that the desired welfare measures, the *CV* or the *EV*, are based on the unobservable Hicks-compensated demand functions, while the one measure based on the observed Marshallian demand functions is flawed as a welfare indicator. The typical practice had been to use the Marshallian surplus anyway, and to offer such justifications as "income effects are likely to be small"; "with only one price change, path dependence is not an issue"; and "it is the only measure we have and it is better than nothing." Then Robert Willig (1976), in a widely cited article, provided a

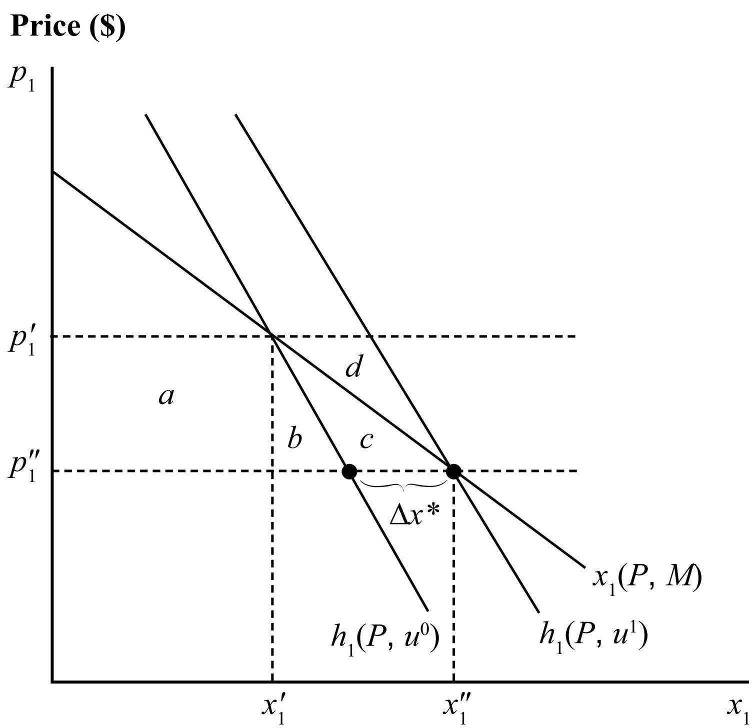

Figure 3.5 Deriving the Willig bounds for *S* as an approximation to *CV*

justification for using the Marshallian surplus by examining the magnitude of the differences between S and CV or EV under different conditions. Willig argued, "In most applications the error of approximation will be very small. In fact the error will often be overshadowed by the errors involved in estimating the demand curve" (1976, 589). Following his work, several authors have developed methods for direct calculation of the CV and EV from information contained in the ordinary demand function, either through a Taylor's series approximation (McKenzie and Pearce 1982; McKenzie 1983), or as exact measures through integration to obtain the indirect utility function and the expenditure function (see for example, Hausman 1981). The second subsection describes Hausman's contribution.

Consumer's Surplus without Apology

Willig (1976) has offered rigorous derivations of expressions relating CV, S, and EV. These expressions provide a way of calculating the magnitude of the differences among the three measures for given prices, quantities, and income. The differences among the three measures depend on the income elasticity of demand for the good in question and consumer surplus as a percentage of income. The differences among the measures appear to be small and almost trivial for most realistic cases. The differences are probably smaller than the errors in the estimation of the parameters of demand functions by econometric methods.

Willig's bounds for the approximation errors are based on the fact that the differences between S and CV or EV arise from an income effect on the quantity demanded; and the size of that effect depends on the change in real income brought about by the price change and on the income elasticity of demand for the good. This can be shown in a nonrigorous way for the case of one price change with the help of Figure 3.5. Although this exposition applies to the case of only one price change, the Willig expressions can be generalized to accommodate multiple price changes (Willig 1979), provided that a specific path of integration is chosen. In Figure 3.5, the ordinary and compensated demand curves are assumed to be linear. Let S represent the area $a + b + c$. So:

$$CV = a + b = S - c, \tag{3.28}$$

and

$$EV = a + b + c + d = S + d. \tag{3.29}$$

The errors in using S to approximate CV and EV are equal to the areas c and d respectively. For a price change from p_1' to p_1'', the factors influencing the size of the approximation error can be seen by examining the determinants of the area c:

$$CV - S = -c = -\tfrac{1}{2}\Delta p \cdot \Delta x^*, \tag{3.30}$$

where Δx^* is the income effect on the quantity demanded of x, which is associated with reducing income sufficiently to hold utility at u^0. Let ΔM^* represent this income change. By definition, ΔM^* is CV. The definition of income elasticity of demand is

$$E_M = \frac{\Delta x}{\Delta M} \cdot \frac{M}{x} \cdot \tag{3.31}$$

Solving this expression for Δx^* gives

$$\Delta x^* = E_M \cdot x \cdot \frac{\Delta M^*}{M} = E_M \cdot x \cdot \frac{CV}{M} \cdot \tag{3.32}$$

Substituting this into equation (3.30), we obtain

$$CV - S = -\frac{\Delta p \cdot x \cdot E_M \cdot CV}{2M} \cdot \tag{3.33}$$

In general, for small changes in p, $\Delta p \cdot x \approx S$. This is strictly true for the linear demand curve when x is evaluated at the midpoint between $x'$and x''. Finally, dividing both sides by CV to express the error in percentage terms gives

$$\frac{CV - S}{CV} \simeq -\frac{E_M}{2} \cdot \frac{S}{M} \cdot \tag{3.34}$$

This is similar to the Willig expression for the approximation error. The principal difference is that it expresses the error as a percentage of CV, while Willig's term makes the error a percentage of S. It says that the error is proportional to the income elasticity of demand and consumer surplus as a percentage of income. A similar line of reasoning can be used to derive the relationship between EV and S.

Willig's analysis is more rigorous than this in that it takes into account the possibility that for finite changes in price and quantity, the income elasticity of demand may vary over the range of the price change. Willig derived rules of thumb for calculating the maximum error in using S as an approximation for EV or CV. The rules of thumb are applicable if the following conditions are met:

$$\begin{aligned} &\frac{S}{M} \cdot \frac{\underline{E}_M}{2} \leq 0.05 \\ &\frac{S}{M} \cdot \frac{\overline{E}_M}{2} \leq 0.05 \end{aligned} \tag{3.35}$$

and

$$\frac{S}{M} \leq 0.9 , \tag{3.36}$$

where $\underline{E}_M$ and $\overline{E}_M$ are the smallest and largest values, respectively, of the income elasticity of demand for the good in the region under consideration.

Given these conditions, the rule of thumb for CV is

$$\frac{S}{M} \cdot \frac{\underline{E}_M}{2} \leq \left| \frac{CV - S}{S} \right| \leq \frac{S}{M} \cdot \frac{\overline{E}_M}{2} , \tag{3.37}$$

and the rule of thumb for EV is

$$\frac{S}{M}\cdot\frac{\underline{E}_M}{2}\leq\left|\frac{S-EV}{S}\right|\leq\frac{S}{M}\cdot\frac{\overline{E}_M}{2}\ . \tag{3.38}$$

The first thing to note is the conditions under which these rules of thumb are valid. Consider equation (3.36) first. The change in consumer surplus as a percentage of income depends on the size of the price change, the price elasticity of demand, and expenditure on this good as a percentage of total income. The smaller the price change and the smaller the proportion of income spent on the good, the smaller S/M becomes. It can readily be shown that

$$\frac{S}{M}\leq\left|\frac{\Delta p}{p}\right|\cdot\frac{px}{M}\cdot \tag{3.39}$$

From a given initial situation, S is largest when the demand curve is perfectly inelastic. Then $S=|x\cdot\Delta p|$ and (3.39) holds as an equality. With more elastic demand, $S<|x\cdot\Delta p|$ and the condition follows. For example, it shows that for a good absorbing 50 percent of total income and for a 100 percent price change, S/M cannot exceed 0.5, while for a 10 percent price change for a good absorbing 10 percent of income, S/M will be less than 0.1. Thus, condition (3.36) is likely to be satisfied except for very large price increases for goods with low price elasticities that also absorb a large proportion of the total budget.

As for the first condition, the smaller consumer surplus is as a percentage of income, and the smaller the income elasticity of demand is, the more likely it is that (3.35) be satisfied. For example, if consumer surplus is 5 percent of income, the income elasticity of demand can be as high as 2.0 and still satisfy (3.35). If S/M just barely satisfies condition (3.36), the income elasticity cannot exceed 0.11 to satisfy (3.35).

Assuming that conditions (3.35) and (3.36) hold, then let us turn to the rules of thumb. First, according to (3.35), the maximum error involved in using S as an approximation for either CV or EV is 5 percent. Second, the smaller the change in income elasticity over the range being considered, the more precise (3.37) and (3.38) are as statements of the error involved in using S rather than CV or EV. If the income elasticity of demand does not change over the range being considered, the left-hand and right-hand sides of (3.37) and (3.38) are equal to each other and the errors are zero, as discussed above. Finally, as the income elasticity of demand for the good decreases, the differences among ordinary consumer surplus, CV, and EV decrease, disappearing as E_M goes to zero.

Willig's analysis has been interpreted as providing a justification for using consumer surplus as an approximation of the CV or the EV. However, there are two reasons why one should be cautious about adopting the Willig approach to welfare measurement. The first has to do with limitations on the applicability of the Willig conditions to some kinds of problems of welfare measurement, including some of specific interest to environmental and resource economists.

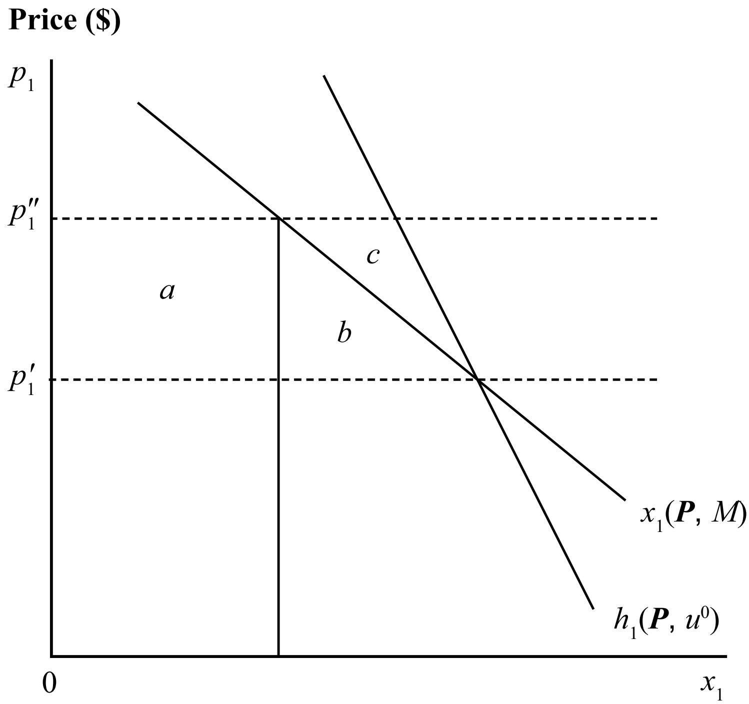

Figure 3.6 The Willig approximation and the error in estimating dead-weight losses

The second arises because of the recent development of new methods for obtaining exact measures of *CV* and *EV* from the same information that is required to use the Willig approximation.

The Willig conditions for valid approximation were developed for changes in *S* resulting from changes in the price of some market good. However, many environmental and resource policy issues require information on the total value of some environmental service as a measure of what would be lost if the resource were destroyed or diverted to some other use. For example, the economic cost of damming a river that provides whitewater canoeing and trout fishing would be measured by the total areas under the Hicks-compensated demand curves for these activities. This is equivalent to measuring the change in consumer surplus for a price increase from the present price to the vertical intercept of the Hicks-compensated demand curve. Bockstael and McConnell (1980) pointed out that for the linear demand function, the income elasticity of demand goes to infinity as the price approaches the vertical intercept; and thus, the approximation error cannot be calculated. In a comment on Bockstael and McConnell (1980), Hanemann (1980) showed that if the parameters of the Marshallian demand function were known, it was unnecessary to compute the Willig approximation error, since the *CV* could be calculated directly. In this, Hanemann (1980) apparently anticipated the analysis of Hausman (1981), discussed below.

For some questions, the variable of interest to policymakers is not *CV* but some fraction of *CV*—for example, the dead weight loss associated with a tax on a commodity. Suppose an excise tax raises the price of a good from p' to p'', as shown in Figure 3.6. The consumer's loss as measured by *CV* is the area $a + b + c$, but only $b + c$ is an efficiency loss, since a is a revenue transfer to the government. If the ordinary demand curve is used to approximate the consumer loss, the area c is the error. If the Willig conditions are satisfied, c is an acceptably small percentage of S and *CV*; but it can be an unacceptably large percentage of the true dead weight loss.

The second reason for being cautious about using the Willig approximation is that better methods of welfare measurement now exist. If the demand functions being used to calculate S reflect utility maximizing behavior on the part of individuals, they should satisfy the integrability conditions. If this is the case, it is possible to calculate *CV* and *EV* directly without approximation. On the other hand, if the demand functions do not satisfy the integrability conditions, then it is inappropriate to use the Willig approximations, since their derivation was also based on the assumption of utility-maximizing behavior.

Exact Welfare Measurement

Hausman (1981) presented a procedure for exact welfare measurement based on the recovery of the parameters of the utility function from data on consumers' demand. His procedure, which was developed for the case of only one price change, involves four steps. The first involves combining the ordinary demand function and Roy's identity to obtain a partial differential equation:

$$x_1(\boldsymbol{P}, M) = -\frac{\partial v(\boldsymbol{P}, M)/\partial p_1}{\partial v(\boldsymbol{P}, M)/\partial M} \tag{3.40}$$

If the utility function is separable so that the demand function contains only its own price argument, and if the demand function is linear, this becomes:

$$a - (b \cdot p_1) + (c \cdot M) = -\frac{\partial v(\boldsymbol{P}, M) / \partial p_1}{\partial v(\boldsymbol{P}, M) / \partial M}, \tag{3.41}$$

where the parameters a, b, and c are estimated econometrically, and where p_1 and M are deflated by an appropriate index of the other prices. Changes in p_1 and M that involve moving along an indifference curve must satisfy

$$\left[\frac{\partial v(\cdot)}{\partial p_1(t)} \frac{dp_1(t)}{dt}\right] + \left[\frac{\partial v(\cdot)}{\partial M(t)} \frac{dM(t)}{dt}\right] = 0, \tag{3.42}$$

where t defines a path of price changes. Rearranging this expression, substituting into (3.41), and using the implicit function theorem gives

$$\frac{dM(p_1)}{dp_1} = a - (b \cdot p_1) + (c \cdot M), \tag{3.43}$$

the solution of which is

$$M(p_1) = k \cdot \exp(cp_1) - \frac{1}{c}\left[a - (b \cdot p_1) - \frac{b}{c}\right], \tag{3.44}$$

where k is the constant of integration, which depends on the initial level of utility. If units are arbitrarily chosen so that k is the initial utility level, the quasi-indirect utility function and quasi-expenditure function follow directly:

$$u = k^0 \cdot \exp(-cp_1)\left\{\left(M + \frac{1}{c}\right)\left[a - (b \cdot p_1) - \frac{b}{c}\right]\right\} \tag{3.45}$$

and

$$e = k^0 \cdot \exp(cp_1) - \frac{1}{c}\left[a - (b \cdot p_1) - \frac{b}{c}\right]. \tag{3.46}$$

These expressions are termed "quasi" functions because they do not contain information about the effects of the prices of other goods on utility or expenditure. Hausman's method depends on the ability to solve the differential equation that is obtained from Roy's identity. Hausman has shown a method of solution for the case when only one price changes, and has discussed in general terms the solution in the case of multiple price changes.

Conclusions

Selection of a welfare measure has long involved questions both of appropriateness and of practicality. The Marshallian surplus measure was frequently chosen on the grounds of practicality, even though it was recognized that the measure was inappropriate in that it did not answer any specific well-formed welfare question. Willig's development of the bounds for the errors of approximation in using S gave encouragement to this practice. However, quickly on its heels have come new approaches to exact welfare measurement that offer the opportunity to calculate the more appropriate CV and EV measures directly.

One question related to practicality remains, however—do we know enough about the functional form of the utility function to implement the exact measurement methods? Assuming a functional form for the system of demand functions for purposes of estimation is equivalent to assuming the functional form of the underlying utility function. One approach is to assume a specific functional form for the utility function or indirect utility function, and to derive the demand functions for estimation. If this is the approach taken, then plugging the estimated parameters back into the utility function to calculate welfare changes is straightforward, provided the parameter estimates of the demand function satisfy the integrability conditions. Since researchers have been reluctant to specify the functional form of the utility function, one alternative has been to specify so-called *flexible forms* for the indirect or direct utility function (for example, Deaton and Muellbauer 1980). Again, if the integrability conditions are satisfied, deriving "exact" welfare measures from the "approximate" flexible functional form of the utility function is straightforward. The alternative is to seek guidance from the data by selecting the functional form for the demand functions based on goodness-of-fit and consistency with the restrictions imposed by theory.

Welfare Measures for Continuous Goods: Quantity Changes

Many environmental policy proposals involve changes in either the quantities or the qualities of nonmarket environmental goods and services, rather than changes in the price of a marketable good. From the individual's point of view, the most important characteristic of some environmental goods is that they are available only in fixed, unalterable quantities. These quantities act as constraints on each individual's choice of a consumption bundle. The analysis of this class of problems is often referred to as the theory of choice and welfare under quantity constraints (Johansson 1987). The imposition of quantity constraints raises some new issues in the theory of choice and welfare measurement. The analysis of these problems has evolved out of the theory of rationing as initially developed by Tobin and Houthakker (1950/1), and Neary and Roberts (1980).

This section provides a brief description of the model of individual preferences and choice under imposed quantity constraints. The corresponding measures of welfare impacts for changes in the quantities of imposed goods are then derived. These measures are essentially similar to the compensating and equivalent surplus measures for price changes presented in Hicks (1943), but the change being considered is one of a quantity or quality change, rather than price. As mentioned earlier, Hicks referred to these measures as compensating or equivalent "surplus," and this terminology convention is continued in this chapter. The section closes with a brief discussion of the value of changes in q when q is a bad.

The Basic Model

Consider an individual whose utility function has the following form:

$$u(\boldsymbol{X}, \boldsymbol{Q}), \tag{3.47}$$

where $\boldsymbol{X} = \left(x_1, \ldots, x_J\right)$ is the vector of private goods quantities, and $\boldsymbol{Q} = \left(q_1, \ldots, q_K\right)$ is a vector of environmental and resource service flows (unpriced public goods) that is exogenous to the individual. It is possible that there is a positive price for at least some of the elements in $\boldsymbol{Q}$; but to keep the exposition simple, all prices for elements of $\boldsymbol{Q}$ are assumed to be zero. Let $\boldsymbol{P} = \left(p_1, \ldots, p_J\right)$ be the vector of prices for $\boldsymbol{X}$. The individual maximizes utility subject to a budget constraint

$$\boldsymbol{P} \cdot \boldsymbol{X} = M , \tag{3.48}$$

where M is money income. This yields a set of conditional demand functions for the marketed goods:

$$x_j = x_j \left(\boldsymbol{P}, M, \boldsymbol{Q} \right). \tag{3.49}$$

In general, $\boldsymbol{Q}$ will be an argument in these conditional demand functions, along with prices and income. The term "conditional" refers to the fact that these functions are conditioned upon the imposed $\boldsymbol{Q}$.

Inserting the conditional demand functions into the utility function gives the conditional indirect utility function

$$v = v\left(\boldsymbol{P}, M, \boldsymbol{Q} \right). \tag{3.50}$$

Inverting the conditional indirect utility function for M yields a conditional expenditure function that gives the minimum expenditure on market goods required to produce utility level u, given $\boldsymbol{P}$ and $\boldsymbol{Q}$. This is

$$e = M = e\left(\boldsymbol{P}, \boldsymbol{Q}, u \right). \tag{3.51}$$

For simplicity, in what follows $\boldsymbol{Q}$ is assumed to consist of only one element, q. In order to make graphic presentations of some of the key points, it is assumed that $\boldsymbol{X}$ is the numeraire, represented as x with a price of 1. Finally, it is assumed that at the given prices and income, the individual would choose more of q if given the option (i.e., q is a "good").

To begin with, the marginal value of a small increase in q is the reduction in income that is just sufficient to maintain utility at its original level. If w_q is the marginal value or marginal willingness to pay for a change in q, it is given by the derivative of the restricted expenditure function with respect to q or

$$w_q = -\frac{\partial e}{\partial q}. \tag{3.52}$$

The right-hand side of this expression is also equal (in absolute value) to the slope of the indifference curve through the point at which the welfare change is being evaluated. There are several ways to present compensating surplus (CS) and equivalent surplus (ES) for changes in quantity-constrained goods.

The first way is based on the conditional indirect utility function. The CS and ES measures are defined implicitly as the solutions to the following expressions:
CS is the solution to

$$v\left(\boldsymbol{P}, M, q^0 \right) = v\left(\boldsymbol{P}, M - CS, q^1 \right), \tag{3.53}$$

and ES is the solution to

$$v\left(\boldsymbol{P}, M + ES, q^0 \right) = v\left(\boldsymbol{P}, M, q^1 \right). \tag{3.54}$$

These two measures can also be defined in terms of the conditional expenditure function. For a change in q, CS is

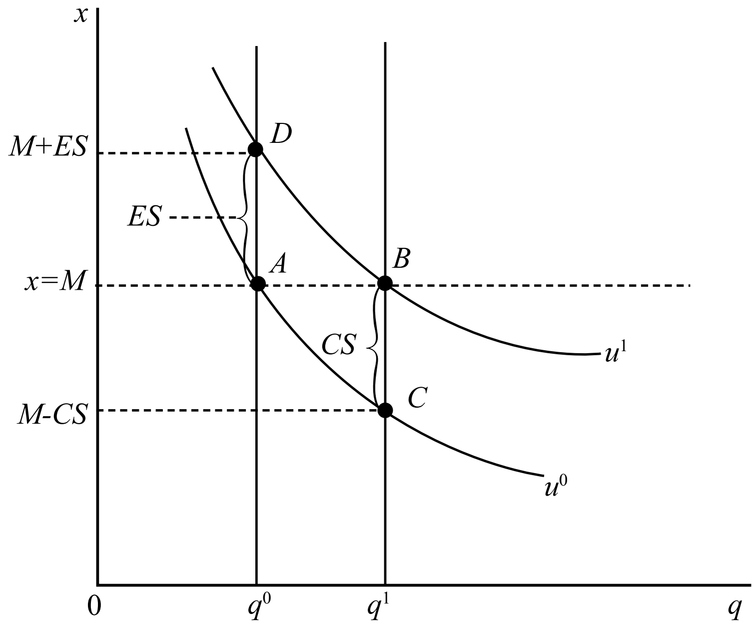

Figure 3.7 Compensating and equivalent surpluses for a change in *q*

$$CS = e\left(\boldsymbol{P},q^0,u^0\right) - e\left(\boldsymbol{P},q^1,u^0\right) = M - e\left(\boldsymbol{P},q^1,u^0\right). \tag{3.55}$$

The *ES* measure given by the conditional expenditure function is

$$ES = e\left(\boldsymbol{P},q^0,u^1\right) - e\left(\boldsymbol{P},q^1,u^1\right) = e\left(\boldsymbol{P},q^0,u^1\right) - M. \tag{3.56}$$

ES and *CS* are shown graphically in Figure 3.7. The increase in q enables the individual to reach point *B* with utility equal to u^1. The *CS* is the distance *B-C*. Alternatively, if income increased by the *ES* value while holding q constant, the individual could achieve u^1 at point *D*. Thus, *ES* is the distance *A-D*.

A second way to derive the *ES* and *CS* measures is also based on the conditional expenditure function. The value of a nonmarginal change in q is the integral of this function taken over the relevant range, or

$$W_q = -\int_{q^0}^{q^1} \frac{\partial e\left(\boldsymbol{P},q,u^t\right)}{\partial q} dq. \tag{3.57}$$

This is either a *CS* or an *ES* measure, depending on whether $t = 0$ or $t = 1$.

Before leaving this section, note that there are two ways in which more q could be a bad, rather than a good, for an individual. The first way is when q has a price greater than zero and the individual would prefer to have less than the quantity being imposed given that price. The welfare measures *ES* and *CS* are still defined in the same way, but now they are negative for increases in q and positive for decreases in q.

The second way in which q can be a bad is the more fundamental one—it is when the marginal utility of q is negative. Even at a zero price, the individual would prefer to receive a smaller quantity. In both cases, the welfare measures *ES* and *CS* are defined in the same way, and again they are negative for increases in q and positive for decreases in q. In addition, all of the discussion of exact welfare measurement techniques and approximations carries over with appropriate changes to the case of q as a bad.

Welfare Measures for Discrete Goods

In the first part of this chapter the models described for changes in price exploited the marginal equalities revealed when individuals optimize over choice variables that are continuously variable. This is not always a realistic way to model the individual choice problem. Some problems are better viewed as involving the choice of one option from a range of discrete alternatives. For example, the choice might be whether or not to take a once-in-a-lifetime cruise around the world, or whether to travel to work by private auto, bus, or on foot. The solutions of discrete choice problems of this sort are essentially corner solutions. Consequently, there are no tangencies from which a marginal rate of substitution can be inferred. Discrete choice models have been developed both to predict individuals' behaviors in these choice contexts and to draw inferences about welfare change on the basis of observed choices.

In this section, a simple discrete choice model is presented, and measures of welfare change and value are derived from the model. Welfare measures for both price changes and quantity changes are considered. Subsequent chapters present detailed discussions of applications and estimation approaches. A wide range of environmental problems and decision making can be represented in a discrete choice setting, including: voting yes or no on a referendum question; accepting or rejecting a hypothetical offer for an environmental commodity; the choice of which of several alternative houses to live in based in part on environmental quality in their vicinity; and the choices of whether or not to undertake a specific recreation activity or to visit a specific recreation site. For expositions of the specification, estimation, and interpretation of discrete choice models generally, see Ben-Akiva and Lerman (1985) and Train (2009). Hanemann (1999) gave a more advanced exposition in the context of valuing environmental changes. See also Johansson, Kriström, and Mäler (1989) and Hanemann (1989).

Consider an individual's decision regarding which one of several alternative goods to purchase. The individual can choose one good from a set of J alternatives ($j = 1, \ldots, J$), where each good has a vector of environmental quality attributes $\boldsymbol{Q}_j$ associated with it. The price for good j is p_j. The individual gets utility from the discrete good chosen and the consumption of a numeraire good.

With this construction, a conditional utility function associated with each alternative can be written as:[3]

$$u_j = u_j(M, p_j, \boldsymbol{Q}_j), j = 1,...,J \ . \tag{3.58}$$

The individual will choose to consume the alternative that yields the highest utility; that is, the chosen alternative "j" will satisfy

$$u_j(M, p_j, \boldsymbol{Q}_j) > u_k(M, p_k, \boldsymbol{Q}_k), \quad j,k = 1,...,J \ . \tag{3.59}$$

It is straightforward at this point to implicitly define the compensating and equivalent variation associated with a price change for one or more of the alternatives. Specifically, the compensating variation associated with a decrease in all prices of the discrete alternatives can be written implicitly as:

$$\underset{j}{\text{Max}}\ u_j\left(M, p_j^0, \boldsymbol{Q}_j\right) = \underset{j}{\text{Max}}\ u_j\left(M - CV, p_j^1, \boldsymbol{Q}_j\right), \tag{3.60}$$

where superscript "0" indicates the original price, and superscript "1" indicates the new, lower set of prices. The expression makes clear that the option chosen after compensation is paid could also differ from either the original alternative or the choice without compensation. Likewise, equivalent variation can be written as:

$$\underset{j}{\text{Max}}\ u_j\left(M + EV, p_j^0, \boldsymbol{Q}_j\right) = \underset{j}{\text{Max}}\ u_j\left(M, p_j^1, \boldsymbol{Q}_j\right) \ , \tag{3.61}$$

where the base level of utility is the utility associated with the new price vector rather than the original.

It is also straightforward to construct the compensating and equivalent surplus measures associated with a change in the vector of quality attributes associated with each alternative:

$$\begin{aligned} &\underset{j}{\text{Max}}\ u_j\left(M, p_j, \boldsymbol{Q}_j^0\right) = \underset{j}{\text{Max}}\ u_j\left(M - CS, p_j, \boldsymbol{Q}_j^1\right), \\ &\underset{j}{\text{Max}}\ u_j\left(M + ES, p_j, \boldsymbol{Q}_j^0\right) = \underset{j}{\text{Max}}\ u_j\left(M, p_j, \boldsymbol{Q}_j^1\right). \end{aligned} \tag{3.62}$$

A common representation of the utility function is additive. By also recognizing that the budget constraint implies that the amount of the numeraire that can be consumed when alternative "j" is chosen is $M - p_{j,}$ the conditional utility function can be written as:

$$u_j = \beta(M - p_j) + \tilde{u}_j(\boldsymbol{Q}_j), \quad j = 1,...,J \ , \tag{3.63}$$

where β can be interpreted as the marginal utility of income, and $\tilde{u}_j(\boldsymbol{Q}_j)$ is a function representing the utility associated with the quality aspects of the alternative. With this specification, the compensating and equivalent surpluses for

3 Note that this is a conditional *indirect* utility function. We depart from our standard notation used throughout the rest of the book and use $u(\cdot)$ to denote an indirect utility function in this case for consistency with the established literature in this area.

a quality change are identical (a direct result of the constant marginal utility of income, β) and can be written as:

$$CS = ES = \frac{1}{\beta}\{\underset{j}{\mathrm{Max}}\, u_j(M, p_j, \boldsymbol{Q}_j^1) - \underset{j}{\mathrm{Max}}\, u_j(M, p_j, \boldsymbol{Q}_j^0)\}, \quad j = 1,...,J. \tag{3.64}$$

Expression (3.64) is intuitively appealing, as it says that the compensating and equivalent variation associated with a quality change is simply the difference in utility from the most desirable alternatives before and after the change, divided by the marginal utility of income. The marginal utility of income acts to monetize the utility difference.

Thus far, the discrete choice behavioral model and associated welfare measures have been presented in a deterministic form, just as the behavioral model underlying the continuous demand functions and their associated welfare measures were presented earlier in this chapter. Typically, however, analysts employing the discrete choice model recognize that there are individual characteristics and/or omitted variables that are not observable to the researcher, but are known to the individual making the decision. To incorporate this idea, an additive error can be added to the observable component

$$u_j = v_j(M, p_j, \boldsymbol{Q}_j) + \varepsilon_j,\ j = 1,...,J, \tag{3.65}$$

where ε_j is a random, unobservable component of utility. As before, utility maximizers will choose the alternative that yields the highest utility, but from the perspective of the analysis, the utility is now random. This "random utility maximization" model, or RUM model (Thurstone 1927; Marschak 1960; McFadden 1974, 1978, 1981), implies the probability that the individual chooses to purchase alternative "k" can be expressed as the probability that the utility associated with k is greater than the utilities associated with all the other alternatives:

$$\Pr(k) = \Pr\begin{bmatrix} v_k(M - p_k, \boldsymbol{Q}_k) \\ + \varepsilon_k > v_j(M - p_j, \boldsymbol{Q}_j) + \varepsilon_j \end{bmatrix}, \forall j \neq k. \tag{3.66}$$

McFadden (1974) demonstrated that if the error terms are independently and identically distributed (i.i.d.) with a Type I Extreme Value distribution, a logistic distribution results and this probability can be written simply, as follows:

$$\Pr(k) = \frac{e^{v_k(M-p_k, \boldsymbol{Q}_k)}}{e^{v_k(M-p_k, \boldsymbol{Q}_k)} + \sum_{j \neq k}^{J} e^{v_j(M-p_j, \boldsymbol{Q}_j)}} = \left(1 + \sum_{j \neq k}^{J} e^{-\Delta v_{jk}}\right)^{-1} \tag{3.67}$$

$$\text{where} \quad \Delta v_{jk} = v_j\left(M - p_j, \boldsymbol{Q}_j\right) - v_k\left(M - p_k, \boldsymbol{Q}_k\right).$$

The logit model of choice implies certain restrictions on individuals' choices and preferences. The most notable is that choices must have the property of the

Independence of Irrelevant Alternatives (IIA). This issue, and a host of additional topics related to interpretation and estimation of RUMs, will be discussed in later chapters.

The introduction of an error term complicates welfare computation since only the probability of choosing a particular alternative under a price or quality change can be considered. A general expression for the compensating variation associated with a price change is:

$$\operatorname*{Max}_{j}\left[v_j(M, p_j^0, \boldsymbol{Q}_j) + \varepsilon_j\right] = \operatorname*{Max}_{j}\left[v_j(M - CV, p_j^1, \boldsymbol{Q}_j) + \varepsilon_j\right], \tag{3.68}$$

where $CV = CV\left(M, \boldsymbol{P}^0, \boldsymbol{P}^1, \boldsymbol{Q}_j, \varepsilon\right)$ and $\varepsilon = \left(\varepsilon_1, \ldots, \varepsilon_{\mathcal{J}}\right)$ denotes the full vector of error terms. A corresponding equivalent expression for EV can be written. As the notation indicates, this welfare measure will itself be a random variable and its expected value can be computed (Small and Rosen 1981; Hanemann 1984). Using the linear functional form identified in (3.63), compensating and equivalent variations are equal to each other. If in addition, the error terms are Type I Extreme Value, then the mean CV and EV terms take a particularly simple form, with:

$$\overline{CV} = \overline{EV} = \frac{1}{\beta}\left\{\ln\left[\sum_{j=1}^{\mathcal{J}} e^{v_j^1}\right] - \ln\left[\sum_{j=1}^{\mathcal{J}} e^{v_j^0}\right]\right\}, \tag{3.69}$$

where $v_j^t = v_j\left(M, p_j^t, \boldsymbol{Q}_j\right)$ for $t = 0, 1$. Similar calculations can be used to obtain the value of adding or deleting a site with a specified set of characteristics from the individual's choice set. For the addition of site $\mathcal{J} + 1$, the expression is

$$\overline{CV} = \overline{EV} = \frac{1}{\beta}\left\{\ln\left[\sum_{j=1}^{\mathcal{J}+1} e^{v_j}\right] - \ln\left[\sum_{j=1}^{\mathcal{J}} e^{v_j}\right]\right\}, \tag{3.70}$$

and for deleting site $\mathcal{J}$, the expression is

$$\overline{CV} = \overline{EV} = \frac{1}{\beta}\left\{\ln\left[\sum_{j=1}^{\mathcal{J}-1} e^{v_j^1}\right] - \ln\left[\sum_{j=1}^{\mathcal{J}} e^{v_j^0}\right]\right\}. \tag{3.71}$$

These measures are examples of compensating and equivalent variation approaches to defining a welfare measure using a random utility framework. Hanemann described two such approaches and examined the relationships among them (Hanemann 1999, 43–48).

When CV and EV Diverge: Willingness to Pay versus Willingness to Accept Compensation

The results from Willig discussed earlier imply that measures of compensating variation (or surplus) should in theory generally be very close to their associated equivalent variation (or surplus) measures. Since these measures have willingness to pay for and willingness to accept compensation interpretations, another way to say the same thing is that WTP to acquire a good or price change should typically approximately equal WTA to do without the change. However, there is a substantial body of evidence from stated preference studies, laboratory experiments, and field experiments that suggests that differences between WTP and WTA for the same good can be quite large (Horowitz and McConnell 2002; Sayman and Onculer 2005). Efforts at explaining these differences have taken several paths.

One argument is that these divergences do not represent actual divergences in preferences but reflect experience with the good and the trading environment in which the values are elicited. List (2003, 2004) studied the divergence in an actual marketplace and found that the disparity is highly correlated with experience in the market: those who have extensive experience in buying and selling the good (sports memorabilia at trade shows) exhibit no meaningful disparity. Focusing on the experimental environment in which these values are elicited, Plott and Zeiler (2005) argued that when a full suite of experimental controls is employed, the divergence between WTP and WTA disappears. They presented findings from three experiments to support their argument and concluded that the differences between WTP and WTA reported in the literature relate to misconceptions that subjects have about the task they faced in the experiment, rather than representing a reflection of true value disparity. Fudenberg, Levine, and Maniadis (2012) undertook a similar set of experiments (though their focus was on "anchoring" effects) and found evidence for the existence of the disparity, albeit of smaller size than many previous studies. Other authors have suggested and studied explanations that relate to the value elicitation environment (Hoehn and Randall 1987; Kolstad and Guzman 1999; Guzman and Kolstad 2007).

A second path involves examining the theory of preferences and value more closely to see whether theory predicts the large disparities between true WTA and WTP. One example of this is in the work of Hanemann (1991, 1999). He has shown that the price flexibility of income can be expressed as the ratio of two other terms:

$$E_q = \frac{E_M}{\sigma_q}, \tag{3.72}$$

where σ_q is the aggregate Allen–Uzawa elasticity of substitution between q and the composite commodity X and E_M is the income elasticity of demand for q. If the elasticity of substitution (a measure of the curvature of the indifference curve between q and private goods) is low, σ_q can be close to zero. This can lead to a high value for E_q and a large difference between CS and ES. However, Hanemann's analysis does not explain the persistent differences between the two measures in experiments with simulated markets involving commonplace goods such as lottery tickets, coffee mugs, and pens (see Knetsch and Sinden 1984; Kahneman, Knetsch, and Thaler 1990).

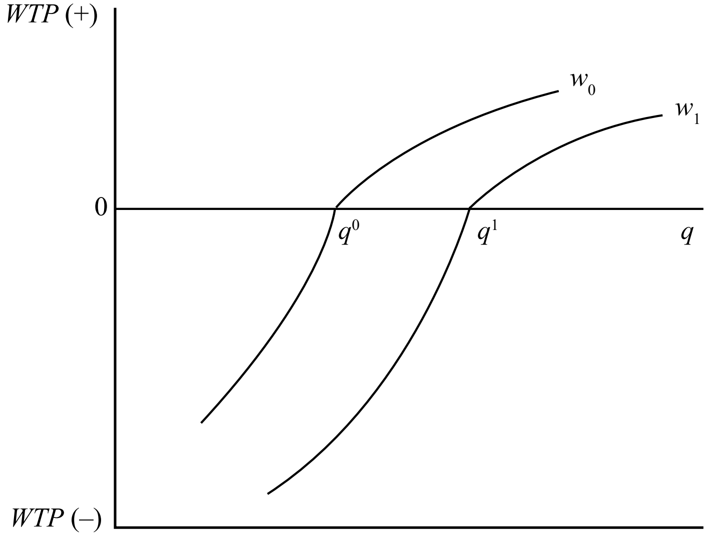

Figure 3.8 The value function and the endowment effect

Zhao and Kling (2001, 2004) have also offered a possible explanation that is largely consistent with the standard paradigm. They considered consumers who make decisions about whether to buy or sell goods whose value is uncertain to them when they have the opportunity to delay the decision and gather more information in the meantime. They demonstrated that there are conditions under which this will lead to lower WTP values and higher WTA values than theory would predict in the absence of this potential for learning. The dynamic welfare measures they derived will be further discussed in Chapter 5.

A final approach, and the one that seems to have gained the most traction, has been to move further from standard economic theory. Thaler (1980) proposed that the reconciliation of theory with observation can be brought about by postulating an "endowment effect" on individuals' valuation functions and a kink in this function at the status quo point. He suggested that this is a reasonable extension and generalization of the prospect theory of Kahneman and Tversky (1979) to choices not involving uncertainty. The idea of the endowment effect and the differential valuation of gains and losses can be shown with the aid of Figure 3.8. The horizontal axis shows the quantity of an environmental good q. The vertical axis shows the compensating welfare measure for changes in q. This measure is positive (WTP) for increases and negative (WTA) for decreases from some status quo point. Suppose that the status quo is q_0. The associated valuation function w_0 shows the monetary payment (compensation) that holds utility constant for a given increase (decrease) in q from q_0. This function is kinked at the status quo point of q_0, showing that the marginal valuation of increases in q is substantially lower than the marginal valuation of losses from q_0. A change in the endowment of q from q_0 to q_1 shifts the valuation function. In addition, as Figure 3.8 shows, the willingness to pay for an increase from q_0 to q_1 is substantially less than the required compensation for the decrease from q_1 to q_0.

In conclusion, although the observed large differences between WTP and WTA can be explained by replacing the standard utility model with one that incorporates an endowment effect, it is not clear that this is always necessary. These differences can also be explained by the absence of close substitutes in the case of unique and perhaps irreplaceable resources and as the rational response to uncertainty and the high cost of information about preferences.

Aggregation and Social Welfare

Assume now that we have obtained measures of the welfare changes, either plus or minus, for all individuals. How can we use that information to make choices about public policy alternatives? To put the question in its most profound sense, what is the appropriate relationship between the welfare of individuals and the social welfare? What follows is a brief review of alternative social welfare criteria. Since the main concern of this book is with measurement, the question of social welfare criteria—that is, how to use the measures—is off the main track. For a more extensive discussion of the problem, see Mishan (1960), especially section III, and Boadway and Bruce (1984).

In the literature on welfare economics there are basically four ways to approach this question. The first approach to the question is the so-called Pareto criterion. Only policy changes that make at least one person better off (that is, an individual experiences a positive welfare change) and make no individual worse off (that is, no individual experiences a negative welfare change) pass this criterion. This criterion deliberately rules out any attempts to add up, or otherwise make commensurable, the welfare measures of different individuals. Since virtually all actual public policy proposals impose net costs on at least some individuals, most policy actions by the state could not be accepted under this criterion. This would be particularly true in the environmental area, where environmental management costs are often channeled through the production sector while benefits accrue to households in the form of increased levels of environmental services. It is unlikely that this would result in a pattern of incidence of benefits and costs in which no one would lose. The restrictive features of the Pareto criterion have stimulated an ongoing search for a welfare criterion that would justify the state doing certain things that at least some people feel it should be able to do.

The second approach to the question was proposed in slightly different forms by Kaldor (1939) and Hicks (1939)—these are the two different forms of a potential compensation test discussed earlier. Let us review these tests in the present context of aggregation and social welfare.

As noted earlier, the Kaldor version of the test asks whether those who gain by the policy can fully compensate for the welfare losses of those who lose by the policy. The Kaldor version of the test would be satisfied if the sum of all individual *CV* and *CS* measures of welfare changes were greater than zero. The criterion is essentially one of potential Pareto improvement, since if the compensation were actually paid no one would lose from the policy.

The Hicks version of the potential compensation test asks whether those who lose from the policy could compensate the gainers for a decision not to proceed with the policy. If the answer is yes, the policy should be rejected according to the Hicks criterion. If the policy was rejected and compensation was actually paid, those who would have gained from the policy would be just as well off as if the policy had been adopted, and those who would have lost are at least as well off as they would have been with the policy. The Hicks version of the test takes acceptance of the project as its reference point. In effect, it is a decision to forgo the project that creates the gains and losses that are relevant to the Hicksian version of the potential Pareto improvement criterion.

Should compensation actually be paid in either the Kaldor or Hicks cases? If one thinks the answer should be yes, then the compensation test is transformed into a variation of the Pareto criterion in which the state serves to enforce the taxes and transfer payments that are necessary to ensure that no one actually experiences a welfare loss, assuming that such taxes and transfers would be costless. If one thinks that the answer should be no, this is equivalent to, in effect, assuming that all individual welfare changes are commensurate and can be summed together into an aggregate measure of welfare change. This is the efficiency criterion of the new welfare economics. According to the efficiency criterion, the objective of social policy is to maximize the aggregate value of all of the goods and services people receive, including environmental and resource services. One justification for the Hicks–Kaldor potential compensation test is that a large number of efficient projects will spread benefits sufficiently wide so that everyone is a net gainer from the set of projects taken as a whole, even though some might be losers on individual projects. See Polinsky (1972) for an interesting development of this line of reasoning.

Alternatively, one might believe that whether compensation should be paid depends upon who has to pay and who gets the benefits. This requires consideration of the equity (fairness) in the distribution of income as an element in the evaluation of social policy. The third approach to the question of social welfare criteria, proposed by Little (1957), makes explicit the concern for equality. He proposed a twofold test. First, does the policy pass the Kaldor test? Second, does the resulting change improve the distribution of income? The Little criterion legitimizes a concern with the distributional effects of changes in resource allocation, but it does not resolve the question of what constitutes an improvement.

The fourth approach to the question involves an attempt to make specific social judgments regarding equity, and to introduce equity considerations systematically into the evaluation of social policy. The most common proposal calls for the establishment of a social welfare function that gives different weights to individual welfare changes according to the relative deservingness of the different individuals (Eckstein 1961; Haveman and Weisbrod 1975). Of course, the main problem with the social welfare weight approach is the determination of the weighting function (Freeman 1971).

Nevertheless, willingness to make explicit value judgments about equity makes it possible to consider a wider range of policy choices. For example, if one opts for the Pareto criterion or the potential compensation version of the Hicks–Kaldor test, one rules out the possibility of accepting a project that has a sum of individual welfare changes that is less than zero, but would substantially improve the distribution of income. An example of such a policy would be one that imposes a welfare loss of $1,000 on a millionaire while bringing benefits of $99 to each of ten impoverished orphans. A welfare-weighting function could approve negative sum policies like this, provided that the weights given to the beneficiaries were sufficiently greater than the welfare weights of the losers. In addition, neither of these criteria would reject a project that imposes costs on no one, but distributes benefits only to the richest in our society. Some might make the value judgment that this, in itself, is undesirable. A social welfare function that included some measure of inequality of the aggregate distribution as an argument might reject inequality-creating projects like this, and it would also be likely to accept negative sum projects that reduced inequality.

The potential compensation test criterion is perhaps the most controversial feature of standard welfare economics. On the one hand, it has been criticized as being incompatible with the Pareto criterion since it allows for a ranking of projects that are Pareto noncomparable. On the other hand, many economists argue that lump sum transfers or other means of transferring wealth are a more appropriate way for addressing equity concerns. Thus, one should adopt projects that pass the potential compensation test *and* also take steps to efficiently address distributional concerns. In any case, these concerns have not deterred governments from using it for some kinds of policy choices, and economists from advocating greater use of it in a wider range of environmental and resource policy questions.

Summary

This chapter has provided a derivation and explanation of the compensating and equivalent measures of individual welfare change for changes in prices and quantities for both discrete and continuous goods. The compensating and equivalent measures answer different kinds of policy-relevant questions because they make different implicit assumptions about the relevant status quo. It is interesting to examine some hypothetical examples.

Suppose that the question is whether to locate a landfill in a particular neighborhood. The neighbors are likely to oppose this proposal, and suppose that it is accepted that the neighbors have a right to an undisturbed neighborhood. Then the relevant measure of the harm for locating the landfill in their neighborhood would be the sum of their compensating measures of loss (*CV* and *CS*). The appropriate measure of the gain to those who would use the landfill would be their willingness to pay to locate it in this neighborhood—also a compensating

Table 3.2 Implied property rights and associated welfare measures

Implicit "rights"	*Policy question*	*Gainers*	*Losers*
To the present polluter	Require cleanup?	Neighbors, compensating measure (WTP)	Polluter, compensating measure (WTA)
To the potential polluter	Allow pollution?	Polluter, equivalent measure (WTA)	Neighbors, equivalent measure (WTP)
To the neighbors	Require cleanup?	Neighbors, equivalent measure (WTA)	Polluter, equivalent measure (WTP)
To the neighbors	Allow pollution?	Polluter, compensating measure (WTP)	Neighbors, compensating measure (WTA)

measure. Alternatively, if it is argued that the larger society has a right to locate the landfill anywhere, then what is relevant is the neighbors' willingness to pay to keep it out of their neighborhood. This is an equivalent measure of the potential loss (*EV* and *ES*). For the users of the landfill, the value of locating the landfill in this neighborhood is what its users would require to compensate them for locating it in a less desirable place—an equivalent measure of benefit.

Suppose, instead, that the offending facility is a polluting factory that has been in the neighborhood for a long time. If the neighbors are deemed to have a right to a clean neighborhood, then the appropriate reference point for welfare measurement is their utility levels after the factory has stopped polluting. This implies an equivalent measure of welfare change (*EV* and *ES*). Specifically, this is a measure of the compensation that the neighbors would require to forgo having the pollution stopped, and a measure of the factory owners' willingness to pay to continue to pollute. Alternatively, if the factory has a right to pollute, compensating measures of the gain from stopping the pollution are appropriate (*CV* and *CS*).

In each case, the appropriate welfare measure can be found by examining the nature of the social transaction that is implied by the policy decision at hand, and by the implicit rights to the services of the environment presumed to be held by the various parties to the transaction. The results for the examples discussed here can be summarized in Table 3.2.

References

Ben-Akiva, Moshe, and Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge: MIT Press.

Bergson, Abram. 1966. *Essays in Normative Economics.* Cambridge: Harvard University Press.

Boadway, Robin, and Neil Bruce. 1984. *Welfare Economics.* Oxford: Basil Blackwell.

Bockstael, Nancy E., and Kenneth E. McConnell. 1980. "Calculating Equivalent and Compensating Variation for Natural Resource Facilities." *Land Economics* 56(1): 56–63.

Bockstael, Nancy E., and Kenneth E. McConnell. 2007. *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models.* Dordrecht: Springer.
Camerer, Colin F., George Loewenstein, and Matthew Rabin, eds. 2004. *Advances in Behavioral Economics.* New York: Princeton University Press.
Currie, John M., John A. Murphy, and Andrew Schmitz. 1971. "The Concept of Economic Surplus." *Economic Journal* 81(324): 741–799.
Deaton, Angus, and John Muellbauer. 1980. *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.
Eckstein, Otto. 1961. "A Survey of the Theory of Public Expenditures Criteria." In National Bureau of Economic Research *Public Finances: Needs, Sources, and Utilization*. Princeton: Princeton University Press.
Freeman, A. Myrick, III. 1971. "Project Design and Evaluation with Multiple Objectives." In *Public Expenditures and Policy Analysis*, edited by R. Haveman and J. Margolis. Chicago: Markham.
Fudenberg, Drew, David K. Levine, and Zacharias Maniadis. 2012. "On the Robustness of Anchoring Effects in WTP and WTA Experiments." *American Economic Journal: Microeconomics* 42(2): 131–145.
Guzman, Rolando M., and Charles D. Kolstad. 2007. "Researching Preferences, Valuation and Hypothetical Bias." *Environmental and Resource Economics* 37(3): 465–487.
Hanemann, W. Michael. 1980. "Measuring the Worth of Natural Resource Facilities: Comment." *Land Economics* 56(4): 482–490.
Hanemann, W. Michael. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66(3): 332–341.
Hanemann, W. Michael. 1989. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data: Reply." *American Journal of Agricultural Economics* 71(4): 1057–1061.
Hanemann, W. Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 81(3): 635–647.
Hanemann, W. Michael. 1999. "The Economic Theory of WTP and WTA." In *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, edited by Ian J. Batemen and Kenneth G. Willis. Oxford: Oxford University Press.
Hausman, Jerry A. 1981. "Exact Consumer's Surplus and Dead Weight Loss." *American Economic Review* 71(4): 662–676.
Haveman, Robert H., and Burton A. Weisbrod. 1975. "The Concept of Benefits in Cost-Benefit Analysis: With Emphasis on Water Pollution Activities." In *Cost-Benefit Analysis and Water Pollution Policy*, edited by Henry M. Peskin and Eugene P. Seskin. Washington, D.C.: The Urban Institute.
Hicks, John R. 1939. "The Foundations of Welfare Economics." *Economic Journal* 49(196): 696–712.
Hicks, John R. 1943. "The Four Consumer Surpluses." *Review of Economic Studies* 11(1): 31–41.
Hoehn, John P., and Alan Randall. 1987. "A Satisfactory Benefit Cost Indicator from Contingent Valuation." *Journal of Environmental Economics and Management* 14(3): 226–247.
Horowitz, John K., and Kenneth E. McConnell. 2002. "A Review of WTA/WTP Studies." *Journal of Environmental Economics and Management* 44(3): 426–447.
Hurwicz, Leonid, and Hirofumi Uzawa. 1971. "On the Integrability of Demand Functions." In *Preferences, Utility and Demand,* edited by John Chipman et al. New York: Harcourt Brace Jovanovich.

Johansson, Per-Olov. 1987. *The Economic Theory and Measurement of Environmental Benefits*. Cambridge: Cambridge University Press.

Johansson, Per-Olov, Bengt Kriström, and Karl-Göran Mäler. 1989. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data: Comment." *American Journal of Agricultural Economics* 71(4): 1054–1056.

Just, Richard E., Darrell L. Hueth, and Andrew Schmitz. 1982. *Applied Welfare Economics and Public Policy*. Englewood Cliffs: Prentice-Hall.

Just, Richard E., Darrell L. Hueth, and Andrew Schmitz. 2004. *The Welfare Economics of Public Policy: A Practical Approach to Project and Policy Evaluation*. Cheltenham: Edward Elgar.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decisions Under Risk." *Econometrics* 47(1): 263–291.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* 98(6): 1325–1347.

Kaldor, Nicholas. 1939. "Welfare Propositions of Economics and Interpersonal Comparisons of Utility." *Economic Journal* 49(196): 549–552.

Knetsch, Jack L., and John A. Sinden. 1984. "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Values." *Quarterly Journal of Economics* 99(3): 507–521.

Kolstad, Charles D., and Rolando M. Guzman. 1999. "Information and the Divergence Between Willingness to Accept and Willingness to Pay." *Journal of Environmental Economics and Management* 38(1): 66–80.

Krutilla, John V. 1967. "Conservation Reconsidered." *American Economic Review* 57(4): 777–786.

List, John A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118: 41–71.

List, John A. 2004. "Substitutability, Experience, and the Value Disparity: Evidence from the Marketplace." *Journal of Environmental Economics and Management* 47: 486–509.

Little, I.M.D. 1957. *A Critique of Welfare Economics*. 2d ed. Oxford: Clarendon Press.

Marschak, Jacob. 1960 "Binary Choice Constraint on Random Utility Indications." In *Stanford Symposium on Mathematical Methods in the Social Sciences*, 312–319, edited by K. Arrow. Stanford: Stanford University Press.

Marshall, Alfred. 1920. *Principles of Economics: An Introductory Volume*. 8th ed. London: Macmillan.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Discrete Choice Behavior." In *Frontiers of Econometrics*, edited by Paul Zarembka. New York: Academic Press.

McFadden, Daniel. 1978. "Modeling the Choice of Residential Location." In *Spatial Interaction and Planning Models*, 75–96, edited by A. Karlqvist. Amsterdam: North Holland.

McFadden, Daniel. 1981. "Econometric Models of Probabilistic Choice Models." In *Structural Analysis of Discrete Choice Data with Applications*, edited by Charles Manski and Daniel McFadden. Cambridge: MIT Press.

McKenzie, George W. 1983. *Measuring Economic Welfare: New Methods*. Cambridge: Cambridge University Press.

McKenzie, George W., and I.F. Pearce. 1982. "Welfare Measurement: A Synthesis." *American Economic Review* 72(4): 669–682.

Mishan, Ezra J. 1960. "A Survey of Welfare Economics, 1939–1959." *Economic Journal* 70(278): 197–256.

Mishan, Ezra J. 1976. "The Use of Compensating and Equivalent Variations in Cost-Benefit Analysis." *Economica* 43(170): 185–197.

Morey, Edward R. 1984. "Confuser Surplus." *American Economic Review* 74(1): 163–173.

Mullainathan, S., and R.H. Thaler. 2001. "Behavioral Economics." In *International Encyclopedia of the Social and Behavioral Sciences, vol. 2*, edited by Neil J. Smelser and Paul B. Baltes. Amsterdam: Elsevier Press.

Neary, J.P., and K.W.S. Roberts. 1980. "The Theory of Household Behavior Under Rationing." *European Economic Review* 13(2): 25–42.

Plott, Charles R. and Kathryn Zeiler. 2005. "The Willingness to Pay–Willingness to Accept Gap, the 'Endowment Effect,' Subject Misconceptions, and Experimental Procedures for Elicity Valuations." *American Economic Review* 95(3): 530–545.

Polinsky, A. Mitchell. 1972. "Probabilistic Compensation Criteria." *Quarterly Journal of Economics* 86(3): 407–425.

Samuelson, Paul A. 1942. "Constancy of the Marginal Utility of Income." In *Studies in Mathematical Economics and Econometrics, in Memory of Henry Schultz*, edited by Oscar Lange, Francis McIntyre, and Theodore O. Yntema. Chicago: University of Chicago Press. Reprinted in *The Collected Scientific Papers of Paul A. Samuelson, vol. 1*, edited by Joseph E. Stiglitz. Cambridge: MIT Press, 1966.

Sayman, Sarder, and Ayse Onculer. 2005. "Effects of Study Design Characteristics on the WTA–WTP Disparity: A Meta Analytical Framework." *Journal of Economic Psychology* 26(2): 289–312.

Shogren, Jason, and Laura Taylor. 2008. "On Behavioral Environmental Economics." *Review of Environmental Economics and Policy* 2(1): 26–42.

Silberberg, Eugene. 1972. "Duality and the Many Consumer's Surpluses." *American Economic Review* 62(5): 942–952.

Silberberg, Eugene. 1978. *The Structure of Economics: A Mathematical Analysis*. New York: McGraw-Hill.

Small, Kenneth A., and Harvey S. Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." *Econometrics* 49(1): 105–130.

Thaler, Richard. 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization* 1(1): 39–60.

Thurstone, Louis. 1927. "A Law of Comparative Judgement." *Psychological Review*. 34: 273–286.

Tobin, James, and Hendrik S. Houthakker. 1950/1. "The Effects of Rationing on Demand Elasticities." *Review of Economic Studies* 18(3): 140–153.

Train, Kenneth. 2009. *Discrete Choice Models with Simulation.* 2d ed. Cambridge: Cambridge University Press.

Varian, Hal R. 1992. *Microeconomic Analysis*. 3d ed. New York: Norton.

Willig, Robert D. 1976. "Consumer's Surplus Without Apology." *American Economic Review* 66(4): 589–597.

Willig, Robert D. 1979. "Consumer's Surplus Without Apology: Reply." *American Economic Review* 69(3): 471–474.

Zhao, Jinhua, and Catherine L. Kling. 2001. "A New Explanation for the WTP/WTA Disparity." *Economic Letters* 73(3): 293–300.

Zhao, Jinhua, and Catherine L. Kling. 2004. "Willingness-to-Pay, Compensating Variation, and the Cost of Commitment." *Economic Inquiry* 42(2004): 503–517.

Chapter 4

Welfare Measures

Theoretical Basis for Empirical Assessment

For market goods, welfare effects due to changes in prices have been defined in terms of the area under the appropriate Hicks-compensated demand curve. For nonmarket goods, welfare effects due to changes in quantities have been defined in terms of the area under the marginal willingness-to-pay curve for the good or service. The marginal willingness-to-pay curves exist for public goods and nonmarketed goods such as the services of the environment; but they cannot be estimated from direct observations of transactions in these goods. Given the absence of markets for public goods and environmental goods, how can information on demand and benefits be obtained?

As described in Chapter 2, there are basically two approaches to obtaining demand and value information for changes in the quantities of nonmarket goods. They are the revealed preference methods that involve the estimation of value from observations of behavior in the markets for related goods and the stated preference methods for deriving values from responses to hypothetical questions. This chapter explores some of the possible relationships between demands for private goods and demands for environmental services in an effort to determine under what circumstances the demands for environmental services can be inferred from information on market transactions for a related private good. Let q denote some measure of environmental or resource quality. The task is to estimate in monetary terms the changes in individuals' welfares associated with changes in q.

The basic thesis of this chapter is that the degree to which inferences about the benefits of increases in q can be drawn from market observations and the appropriate techniques to be used in drawing these inferences, both depend upon the way in which q enters individual utility functions. Broadly speaking, there are three ways that q can affect an individual's utility: (a) q can produce utility indirectly as a factor input in the production of a marketed good that yields utility; (b) q can be an input in the household production of utility-yielding commodities; or (c) q can produce utility directly by being an argument in an individual's utility function. In the third case, there are a variety of ways in which q can interact with one or more market goods in the individual's preference structure. For example, there may be a substitution or complementary relationship between q and some private good. If the nature of the household production process or the forms of interaction between q and private goods can be specified, it may be possible to infer the value of q to the individual from observations of choices of the related market goods.

The next section briefly explores the case where q is a factor of production for a market good. The bulk of the chapter, however, is devoted to examining models of the ways in which q can affect utility more directly. The exploitation of possible relationships between environmental goods and private goods leads to several empirical techniques for estimating environmental and resource values. These techniques have the following characteristics: (a) they are consistent with the basic theory of demand and consumer preferences; (b) they provide a means for estimating the indirect utility function, the expenditure function, or the compensated demand function for the environmental service; and (c) they are practical in the sense of imposing realistic data and computational requirements. The chapter concludes by examining those sources of value that are potentially missed by relying on revealed preference data alone, broadly classified as "nonuse" or "passive use" values.

Environmental Quality as a Factor Input

When q is a factor of production, changes in q lead to changes in production costs, which in turn affect the price and quantity of output or the returns to other factor inputs, or both. The benefits of changes in q can be inferred from these changes in observable market data. There are several examples where q can be interpreted as a factor input. The quality of river water diverted for irrigation affects the agricultural productivity of irrigated land. The quality of intake water may influence the costs of treating domestic water supplies and the costs of production in industrial operations that utilize water for processing purposes. Agricultural productivity is impaired by some forms of air pollution, and to the extent that air pollution causes materials damages, it can affect the costs of production for a wide variety of goods and services.

Assume that good x is produced with a production function,

$$x = x(k, l, q), \tag{4.1}$$

where k and l are capital and labor, respectively, and where the marginal product of q is positive. With given factor prices, and assuming cost-minimizing behavior, there is a cost function:

$$C = C(p_k, p_w, q, x). \tag{4.2}$$

Since q affects the production and supply of a marketed good, the benefits of changes in q can be defined and measured in terms of changes in market variables related to the x industry. A change in q will cause shifts in both cost curves and factor demand curves. The consequences of these shifts depend on conditions in factor and product markets. Changes in q can produce benefits through two channels. The first is through changes in the price of x to consumers. The second is through changes in the incomes and profits received by owners of factor inputs used in x production.

To illustrate the first channel, assume that x is produced in a competitive industry under conditions of constant cost—that is, factor supplies to this industry are infinitely elastic. Assume that the change in q affects the cost curves of a significant proportion of producers in the market. As a result, the supply curve shifts downward, causing a fall in the price and an increase in total quantity. The benefit of the price reduction accrues to consumers and can be measured by the methods described in Chapter 3.

To illustrate the second channel—changes in the incomes received by factors of production—consider only one producer who is a price taker in all markets. If the change in q affects only this producer, output price will not be changed. Since the change in q affects the marginal costs of production, the firm's marginal cost and supply curves are shifted down. In this case, the benefit is equal to the increase in quasi-rents to the firm. This benefit will accrue to the owner of a fixed factor—land, for example—or to the residual income claimant as profit. In either case, benefits can be measured by changes in profits and fixed factor incomes. However, if the producers affected by changes in q face less than perfectly elastic factor supply curves, at least some of the benefits will be passed on to factors through changes in factor prices and incomes. The factors' shares of benefits can be approximated by the areas to the left of factor supply curves.

The effects of these two channels are combined in Figure 4.1. When the supply curve of the industry is shifted down to S'', the price decreases to p''. The benefit to consumers of x is approximated by the change in consumers' surpluses, the area $p'BCp''$. Part of this benefit, $p'BFp''$, is at the expense of a reduction in producer and factor surpluses, so the net gain from the lower price is BCF. The lower supply curve results in factor surpluses and quasi-rents equal to $p''CE$. The net increase to producers and factors is $AFCE$, so total benefits are equal to $ABCE$.

Implementation of these measures requires knowledge of the effects of changes in q on the cost of production, the supply conditions for output, the demand curve for good x, and factor supplies. There are two special cases where the estimation of benefits is relatively straightforward.

The first is the case where q is a perfect substitute for other inputs in the production of a good. An increase in q leads to a reduction in factor input costs. If the substitution relationship is known, the decrease in per unit production costs is readily calculated. For example, if water quality improvement results in a decrease in chlorination requirements for drinking water supplies, the decrease in chlorination costs per unit of output can be readily calculated. Where the change in total cost does not affect marginal cost and output, the cost saving is a true measure of the benefit of the change in q. If the change in q affects marginal cost, the benefits should include the effect of the lower cost on output and price. However, if the percentage reduction in marginal costs is small or the marginal cost curve is inelastic, or both, the corresponding increase in output would be relatively small. Thus, the decrease in total cost could still be used to provide a

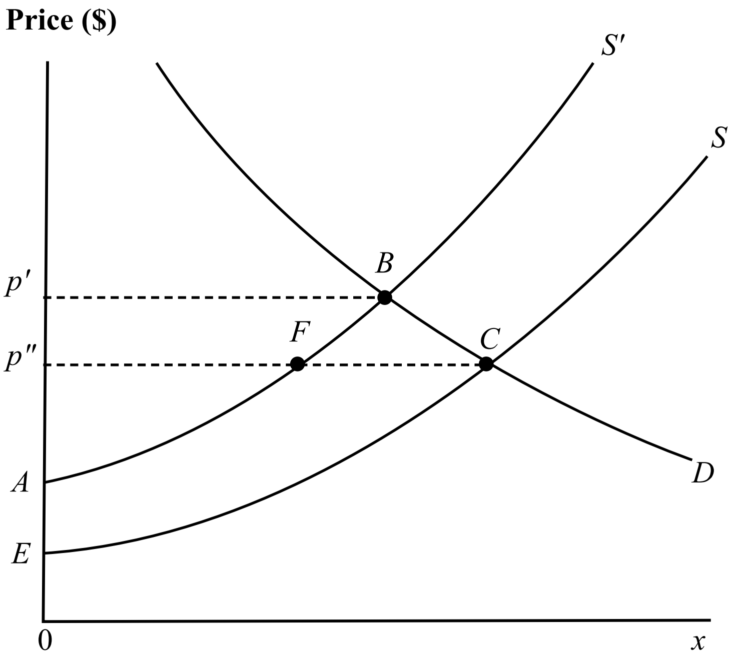

Figure 4.1 The welfare measure when *q* affects the production of *x*

rough approximation of true benefits. This approach, sometimes referred to as the "damage function" approach, has been the basis of a number of estimates of the materials, household cleaning, and agricultural crop-loss benefits of air pollution control, and of the benefits to municipalities, industries, and households of reduced contamination of intake water supplies.

The second case that makes the estimation of benefits relatively straightforward is where knowledge of cost, demand, and market structure suggests that the benefits of a change in q will accrue to producers. Then benefits may be estimated from observed or predicted changes in the net income of factor inputs. If the production unit in question is small relative to the market for the final product and for variable factors, it can be assumed that product and variable factor prices will remain fixed after the change in q. The increased productivity then accrues to the fixed factors of production in the form of profit or quasi-rent.

More generally, however, estimates of the value of q require knowledge of the cost and demand functions. In some studies, it has been possible to use econometric methods to estimate a cost function that includes an environmental quality variable (for example, Mjelde et al. 1984; Garcia et al. 1986; Neeliah and Shankar 2010). Other studies have used various simulation approaches to model the behavior of producers and their responses to changes in an environmental variable. Models and techniques for valuing the effects of q on production are discussed in more detail in Chapter 8.

An Individual's Demand for Environmental Quality

In order to analyze those cases where q affects individuals directly, the basic model of individual preference and demand, with environmental quality included as an argument in the utility function, must first be reviewed. The implications of different forms of utility functions for estimation of the demand for q can then be examined.

Consider a single individual who has a utility function

$$u = u(\boldsymbol{X}, q), \tag{4.3}$$

where $\boldsymbol{X} = \left(x_1, \ldots, x_J\right)$ is a vector of private goods quantities. In entering environmental quality as an argument in the utility function, it is assumed here that the individual perceives the effects of changes in environmental quality. For example, if high ozone levels cause respiratory irritation, the individual is assumed to be aware of the irritation, so that he feels "better" when it is reduced. He need not know the cause of the irritation or the actual levels of air pollution. If the individual is not aware of the effects of changes in q, the revealed preferences methods of benefit estimation cannot be applied. For example, individuals may not perceive the effects of long-term exposure to air pollutants on their probability of chronic illness or death. If that is the case, changes in q will not affect their behavior and observations of market behavior will yield no information about the value of reducing risks to health.

Assume that the individual maximizes utility subject to a budget constraint

$$\sum_j p_j \cdot x_j \leq M, \tag{4.4}$$

where M is money income. The individual takes q as given and does not have to pay a price for this freely provided quantity. The solution to this problem yields a set of ordinary demand functions

$$x_j = x_j(\boldsymbol{P}, M, q), \tag{4.5}$$

where $\boldsymbol{P} = \left(p_1, \ldots, p_J\right)$ is the vector of private good prices. Note that in general q could be an argument in all private good demand functions.

The dual to the utility maximization problem can be stated as follows: minimize expenditure ($\sum_{j=1}^{J} p_j x_j$) subject to the constraint that utility equals or exceeds some stated level, say u^0. The solution to this problem gives the expenditure function

$$e(\boldsymbol{P}, q, u^0) = M. \tag{4.6}$$

The expenditure function has a number of useful properties for applied welfare analysis. First, as shown in Chapter 3, the derivative of the expenditure function with respect to any price gives the Hicks-compensated demand function for that good—that is,

$$\frac{\partial e}{\partial p_j} = h_j\left(\boldsymbol{P}, q, u^0\right). \tag{4.7}$$

Similarly, the derivative of equation (4.6) with respect to q (with the appropriate change of sign) gives the Hicks-compensated inverse demand function, or marginal willingness to pay for changes, in q. Let w_q be the marginal willingness to pay or marginal demand price for q, then

$$w_q = -\frac{\partial e\left(\boldsymbol{P}, q, u^0\right)}{\partial q}. \tag{4.8}$$

An alternative expression for the marginal willingness to pay can be obtained by setting the total differential of the indirect utility function equal to zero and solving for the compensating change in income associated with the change in q. Specifically,

$$v = v(\boldsymbol{P}, M, q), \tag{4.9}$$

$$dv = \frac{\partial v}{\partial M} dM + \frac{\partial v}{\partial q} dq = 0, \tag{4.10}$$

and

$$\frac{dM}{dq} = -\frac{\partial v / \partial q}{\partial v / \partial M}, \tag{4.11}$$

where $d\boldsymbol{P}$ is zero by assumption.

If the value of the derivative of the right-hand side of equation (4.8) can be inferred from observed data, then we have a point estimate of the marginal willingness to pay for q. If this derivative can be estimated as a function of q, then we have the marginal willingness to pay function for q. Let W_q represent the benefit to the individual of a nonmarginal increase in the supply of q. W_q is the integral of the marginal willingness-to-pay function, or

$$\begin{aligned} W_q &= -\int_{q'}^{q''} e_q\left(\boldsymbol{P}, q, u\right) dq \\ &= e\left(\boldsymbol{P}, q', u\right) - e\left(\boldsymbol{P}, q'', u\right). \end{aligned} \tag{4.12}$$

This is either a compensating surplus (*CS*) or an equivalent surplus (*ES*) measure of welfare change, depending on the level of utility at which equation (4.12) is evaluated. The question to be discussed in the next section is whether there are any circumstances in which information about equations (4.8) or (4.12) can be derived from observations of market prices and quantities for private goods.

The Structure of Preferences and Measures of Value

The main purpose of this section is to describe the available techniques for revealing these welfare measures, or approximations of them, using observable data on related behavior and individual choice. The strategy will be to explore credible a priori assumptions that support restrictions on the form of the utility function and/or demand functions for market goods or household produced goods that, in turn, aid in revealing the individual's preferences for environmental quality.

Different types of restrictions have different implications for the measurability of the demand for environmental quality. Developing a careful taxonomy of the methods for teasing out the welfare effects of interest helps reveal the broader and more general basis for the welfare economics of environmental valuation.

There are several ways additional structure can be given to the general model of preferences and choice discussed above. Each of these alternatives involves some kind of assumption about the structure of preferences and/or the constraints on individual choice. Each of the assumptions implies some kind of connection between observable demands for market goods and the values of environmental services and public goods; and each assumption provides a basis for inferring the marginal willingness to pay for q from observations of the relationships between q and the demands for market goods. It is also important to examine what is required to obtain values of *nonmarginal* changes in q, since these are what are required for most real-world policy questions. For each of the alternatives, the specific restrictions are identified, along with a discussion as to how inferences about value can be drawn from observations of individual choices.

The relationships between q and other goods that have been found to be of use involve, broadly speaking, either substitution or complementarity relationships between q and other goods. Exactly how these relationships work out methodologically, however, depends on a number of other considerations. In what follows, consideration is first given to those cases in which the environmental good is a substitute for a marketed good that enters the utility function. A fundamentally equivalent construct is one in which the environmental good is an input into a household production function and has marketed-good substitutes in the production process. The latter is perhaps the more general and more useful way of conceptualizing the problem.

Attention is then given to a second category of models, namely those in which the environmental good is in some way complementary to another good. The complementarity is often most usefully conceived such that the environmental good is a quality characteristic of the related good. There are two often-used derivatives of this construct. In one, the related good is itself a nonmarket good produced by the household using a household production process. The second is one in which the related good is marketed, but units of the good are heterogeneous and quality-differentiated. Because the good is marketed, the prices of units with higher levels of quality embodied in them are bid up.

Several of the alternatives involve making some kind of assumption about the separability of the utility function. Thus, it will be useful to first review the concept of separability and the implications of various forms of separability for observable demands for market goods. For a review of concepts of separability, see Goldman and Uzawa (1964), Katzner (1970), Blackorby, Primont, and Russell (1978), or Deaton and Muellbauer (1980).

Separability refers to the possible effect of partitioning the goods entering into the utility function into subsets, and the relationships among these subsets. Suppose that there are three types of market goods, so that

$$u = u(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}), \tag{4.13}$$

where $\boldsymbol{X} = (x_1, \ldots, x_J)$, $\boldsymbol{Y} = (y_1, \ldots, y_K)$, and $\boldsymbol{Z} = (z_1, \ldots, z_S)$. A utility function is (*directly*) *weakly separable* if the marginal rate of substitution (MRS) between any pair of goods within the same subset is independent of the quantities of goods in any other subset. The following utility function is weakly separable:

$$u = u[u^1(\boldsymbol{X}), u^2(\boldsymbol{Y}), u^3(\boldsymbol{Z})]. \tag{4.14}$$

This means that the demand functions for goods in one subset can be written as depending only on the prices of goods in that subset and the expenditure share of that subset.

A utility function is strongly *separable* if the MRS between two goods in different subsets is independent of the quantity of any good in any other subset. Specifically, with strong separability the MRS_{x_j, y_k} is independent of z_s. Any additive utility function is strongly separable; and any strongly separable utility function is additive in some monotone transform (Deaton and Muellbauer 1980). The following utility function is strongly separable:

$$u = u[u^1(\boldsymbol{X}) + u^2(\boldsymbol{Y}) + u^3(\boldsymbol{Z})]. \tag{4.15}$$

If the utility function is not separable, there are no restrictions on the terms that are arguments in marginal rates of substitution, or in the demand functions for individual goods. The MRS between any pair of goods depends on the quantities of all goods; and the demand function for any good depends upon the prices of all goods as well as income.

In the rest of this section, three major forms of relationships between q and market goods are examined. These major forms are: (a) some form of substitution relationship between q and one or more market goods; (b) some form of complementary relationship; and (c) the case of differentiated goods where the amount of q embodied in, or attached to, a market good is one of its differentiating characteristics. Since some of these relationships can be modeled either implicitly or explicitly as involving some form of household production, a brief overview of the household production framework for modeling individual preferences and choice is first provided.

The Household Production Framework

The household production function model provides a framework for examining interactions between demands for market goods and the availability of a public good such as environmental quality. This framework is based on the assumption that there is a set of technical relationships among goods used by households in the implicit production of utility-yielding final services. In the household production function literature, the utility-yielding final services are often termed "commodities," while market goods are simply "goods." The terminology used in this chapter is more descriptive of the relationship between goods being bought in the market and the final service flows that yield utility to individuals.

In the household production framework, utility is a function of the level of final service flows:

$$u = u(\boldsymbol{Z}). \tag{4.16}$$

$\boldsymbol{Z} = (z_1, \ldots, z_S)$ is produced according to a technology common to all households and assumed to be known; and

$$z_s = z_s(\boldsymbol{X}, \boldsymbol{Q}), \text{ for } s = 1, \ldots, S, \tag{4.17}$$

where $\boldsymbol{X} = (x_1, \ldots, x_J)$ represents a vector of market goods available at prices $\boldsymbol{P}_x = (p_{x1}, \ldots, p_{xJ})$, and $\boldsymbol{Q}$ represents a vector of environmental quality attributes. The cost of household production depends on the technology, input prices, and exogenously determined $\boldsymbol{Q}$. That is, $C(\boldsymbol{Z}) = C(\boldsymbol{Z}, \boldsymbol{P}_x, \boldsymbol{Q})$.

Formally, the individual choice problem is to maximize equation (4.16) subject to the constraints provided by equation (4.17), the given level of $\boldsymbol{Q}$, and the budget constraint $\sum_{j=1}^{J} p_{xj} x_j \leq M$. This problem can be solved by a two-step procedure in which the first step is to combine market goods and $\boldsymbol{Q}$ so as to minimize the costs of producing the z_s. This determines the marginal costs of final services $\boldsymbol{C}_z$. These are essentially implicit prices, but unlike market prices, they are not parametric to the individual unless marginal costs are constant. The second step is to maximize (4.16) subject to the budget constraint $C(\boldsymbol{Z}) \leq M$. The observable manifestation of the solution to this problem is a set of derived market goods demands:

$$\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{P}_x, M, \boldsymbol{Q}). \tag{4.18}$$

These demand functions will reflect both the role of $\boldsymbol{Q}$ in the household production technology equation (4.17) and the preferences over the final service flows. As will be seen, some of the existing models for interpreting demands for market goods (for example, the averting behavior/defensive expenditures model and the weak complementarity model) can be interpreted as simpler versions of a household production function model. In fact, as Smith (1991) has argued, the household production framework may be a useful way of thinking about whether a specific market good is likely to be a substitute or a complement for environmental quality. He used the household production framework as an organizing principle to examine a number of models for nonmarket valuation and to interpret several empirical applications of these models. As various forms of complementarity and substitutability relationships are considered, it will be shown how they can also be interpreted as applications of the household production function model.

Substitutes

Mäler (1974, 116–118) has shown that the marginal willingness to pay for q can be expressed in terms of the price of any private good and the marginal utilities of that good and q. The expression is

$$w_q = -\frac{\partial e(\boldsymbol{P}, q, u^0)}{\partial q} = -p_i \left(\frac{\partial u / \partial q}{\partial u / \partial x_i} \right) = p_i \cdot MRS_{qx_i}. \tag{4.19}$$

This would be a useful practical result if it were possible to derive simple expressions for the marginal rates of substitution. Of course, if q could be purchased at a given price the marginal rate of substitution between q and x_i could be inferred from the price ratio. However, when the level of q is determined exogenously, the marginal rate of substitution can only be determined through knowledge of the utility function or household production function.

Consider the case where some assumption about the separability of the utility function is used to isolate q and a good, z, from the other determinants of utility (see Mäler 1974, 178–183). Suppose the utility function is weakly separable and is of the following form:

$$u = u\left\{ u^1(\boldsymbol{X}), u^2(\boldsymbol{Y}), \left[c \cdot z^{-\alpha} + (1-c) q^{-\alpha} \right]^{\frac{-1}{\alpha}} \right\}. \tag{4.20}$$

Given the separability assumption, the marginal rate of substitution between z and q is independent of the quantities of $\boldsymbol{X}$ and $\boldsymbol{Y}$. In this case, the marginal rate of substitution is

$$MRS_{zq} = \frac{c}{1-c} \left(\frac{q}{z} \right)^{\alpha - 1} = \frac{c}{1-c} \left(\frac{q}{z} \right)^{1/\sigma}, \tag{4.21}$$

where σ is the elasticity of substitution, which is constant. From (4.19) we have

$$w_q = -p_z \cdot \left(\frac{c}{1-c} \right) \left(\frac{q}{z} \right)^{1/\sigma}. \tag{4.22}$$

Since, in general, the marginal rate of substitution depends on the ratio q/z, we need to know both the elasticity of substitution, σ, and c in order to compute w_q. This requires knowledge of the utility function.

Perfect Substitutes

There is one special case where the expression for w_q reduces to a usable term. If z and q are perfect substitutes in consumption, the elasticity of substitution between them is infinite, and the expression for the demand price for q reduces to $p_z \cdot s$, where s is the substitution ratio between z and q [$s = c/(1 - c)$]. If perfect substitutability can be assumed, s (or c) should be computable from known or observable technical consumption data or from the household production function.

The perfect substitutability assumption lies behind the simplest application of the defensive expenditures technique for estimating the benefits of pollution control. Defensive expenditures are made either to prevent or to counteract the adverse effect of pollution. They are also referred to as averting expenditures (for example see Courant and Porter 1981). In effect, a defensive expenditure is spending on a good that is a substitute for higher q. An increase in q is assumed to lead to a decrease in spending on the substitute.

As equation (4.19) shows, the marginal change in the spending on z is the correct measure of the marginal willingness to pay for the change in q; and for nonmarginal changes in q, the benefit is $p_z \cdot \Delta q$. However, this will not necessarily be the same as the observed change in spending on q. The intuition behind this statement is straightforward. The benefit of a nonmarginal change in q is the reduction in the spending on z that is required to keep the individual on the original indifference curve. In general, the individual will not actually reduce spending on z by this amount. There is an income effect on z as well as a substitution effect. The increase in q means that the same level of utility can be maintained with a smaller expenditure on z, and consequentially, the individual will reallocate expenditure among all goods, including z, so as to maximize the increase in total utility. This will result in increases in the expenditures on all goods with positive income elasticities of demand. Hence, the observed decrease in spending on z will be less than necessary to hold utility constant, and the reduction in defensive spending will be an underestimate of the benefits of higher q. In fact, as we will see below, actual spending on z could increase.

A General Model of Substitution

Since perfect substitutes represent a special case, it will be useful to explore a more general model in which q and a market good are less than perfect substitutes, and in which the substitution relationship arises because q and the market good contribute to utility through the same mechanism. Following the analysis of Courant and Porter (1981), this case is modeled in the household production framework. Early analyses dealing with this type of model are presented in Shibata and Winrich (1983), Harford (1984), and Harrington and Portney (1987) and more recent examples include Neidell (2009) and Zivin, Neidell, and Schlenker (2011). Suppose that clean air and soap are substitutes in the production of cleanliness. Let

$$u = u(\boldsymbol{X}, z), \tag{4.23}$$

where z is cleanliness, and $\boldsymbol{X}$ is a vector of market goods with prices $\boldsymbol{P}_x = \left(p_{x1}, \ldots, p_{xJ}\right)$ Suppose that z is produced by households by combining a market good, y, and air quality according to the production function

$$z = z(y, q), \tag{4.24}$$

with positive partial derivatives for both arguments. Further, assume that y contributes to utility only through its contribution to cleanliness. This is equivalent to assuming that the utility function is weakly separable, with y and q in one group.

For any given q, the individual chooses $\boldsymbol{X}$ and y so as to maximize

$$u = u[\boldsymbol{X}, z(y, q)], \tag{4.25}$$

subject to the budget constraint

$$M - \sum_{j=1}^{J} p_{xj} x_j - p_y y \geq 0, \tag{4.26}$$

the household production technology and the exogenous q. The first-order conditions are

$$\begin{aligned} &\frac{\partial u}{\partial \boldsymbol{X}} - \lambda \boldsymbol{P}_x = 0 \\ &\frac{\partial u}{\partial z} \cdot \frac{\partial z}{\partial y} - \lambda \cdot p_y = 0. \end{aligned} \tag{4.27}$$

Substituting the expressions for the demands for $\boldsymbol{X}$ and y into the utility function yields the indirect utility function

$$v = v(\boldsymbol{P}_x, p_y, M, q). \tag{4.28}$$

Setting the total differential of this expression equal to zero, assuming prices do not change, and rearranging terms results in the following expression for the marginal value of increasing q:

$$\mathrm{w}_q = -\frac{dM}{dq} = \frac{(\partial v/\partial q)}{\lambda} = \frac{\left(\dfrac{\partial u}{\partial z} \cdot \dfrac{\partial z}{\partial q}\right)}{\lambda}, \tag{4.29}$$

where λ is the marginal utility of income, and the term in brackets is an expansion of $\partial v/\partial q$ making use of the chain rule. Substituting the first-order condition (4.27) gives the marginal willingness to pay for q as a function of the price of the private good and the marginal productivities of y and q:

$$w_q = p_y \cdot \frac{\partial z / \partial q}{\partial z / \partial y}. \tag{4.30}$$

This is the reduction in spending on y, holding z constant—that is, moving along a given isoquant in the household production function. This marginal value can be calculated if the household production function is known.

If y and q jointly produce two goods (or help to avoid two bads), this result does not hold. Suppose $u = u\left[\boldsymbol{X}, z_1(y,q), z_2(y,q)\right]$. Using the procedure described above, we obtain

$$w_q = p_y \left[\frac{(\partial u / \partial z_1)\cdot(\partial z_1 / \partial q) + (\partial u / \partial z_2)\cdot(\partial z_2 / \partial q)}{(\partial u / \partial z_1)\cdot(\partial z_1 / \partial y) + (\partial u / \partial z_2)\cdot(\partial z_2 / \partial y)}\right]. \tag{4.31}$$

The unobservable utility terms do not cancel out. The relationship between the actual change in spending on y and marginal willingness to pay can be obtained by totally differentiating the household production function with respect to q and using the first-order conditions for the optimum choice of y to obtain

$$\frac{dz}{dq} = \frac{\partial z}{\partial y} \cdot \frac{\partial y^*}{\partial q} + \frac{\partial z}{\partial q}, \tag{4.32}$$

where $y^*\left(\boldsymbol{P}_x, p_y, M, q\right)$ is the derived demand function. Rearranging and employing the second line of equation (4.27) yields

$$p_y \cdot \frac{\partial z / \partial q}{\partial z / \partial y} = w_q = \frac{\partial u / \partial z}{\lambda} \cdot \frac{dz}{dq} - p_y \cdot \frac{\partial y^*}{\partial q}. \tag{4.33}$$

If spending on y decreases ($\partial y^*/\partial q < 0$), then the marginal willingness to pay for q is the sum of the observed reduction in spending and the willingness to pay for the induced increase in the final service flow z. The actual savings in spending on y is an underestimate of the marginal value in this case; and if the lower implicit cost of household production of z leads to a sufficiently large increase in the consumption of z, then it is possible that spending on y could increase. Thus, observed changes in spending on substitute goods are not reliable indicators of marginal willingness to pay for changes in q.

Similar results are obtained for the case where z requires several private goods as well as q for its production. The marginal value of a change in q can be measured by the reduction in spending on any of the market goods for movements along the original z isoquant. Observed changes in spending on market goods are not reliable measures of w_q.

An alternative formulation is to let z represent a bad, with y and q mitigating the adverse effects of z. In other words, $u_z < 0$, $z_q < 0$, and $z_y < 0$, where subscripts indicate partial derivatives. Similar conclusions can be reached about willingness to pay and its relationship to changes in mitigating expenditures. This alternative model is typically referred to as an *averting behavior model* or a *mitigating behavior model.*

Bartik (1988) has shown how lower and upper bounds on the welfare gains of nonmarginal changes in q can be calculated from knowledge of only the household production function. The lower bound on the CS measure of benefits of an increase in q is the reduction in the expenditures on y necessary to reach the initial level of z, other things being equal. More formally, the lower bound is given by

$$\text{Lower bound} = p_y\left(y' - y''\right). \tag{4.34}$$

This can be shown to be a lower bound on CS by defining a restricted expenditure function, $e^*(q, z', u^0)$, where z' is the initial level of z, and price terms have been omitted for simplicity. The welfare gain associated with an increase in q from q' to q'' holding z at z', is

$$CS^* = e^*\left(q', z', u^0\right) - e^*\left(q'', z', u^0\right). \tag{4.35}$$

The first term on the right-hand side is $M = \sum_{j=1}^{J} p_{xj} x_j' + p_y y'$. The increase in q means that less must be spent on y to produce z', and since $u(\boldsymbol{X}', z') = u^0$, the second term is $\sum_{j=1}^{J} p_{xj} x_j' + p_y y''$. So

$$CS^* = \sum_{j=1}^{J} p_{xj} x_j' + p_y y' - \sum_{j=1}^{J} p_{xj} x_j' - p_y y'' = p_y \left(y' - y'' \right). \tag{4.36}$$

This has to be less than the true CS, since relaxing the constraint holding z at z' would allow the individual to increase utility, unless the marginal utility of z were already zero. In a similar fashion, Bartik showed that an upper bound on the equivalent surplus measure of welfare gain is the decrease in spending on y that is possible while holding z constant at its new equilibrium level; and since, in general, $ES > CS$, the constrained changes in expenditure bound both of the true measures of welfare change.

We can also consider the case where q contributes to utility directly in addition to its contribution to z. Taking the total differential of the indirect utility function, solving for the compensating change in income, and substituting in the first-order condition for the choice of y gives us

$$w_q = -\frac{dM}{dq} = \frac{\partial u / \partial q}{\lambda} + \left[p_y \cdot \frac{\partial z / \partial q}{\partial z / \partial y} \right]. \tag{4.37}$$

Comparing this expression with equation (4.29), we see that the marginal willingness to pay for q now includes an unobservable marginal utility term for the direct effect of q on utility. Similarly, following the same procedures used to derive equation (4.30), we have

$$w_q = \left[\frac{\partial u / \partial z}{\lambda} \cdot \frac{dz}{dq} \right] + \frac{\partial u / \partial q}{\lambda} - \left(p_y \cdot \frac{\partial y^*}{\partial q} \right). \tag{4.38}$$

Again, if spending on y decreases, it is an underestimate of marginal willingness to pay because it neglects both the utility value of the induced increase in z and the direct utility value of the increase in q.

Perfect Complements

Suppose that the single environmental service q is a perfect complement to a market good, say x_1, where perfect complementarity means that x_1 and q must be consumed in fixed proportions, for example $x_1/q = a$. This means that to utilize effectively the services of one unit of q, the individual must purchase $1/a$ units of x_1. In the context of the household production model, this could imply a production function for z_1 of the following form:

$$z_1 = \min\left(\frac{x_1}{a}, q \right). \tag{4.39}$$

If this is the case, there are conditions under which the benefits of changes in q can be estimated from knowledge of the demand function for the complementary market good. As long as p_1 is less than some critical value, say p_1^*, the quantity demanded of x_1 will be independent of its price and determined solely by the availability of q. If q is increased by one unit, the individual will purchase $1/a$ additional units of x_1 and experience a higher level of utility. Thus, the marginal willingness to pay for q will be positive. However, if p_1 is greater than p_1^*, the individual will purchase fewer units of x_1 than are required to fully utilize the available q. Thus, at p_1 greater than p_1^*, the marginal utility and marginal willingness to pay for q are zero.

Mäler (1974, 180–183) has shown that if the demand functions for x_1 and for other market goods are known, it is possible to compute the expenditure function and the demand price for q when p_1 is less than p_1^*. The exact expression for the demand price for q depends upon the specification of the true demand curves. There is no simple generalization of the technique. Also, as Mäler appeared to realize, it is difficult to imagine examples of perfect complementarity between an environmental service and a market good. But as we will now see, there is a less restrictive form of complementarity that also makes possible the calculation of the marginal willingness to pay for q and that appears to have real-world applications.

Weak Complementarity

Suppose that q enhances the enjoyment the individual derives from consuming x_1 and that an increase in q increases the quantity demanded of x_1, other things being equal. One reasonable interpretation of these assumptions is that q is an exogenously determined characteristic of x_1. Examples of environmental services that might fit these assumptions include water quality as a characteristic associated with visits to a lake, and the number of fish caught as a characteristic of fishing trips to a stream. The model can also be applied to estimate the value of a network where the purchase of a market good, a cell phone for example, is required to connect to the network (Hahn, Tetlock, and Burnett 2000).

In this section, it is shown that when the "enjoyment" of q requires the purchase of a market good, or when q can be treated as a characteristic of a market good, it is possible to identify a measure of the value of a change in q that is based on the demand for the market good. The measure is described, along with the conditions under which it is a valid welfare measure. The section closes with a discussion of some issues concerning implementation of the measure.

Suppose that x_1 and q are Hicksian complements—that is, the compensated demand function

$$h_1 = h_1(\boldsymbol{P}, q, u^0) \tag{4.40}$$

is characterized by $\partial h_1/\partial q > 0$. This implies that the Marshallian demand for x_1,

$$x_1 = x_1(\boldsymbol{P}, M, q), \tag{4.41}$$

has $\partial x_1/\partial q > 0$.

Suppose that the complete system of demand equations for this individual is known (or has been estimated econometrically), and that the system satisfies the Slutsky conditions for integrability. In order to derive a valid welfare measure from this system of demand functions, one must be able to integrate this system to solve for the underlying utility function and expenditure function. However, Mäler (1974, 183–189) has shown that, in general, it is not possible to solve completely for the utility and expenditure functions with the information given. Mathematically, the result of the integration contains unknown terms that are themselves functions of q and the constants of integration. It is necessary to impose additional conditions on the problem in order to solve for the unknown terms and determine the constants of integration. The additional conditions involve what Mäler called weak complementarity.

The Conditions

Weak complementarity requires that the marginal utility or marginal demand price of q be zero when the quantity demanded of the complementary private good x_1 is zero. Mathematically, weak complementarity involves two conditions. The first is that x_1 be nonessential in the sense that the compensating variation associated with the elimination of the good is finite—that is

$$\int_0^{\infty} h_1\left[p_1, q, u^0\right] dp_1 < \infty ,$$

where the terms for all other prices have been omitted for simplicity.

A sufficient condition for this to be the case is that there must exist a finite price $p_1^*(q)$ such that the compensated demand for the good is zero. In other words,

$$h_1\left[p_1^*(q), q, u^0\right] = 0. \tag{4.42}$$

This price is often called the *choke* price. The choke price will usually be an increasing function of q. On a graph, the compensated demand curve has a vertical intercept and there is some level of expenditure on other goods that will sustain u^0 even when x_1 is zero. For ease of exposition, the sufficient condition (a finite choke price) is used in the following discussion rather than the more general, but somewhat tedious, necessary condition.

The second condition is that for values of p_1 at or above $p_1^*(q)$, the derivative of the expenditure function

$$e = e\left(p_1, q, u^0\right) \tag{4.43}$$

must be zero. In other words,

$$\frac{\partial e\left(p_1, q, u^0\right)}{\partial q} = 0 \ \forall \ p_1 \geq p_1^*(q). \tag{4.44}$$

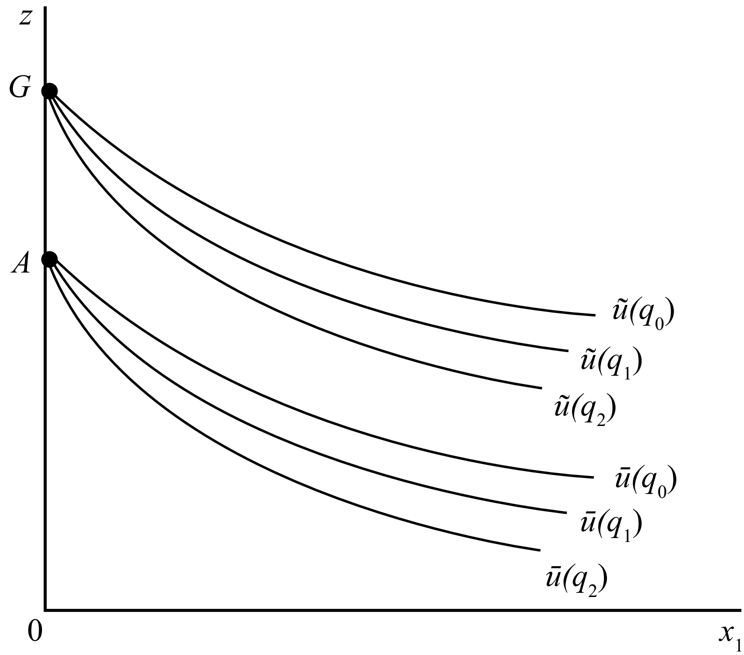

Figure 4.2 Fanned indifference curves

This condition means that at or above the choke price, the marginal utility or marginal willingness to pay for q is zero. Changes in q have no welfare significance unless the price of x_1 is low enough that its compensated demand would be positive. Weak complementarity can also be defined in terms of the underlying utility function, requiring that

$$\left.\frac{\partial u(\boldsymbol{X},q)}{\partial q}\right|_{x_1=0}=0\,. \tag{4.45}$$

Weak complementarity establishes an initial position for the individual that can be used to determine the constants of integration. For proof and a demonstration with a numerical example, see Mäler (1974, 183–189). Bradford and Hildebrandt (1977) also obtained results similar to Mäler. Smith and Banzhaf (2004) provided a graphical exposition of the weak complementarity assumption, replicated here as Figure 4.2. The graph provides an indifference map illustrating the tradeoffs between a market commodity (x_1) and a numeraire good (z) for varying levels of a public good (q). The lower group of indifference curves, $\overline{u}(q)$, depict the same level of utility given different levels of the public good, with $q_0 < q_1 < q_2$. So long as x_1 is consumed, the same level of utility can be reached with less of x_1 when $q = q_2$ than when $q = q_0$. Intuitively, the higher level of q enhances (i.e., complements) the experience associated with consuming x_1, so that, other things

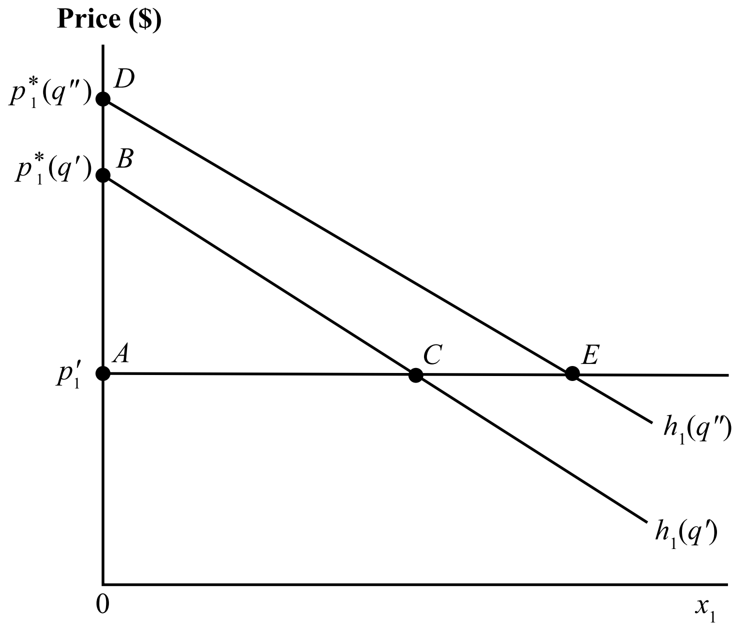

Figure 4.3 The welfare measure for an increase in q when q and x_1 are weak complements

equal, less of x_1 is needed to reach the same level of utility. However, when x_1 is not consumed, the level of q is irrelevant, causing all of the indifference curves $\overline{u}(q)$ to meet at point A. This creates what Smith and Banzhaf referred to as a "fanning" effect on the indifference curves. The second set of indifference curves, $\tilde{u}(q)$, represent a higher level of utility, but retain the fanning pattern in terms of the indifference relationship.

The Welfare Measure

Given the conditions for weak complementarity, the compensating surplus for a change in q can be measured by the area between the two compensated demand curves for x_1. The intuition for this measure can be illustrated graphically. Assume that the compensated demand curve for x_1 has been found for environmental quality level q'. In Figure 4.3, this demand curve is labeled $h_1(q')$. Assume that the price of x_1 is given at p_1' and does not change throughout the analysis. The compensating surplus associated with the use of x_1 is the area ABC under the demand curve. Now assume that quality is improved to q''. The increase in the quality associated with the use of x_1 is assumed to increase the demand for x_1, thus shifting the demand curve outward to $h_1(q'')$. The calculation of the benefit associated with this change is straightforward and can be divided into three steps.

1 Given the initial demand curve $h_1(q')$, postulate a hypothetical increase in price from p_1' to $p_1^*(q')$, the choke price. In order to leave the individual no worse off, that person must be compensated by the area *ABC*.
2 Now postulate the improvement in quality and the shift in the demand curve to $h_1(q'')$. Given the weak complementarity assumption, utility is unaffected since the consumption of x_1 is zero. Therefore, there is no need for compensation, either positive or negative.
3 Now, postulate a return to the old price of p_1'. The individual is made better off by the area *ADE*. In order to restore the individual to the original welfare position, he/she must be taxed by this amount. The net effect of these changes is a gain to the individual (in the absence of the hypothetical compensating payments) of the area *BCED* (= *ADE* – *ABC*). This is the benefit of the change in q.

The role of the two conditions defining weak complementarity can now be made clear. If there were no choke price, there would be no finite compensating variations (*CV*s) in steps 1 and 3 above. In addition, if the derivative of the expenditure function were not zero at the choke price, there would be a welfare change (presumably positive) in step 2 associated with the increase in q, even though the quantity demanded of x_1 were zero. In this case, the area *BCED* would be an underestimate of the benefits of increasing q. If the consumption of x_1 is interpreted as use of an environmental resource (for example, trips to a recreation site), then whatever welfare change is associated with step 2 must be a nonuse value. Nonuse values are discussed in the last section of this chapter.

This result can also be established more rigorously. Recall from Chapter 3 that the compensating welfare measure for a change in q is

$$W_q = e\left(p_1', q', u^0\right) - e\left(p_1', q'', u^0\right). \tag{4.46}$$

The area between the two compensated demand curves is $\Delta = CV_{q''} - CV_{q'}$ where these terms are the *CV*s associated with the ability to purchase x_1 at different levels of q. Specifically,

$$CV_{q'} = e\left[p_1^*\left(q'\right), q', u^0\right] - e\left[p_1', q', u^0\right] \tag{4.47}$$

and

$$CV_{q''} = e\left[p_1^*\left(q''\right), q'', u^0\right] - e\left[p_1', q'', u^0\right]. \tag{4.48}$$

Using these expressions to calculate the area between the compensated demand curves gives

$$\begin{aligned}\Delta = {} & e\left[p_1^*\left(q''\right), q'', u^0\right] - e\left[p_1(q'), q'', u^0\right] \\ & - e\left[p_1^*\left(q'\right), q', u^0\right] + e\left[p_1(q'), q', u^0\right].\end{aligned} \tag{4.49}$$

If the second condition for weak complementarity is satisfied, the first and third terms in (4.49) sum to zero, and we have

$$\Delta = e\left[p_1, q', u^0\right] - e\left[p_1, q'', u^0\right] = W_q. \tag{4.50}$$

Weak Complementarity with More Than One Good

So far, it has been assumed that q is complementary to only one market good. It is possible for q to be complementary to several goods at the same time. For example, if q is water quality in a lake, it could be complementary to several market activities such as fishing, boating, and swimming. As shown in research by Bockstael and Kling (1988), the analysis of weak complementarity can be carried over to the multigood case in a straightforward manner.

Suppose there are two market goods that are complementary to q, with Hicksian demand functions as follows:

$$h_1 = h_1\left(p_1, p_2, q, u^0\right) \tag{4.51}$$

and

$$h_2 = h_2\left(p_1, p_2, q, u^0\right). \tag{4.52}$$

Using the expenditure function, the CS for a change in q is

$$CS = e\left(p_1,\ p_2,\ q',\ u^0\right) - e\left(p_2,\ p_1,\ q'',\ u^0\right). \tag{4.53}$$

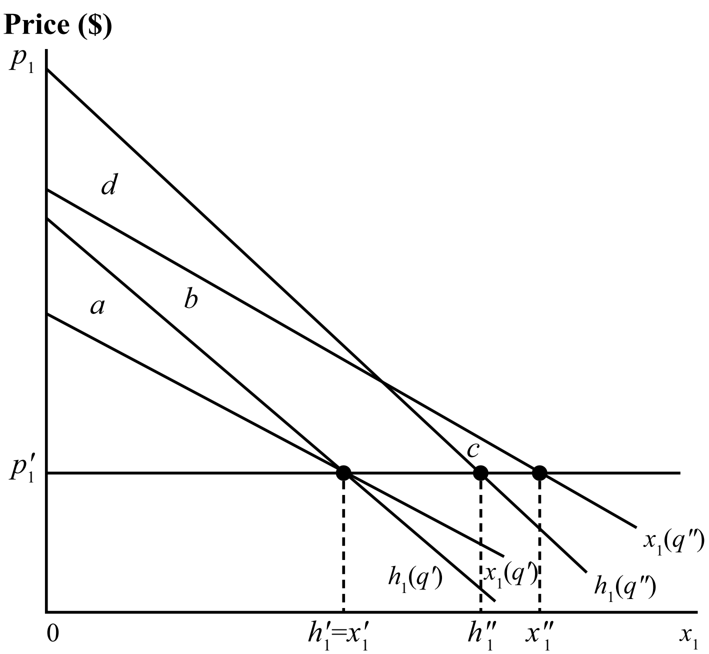

Figure 4.4 Approximating benefits using Marshallian demand curves with weak complementarity

If the conditions for weak complementarity are satisfied, *CS* can be estimated from areas between the compensated demand curves for the two market goods in the following manner. First, calculate the area between the compensated demands for x_1 holding p_2 at the observed price level. Then calculate the area between the compensated demands for x_2 evaluated at the choke price $p_1^*(q'')$ for x_1. Mathematically, this involves calculating the following expression:

$$\begin{aligned}\Delta = &\int_{p_1'}^{p_1^*(q'')} h_1\left[p_1,\, p_2',\, q'',\, u^0\right]dp_1 - \int_{p_1'}^{p_1^*(q')} h_1\left[p_1,\, p_2',\, q',\, u^0\right]dp_1 \\ &+\int_{p_2'}^{p_2^*(q'')} h_2\left[p_1^*\left(q''\right),\, p_2,\, q'',\, u^0\right]dp_2 \\ &-\int_{p_2'}^{p_2^*(q')} h_2\left[p_1^*\left(q'\right),\, p_2,\, q',\, u^0\right]dp_2.\end{aligned} \tag{4.54}$$

The first and third integrals give the value to the individual of consuming x_1 and x_2 when q is equal to q'' evaluated over a consistent path of integration. In other words, starting from (p_1', p_2'), p_1 is increased to $p_1^*(q'')$ and then p_2 is increased to $p_2^*(q'')$. Similarly, the second and fourth integrals give the value to the individual of consuming x_1 and x_2 when q is equal to q'. The result of this calculation is independent of the order in which the prices are changed.

When each of the integrals in equation (4.54) is expressed as a difference in expenditures, we have the following:

$$\begin{aligned}\Delta = &\, e\left[p_1^*\left(q''\right),\, p_2',\, q'',\, u^0\right] - e\left[p_1',\, p_2',\, q'',\, u^0\right] - e\left[p_1^*\left(q'\right),\, p_2',\, q',\, u^0\right] \\ &+e\left[p_1',\, p_2',\, q',\, u^0\right] + e\left[p_1^*\left(q''\right),\, p_2^*\left(q''\right), q'',\, u^0\right] \\ &-e\left[p_1^*\left(q''\right),\, p_2',\, q'',\, u^0\right] - e\left[p_1^*\left(q'\right),\, p_2^*\left(q'\right),\, q',\, u^0\right] \\ &+e\left[p_1^*\left(q'\right),\, p_2',\, q',\, u^0\right].\end{aligned} \tag{4.55}$$

When like terms are canceled out, and if weak complementarity is satisfied, this reduces to $\Delta = CS$ as defined in equation (4.53).

Welfare Measures with Ordinary Demand Functions

The primary empirical requirement for utilizing weak complementarity is to have an estimate of the demand function for the private good as a function of prices, income, and q. Since the compensated demand curves are not directly observable, it is useful to consider whether Willig's (1976) bounds on consumer surplus as an approximation to *CV* and *EV* apply in this situation. Unfortunately, since these bounds were derived for differences in areas under ordinary and compensated demand curves, they do not directly apply to areas between shifting ordinary and compensated demand curves.

This can best be shown graphically, as in Figure 4.4 (Bockstael, McConnell, and Strand 1991). The compensated demand curves for the two levels of q are denoted as $h_1(q')$ and $h_1(q'')$. The ordinary demand curves are denoted by $x_1(q')$ and $x_1(q'')$, and the compensating surplus for a change in q is measured by the area $b + d$. Observe that at the market price of p_1' the increase in q causes the ordinary demand curve to shift out farther to the right than the compensated demand curve. This is because with the ordinary demand curve, there is no compensating reduction in income to hold utility constant. Taking areas between the ordinary demand curves would yield a consumer surplus measure of $a + b + c$. The percentage error arising when the consumer surplus measure is used to approximate the compensating surplus is

$$\%\text{error} = \frac{a+c-d}{b+d}. \tag{4.56}$$

As can be seen by inspection, this error could be positive, negative, or by coincidence, equal to zero. Thus, the practice of using ordinary demand curves to estimate welfare changes with weak complementarity can lead to errors of unknown sign and magnitude.

It is possible to use the exact welfare measurement methods described in Chapter 3 to recover weakly complementary preferences from market demands—see for example Larson (1991) and Bockstael and McConnell (1993). However, to do so on the basis of market demand relationships *alone* requires invoking additional assumptions regarding preferences. The fundamental problem is that, in integrating back from Marshallian demand functions to the underlying indirect utility function, one needs to know how the constant of integration changes with changes in quality. The Willig (1978) conditions provide a means for filling in this missing piece of information. As Palmquist (2005) noted, the Willig conditions can be expressed in several ways. In the case of a single market good (x) and a single public good (q), with a corresponding indirect utility function $V(p,M,q)$, one form of the Willig conditions is that

$$\frac{\partial\left(\dfrac{V_q}{V_p}\right)}{\partial M} = 0. \tag{4.57}$$

Perhaps the more intuitive version of the Willig condition notes that it requires that the marginal willingness to pay for the public good per unit of the weakly complementary private good be independent of income. In other words,

$$\frac{\partial}{\partial M}\left[\frac{\left(V_q/V_M\right)}{x\left(p,M,q\right)}\right] = 0. \tag{4.58}$$

Smith and Banzhaf (2004) also demonstrated that the Willig condition restricts how the indifference fans in Figure 4.2 spread with changes in utility level. Given *both* weak complementarity and the Willig condition, the Marshallian surplus measure approximates the corresponding Hicksian welfare measures in a fashion analogous to the Willig conditions for a price change.

Weak Complementarity and Household Production

As Bockstael and McConnell (1983) have shown, if the individual choice problem is modeled in the household production function framework, there are conditions on the production technology and preferences that yield results equivalent to the weak complementarity model. They showed that if these conditions are satisfied, the welfare value of a change in q can be calculated from knowledge of the market good's compensated demand function even without knowledge of the household production technology. The demand function for the market good is a derived demand. If the conditions established by Bockstael and McConnell are satisfied, the welfare value of the change in q can be calculated by the area between the two compensated demand curves for the market good at the two levels of q.

The individual's choice problem is to maximize

$$u(\boldsymbol{Z}, q) \tag{4.59}$$

subject to

$$\boldsymbol{Z} = \boldsymbol{Z}(\boldsymbol{X}, q) \tag{4.60}$$

and

$$\sum_{j=1}^{J} p_j x_j = M \, . \tag{4.61}$$

The expenditure function is

$$e(\boldsymbol{P}, q, u) = \min\left\{ \sum_{j=1}^{J} p_j x_j \,\middle|\, u^0 = u\left[\boldsymbol{Z}(\boldsymbol{X}, q), q\right] \right\}. \tag{4.62}$$

Suppose that households produce z_1 by combining q and x_1, and that q is used *only* in the household production of z_1. As in the standard model, the compensated demand for x_1 is the derivative of the expenditure function with respect to p_1. In this case, it is a derived demand, so that it reflects features of both preferences and the household production technology.

An increase in q will lower the marginal cost of the household production of z_1 and therefore increase the quantity demanded of x_1. If the change in q is to increase the compensated demand for x_1, then the household production technology and preferences together must result in an increase in the demand for z_1 sufficiently large that more x_1 is required. In other words, restrictions on the household production technology alone are not sufficient to establish the complementarity or substitutability of the derived demand for x_1 with respect to q. A comparative statics analysis of the household maximization problem shows that with

$$\frac{\partial z_1}{\partial x_1} \cdot \frac{\partial x_1}{\partial q} > 0 \tag{4.63}$$

it is possible for an increase in q to reduce both the ordinary and the compensated derived demands for x_1. See, for example, Courant and Porter (1981, 325–326).

The area between the two derived compensated demand curves for x_1 is given by

$$e\left(p_1^*,\ q'',u^0\right)-e\left(p_1',\ q'',u^0\right)-e\left(p_1^*,\ q',u^0\right)+e\left(p_1',\ q',u^0\right). \tag{4.64}$$

If the first and third terms of this expression sum to zero, then this gives the welfare value of the increase in q. The sufficient conditions for these terms to sum to zero are: (a) that x_1 be essential in the production of z_1 so that if $x_1 = 0$, $\partial z_1/\partial q = 0$; and (b) that q not affect utility independent of its contribution to the production of z_1, in other words, that $\partial u/\partial q = 0$ when $x_1 = 0$.

These two conditions are sufficient to make the expenditure function independent of the level of q. Thus, they play a role similar to the second condition of the weak complementarity model. The household production framework requires restrictions on both the household production technology and preferences for final services; but there is no restriction on the Hicksian demand analogous to the existence of a choke price for the market good.

Hedonic Prices and the Value of q

The techniques described so far have been developed for the case where the level of the resource service or environmental quality is fixed and is the same for all individuals. Although this represents the textbook version of the public goods problem, it is not descriptive of all possible cases involving public goods or environmental quality. In some circumstances, the level of q can be considered to be a qualitative characteristic of a differentiated market good. In these cases, individuals have some freedom to choose their effective consumption of the public good or environmental quality through their selection of a private goods consumption bundle. Another way to look at it is that there is a kind of complementarity between the public good and the market good, in that as the quantity of the public good embodied in the market good increases, the demand for the market good increases.

In effect, the market for the differentiated private good functions also as a market for the public good or environmental quality. For example, people can choose the level of consumption of local public goods through their choice of a jurisdiction to reside in; thus, the housing market functions also as a market for the purchase of local public goods.

Where these choices are possible, information on public good demand is embedded in the prices and consumption levels for private goods. For example, if air quality varies across space in an urban area, individuals may choose their exposure to air pollution through their decisions about residential location. Residential housing prices may include premiums for locations in clean areas and discounts for locations in dirty areas. If they do, it may be possible to estimate the demand for public goods such as clean air from the price differentials revealed in private markets. These price differentials are implicit prices for different levels of the public good. A job can also be considered as a differentiated good. Different jobs have different bundles of characteristics and different wage rates. Wage differentials can be interpreted as the implicit prices of job characteristics. These wage and price differentials and the implicit prices they reflect are the subject matter of hedonic price theory.

The hedonic price technique is a method for estimating the implicit prices of the characteristics that differentiate closely related products in a product class. Colwell and Dilmore (1999) cited examples of the technique being applied to prices of farmland as early as 1922. The earliest modern example appears to be Griliches' (1961) application to the prices of automobiles. Rosen (1974) developed the formal theory of hedonic prices in the context of competitive markets. For an early development of the use of hedonic prices for estimating the demand for environmental quality characteristics, see Freeman (1974).

In principle, if the product class contains enough products with different combinations of characteristics, it should be possible to estimate an implicit price relationship that gives the price of any model as a function of the quantities of its various characteristics. This relationship is called the hedonic price function. The partial derivative of the hedonic price function with respect to any characteristic gives its marginal implicit price—that is, the additional expenditure required to purchase a unit of the product with a marginally larger quantity of that characteristic.

More formally, let Y represent a product class. Any model of Y can be completely described by a vector of its characteristics. Let $\boldsymbol{Q} = (q_1, \dots, q_K)$ represent the vector of characteristics of Y. Then any configuration of Y, say y_i, can be described by its characteristics—that is $y_i = y_i(q_{i1}, \dots, q_{iK}) = y_i(\boldsymbol{Q}_i)$, where q_{ik} is the quantity of the kth characteristic provided by configuration i of good Y. The hedonic price function for Y gives the price of any configuration, as a function of its characteristics. Specifically, for y_i,

$$p_y = p_y(q_{i1}, \dots, q_{iK}). \tag{4.65}$$

If $p_y(\cdot)$ can be estimated from observations of the prices and characteristics of different models, the price of any model can be calculated from knowledge of its characteristics.

Before turning to the interpretation of the hedonic price function, it will be useful to describe briefly how the hedonic price function is generated in a competitive market for a differentiated product. Following the analysis of Rosen (1974), assume that each individual purchases only one unit of Y in the relevant time period. An individual's utility depends upon that person's consumption of the numeraire, z, and the vector of characteristics provided by the unit of Y purchased:

$$u = u(z, \boldsymbol{Q}) \tag{4.66}$$

or, substituting out the numeraire good using the budget constraint $p_{yi} + z = M$

$$u = u\left[M - p(\boldsymbol{Q}), \boldsymbol{Q}\right]. \tag{4.67}$$

In order to maximize equation (4.67), the individual must choose levels of each characteristic (q_k) to satisfy

$$\frac{(\partial u / \partial q_k)}{(\partial u / \partial z)} = \frac{\partial p_y}{\partial q_k}. \tag{4.68}$$

That is, the marginal willingness to pay for q_k must just equal the marginal cost of purchasing more of q_k, with other things being equal.

While $p_y(\boldsymbol{Q})$ represents the price that an individual *must* pay for configuration $\boldsymbol{Q}$, Rosen (1974) distinguished this from what the individual would be willing to pay for the configuration. Specifically, he defined the bid function $B\left(M, \boldsymbol{Q}, u^*\right)$ as implicitly solving

$$u\left[M - B\left(M, \boldsymbol{Q}, u^*\right), \boldsymbol{Q}\right] = u^*, \tag{4.69}$$

where u^* is the solution to the constrained utility maximization problem. Since $M = z + p_y(\boldsymbol{Q})$, the bid function can also be viewed as providing the indifference curve mapping, capturing the tradeoffs among the individual attributes of Y and the numeraire good z for a given level of utility u^*.

In order to focus attention on a single attribute, q_k, let

$$B_k(q_k) = B\left(M, q_k, \boldsymbol{Q}^*_{-k}, u^*\right) \tag{4.70}$$

denote the bid function for attribute k, fixing all of the individual attributes *except* q_k at their optimal levels $\boldsymbol{Q}^*_{-k}$. Because of differences in preferences and/or incomes, individuals can have different bid functions. Two such functions for individuals a and b are shown in Figure 4.5. They both show diminishing marginal willingness to pay for q_k or a diminishing marginal rate of substitution between q_k and z. Given the hedonic price function, these two individuals choose q_k^a and q_k^b, respectively.

Turning to the supply side of the market, firms' costs of production depend upon the levels of the characteristics of the models they produce. Assume that firms are heterogeneous so that their cost functions are different. Inverting a firm's profit function yields an offer curve for the characteristic of the form

$$C_k = C_k\left(q_k, \boldsymbol{Q}^*_{-k}, \pi^*\right), \tag{4.71}$$

where π^* is the maximum attainable profit. Offer curves and the optimal quantities of q_k supplied for firms α and β are shown in Figure 4.6 for a given hedonic price function.

For all firms and individuals to be in equilibrium, all of the bid and offer curves for characteristics for each participant in the market must be tangent to the hedonic price function. Thus, the hedonic price function is a double envelope of the two families of bid curves and offer curves (Rosen 1974). As a double envelope, the hedonic price function depends on the determinants of both the supply side

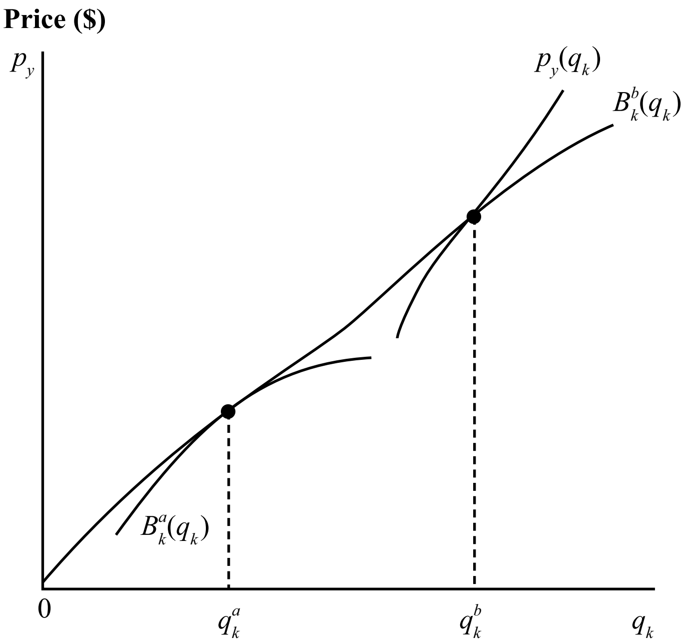

Figure 4.5 Bid curves of buyers in a hedonic market

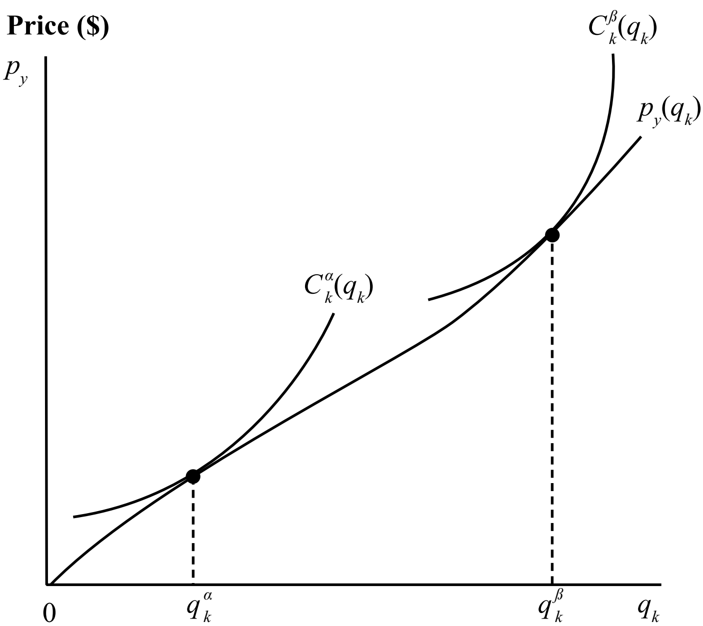

Figure 4.6 Offer curves of sellers in a hedonic market

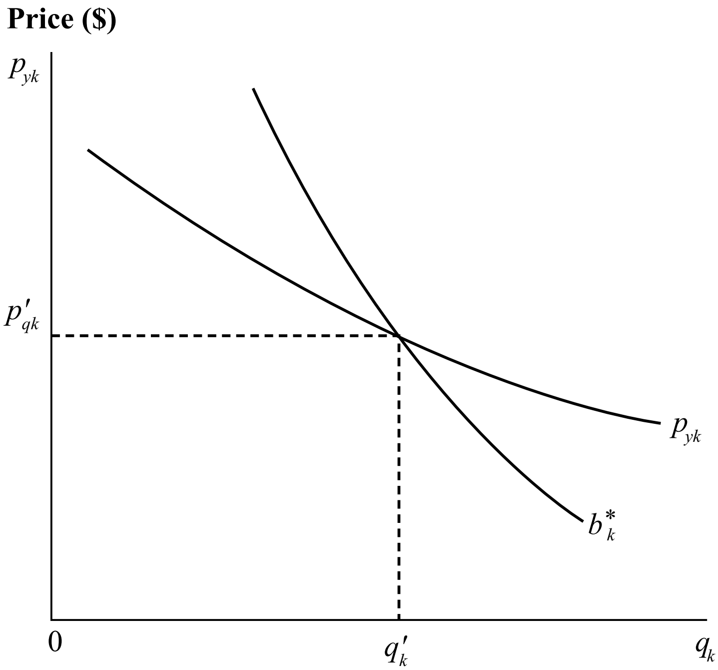

Figure 4.7 The marginal implicit price and inverse demand curves for *q*

and the demand side of the characteristics markets. It is a locus of equilibria between bids for, and offers of, characteristics.

The marginal implicit price of a characteristic can be found by differentiating the hedonic price function with respect to that characteristic—that is,

$$\frac{\partial p_y}{\partial q_k} = p_{yk}\left(q_1, \ldots, q_K\right). \tag{4.72}$$

This gives the increase in expenditure on Y that is required to obtain a model with one more unit of q_k, other things being equal. If equation (4.65) is linear in the characteristics, then the implicit prices are constants for individuals. However, if equation (4.65) is nonlinear, then the implicit price of an additional unit of a characteristic depends on the quantity of the characteristic being purchased.

Equation (4.65) need not be linear. Linearity will occur only if consumers can "arbitrage" attributes by untying and repackaging bundles of attributes (Rosen 1974, 37–38). For example, if individuals are indifferent between owning two two-door cars and one four-door car, other things being equal, they can create equivalents of four-door cars by repackaging smaller units. If both sizes exist on the market, the larger size must sell at twice the price of the smaller one, and the hedonic price of a car will be a linear function of the number of doors. The example suggests that nonlinearity will be a common feature of hedonic price functions.

Another way of looking at the market equilibrium for a characteristic helps to make clearer the welfare implications of the hedonic price model. First, individual a's choice problem can be solved to obtain that person's uncompensated inverse demand function for q_k

$$b_k^{*a} = b_k^{*a}\left(q_k, \boldsymbol{Q}_{-k}^{*}, M - p_y\right), \tag{4.73}$$

where a indexes the individual. This is shown in Figure 4.7 along with the marginal implicit price function for q_k, $p_{yk}\left(\boldsymbol{Q}\right)$. For each individual, the quantity of q_k purchased is known by observation and its implicit price is known from equation (4.72). This point $\left(q_k', p_{yk}'\right)$ can be interpreted as a utility-maximizing equilibrium for this individual resulting from the intersection of the individual's inverse demand curve and the locus of opportunities to purchase q_k, as defined by the marginal implicit price function. Thus, from the first-order conditions of equation (4.68) p_{yk}' can be taken as a measure of the individual's equilibrium marginal willingness to pay for q_k.

Individuals can be viewed as moving out along p_{yk}' as long as their willingness to pay, as reflected in their inverse demand curves, exceeds the marginal implicit price. Thus, p_{yk}' is a locus of individuals' equilibrium marginal willingness to pay.

It is also possible to specify a compensated marginal willingness-to-pay function for q_k,

$$b_k^{a} = b_k^{a}\left(q_k, \boldsymbol{Q}_{-k}^{*}, u*\right). \tag{4.74}$$

This function gives individual a's marginal willingness to pay for q_k, holding utility constant. The value of a nonmarginal change in q_k can be easily calculated if the compensated marginal willingness-to-pay function equation (4.74) is known. This value is given by the integral of equation (4.74) over the range of the change in q_k.

This leads us naturally to the question of whether individuals' marginal willingness-to-pay and/or their inverse demand functions can be identified from observations of marginal implicit prices and quantities. The answer depends on the circumstances of the case. If the implicit price function is linear in q_k, then it is not possible to identify a demand curve for q_k. The price observation is the same for all individuals. However, p_{yk}' can be interpreted as the marginal willingness to pay or marginal benefit for small changes in q_k for each individual.

If the hedonic price function is nonlinear, different individuals selecting different bundles of characteristics will have different marginal implicit prices for q_k. There is one situation where the inverse demand function can be immediately identified—when all individuals have identical incomes and utility functions. When this occurs, the marginal implicit price function is itself the inverse demand function. Recall that the marginal implicit price curve is a locus of equilibrium points on individuals' inverse demand curves. With identical incomes and preferences, all individuals have the same inverse demand curve. Since all the equilibrium points fall on the same inverse demand curve, they fully identify the curve.

In the case where differences in incomes, preferences, or other variables result in individuals having different inverse demand functions, Rosen (1974) argued that implicit price and quantity data from a single market could be used to estimate this inverse demand function, provided that the standard identification problem of econometrics could be solved. It is now clear that this analysis is incorrect. The problem is that the data from a single hedonic market are insufficient to identify how the same individuals would respond to different implicit prices and incomes.

There are at least two ways in which estimates of inverse demand functions for q_k can be obtained from hedonic analysis. The first is to increase the quantity of information obtained from marginal implicit prices by estimating hedonic price functions for several separate markets, and then pooling the cross-sectional data on the assumption that the underlying structure of demand is the same in all markets (Freeman 1974; Brown and Rosen 1982; Palmquist 1984). The second approach is to impose additional structure on the problem by invoking a priori assumptions about the form of the underlying utility function. The nature of the identification problem in hedonic price models, and alternative approaches to dealing with it, are major topics of Chapter 10.

Nonuse Values

Up to this point in the chapter, the values derived for environmental amenities have been obtained by looking for telltale "footprints" in the marketplace associated with how individuals *use* these amenities to enhance their welfare— producing utility-yielding commodities, or using them as substitutes for, or complements to, market goods (such as recreation trips). This section considers the possibility that people value natural resources and environmental characteristics independently of any present or future use they might make of those resources. For example, people may gain utility from their knowledge that the Grand Canyon is preserved even though they expect never to visit the canyon. People may value the survival of whales, eagles, and other endangered species even though they never expect to see one of them. Nonuse values raise two key questions that will be discussed further in this chapter. First, is there a meaningful way to define use so that values that arise from use can be distinguished from those that are independent of use? Second, what methods are available to measure these "nonuse" values and/or test for their existence?

In the economics literature, natural resource values that are independent of people's present use of the resource have been variously termed "existence," "intrinsic," "nonuse," and "passive use" values. This latter term was coined by the D.C. Circuit Court of Appeals in its ruling in *Ohio v. U.S. Department of Interior* (880 F.2d 432, 1989), which legitimized the inclusion of these values in natural resource damage cases brought by the federal government. These values may arise from a variety of motives, including a desire to bequeath certain environmental resources to one's heirs or future generations, a sense of stewardship or responsibility for preserving certain features of natural resources, and a desire to preserve options for future use.

Evidence suggests that nonuse values can be large in the aggregate, at least in some circumstances. For example, a research team working for the state of Alaska estimated that the lower-bound median lost nonuse value for U.S. residents due to the *Exxon Valdez* oil spill was about $30 per household in 1990 dollars or about $2.8 billion for the U.S. as a whole (Carson et al. 2003). If nonuse values are large, ignoring them in natural resource policymaking could lead to serious errors and resource misallocations.

The goal of this section is to examine the role of nonuse values in assessing the *total value* (i.e., the combined use and nonuse value) of a natural resource or environmental amenity. The section begins with a review of some of the earlier literature in which various forms of nonuse values are discussed. A utility theoretic framework is then developed to define nonuse value—though in doing so it becomes clear that the distinction between use and nonuse values is an artifact of the specific functional forms chosen by the analyst. The section concludes with a discussion of the empirical techniques for the measurement of whatever we choose to call nonuse values. The possibility of deriving measures of nonuse values from observed behavior of individuals is considered, but this does not appear to be a fruitful approach. This leaves stated preference methods (alone or in combination with revealed preference methods) as the basis for both measuring nonuse values and for testing for their existence. Some discussion is provided regarding the issues involved in applying stated preference valuation techniques to the measurement of nonuse values, though this is left in large part to Chapter 12.

Background: Motivation and Definitions

John Krutilla (1967) introduced the concept of existence or nonuse values into the mainstream economics literature. In his classic article "Conservation Reconsidered," he argued that individuals do not have to be active consumers of a resource, whose willingness to pay can be captured by a price-discriminating monopoly owner, in order to derive value from the continuing existence of unique, irreplaceable environmental resources. Krutilla wrote that "when the existence of a grand scenic wonder or a unique and fragile ecosystem is involved, its preservation and continued availability are a significant part of the real income of many individuals" (Krutilla 1967, 779). In an accompanying footnote he added that "these [individuals] would be the spiritual descendants of John Muir, the present members of the Sierra Club, the Wilderness Society, the National Wildlife Federation, the Audubon Society and others to whom the loss of a species or the disfigurement of a scenic area causes acute distress and a sense of genuine relative impoverishment" (Krutilla 1967, 779n7).

Krutilla went on to suggest at least two reasons why people might hold values unrelated to their current use of a resource. These reasons were related to bequeathing natural resources to one's heirs and preserving options for future use. In a footnote he suggested that the "phenomenon discussed may have an exclusive sentimental basis, but if we consider the bequest motivation in economic behavior … it may be explained by an interest in preserving an option for one's heirs to view or use the object in question" (Krutilla 1967, 781n11). He also wrote, "an option demand may exist not only among persons currently and prospectively active in the market for the object of the demand, but among others who place a value on the mere existence of biological and/or geomorphological variety and its widespread distribution" (Krutilla 1967, 781). However, literature that is more recent has recognized that the value of preserving options arises out of uncertainty about future demand and/or supply and has different implications for environmental valuation. For a brief discussion of this issue and references to the literature, see Chapter 5 of this book.

Later, Krutilla and Fisher wrote:

> In the case of existence value, we conceived of individuals valuing an environment regardless of the fact that they feel certain they will never demand *in situ* the services it provides … however, if we acknowledge that a bequest motivation operates in individual utility-maximizing behavior … the existence value may be simply the value of preserving a peculiarly remarkable environment for benefit of heirs.
>
> (Krutilla and Fisher 1975, 124)

Whereas Krutilla and Fisher offered a bequest motivation as one of several possible explanations for a pure existence value, McConnell (1983) took a different point of view:

> In most cases, resources are valued for their use. Existence value occurs only insofar as bequest or altruistic notions prevail. We want resources there because they are valued by others of our own generation or by our heirs. Thus, use value is the ultimate goal of preferences that yield existence demand, though the existence and use may be experienced by different individuals.
>
> (McConnell 1983, 258)

Thus, in McConnell's view, an altruistic attitude toward other people's use of a resource is the primary source of existence value. Although the source of existence value is related to someone's use, it is independent of any use made of the resource by the person holding the existence value.

Some have argued that people can have what are essentially existence values out of an ethical or altruistic concern for the status of nonhuman species or proper rules of human conduct (Kopp 1992). While philosophers of ethics are not in agreement as to the validity and proper form of such concerns—for example, see one of the collections on environmental ethics such as Van De Veer and Pierce (1994), DesJardins (1999), or Jamieson (2000)—it is possible that some people hold such values and are willing to commit resources on the basis of those values. Such values could be entirely independent of any use of the environment.

There has also been a lot of discussion about the various forms that nonuse values might take. A typical approach in the literature is to define use value as the economic value associated with the *in situ* use of a resource—for example, through visiting a recreation site or observing a natural wildlife population. Total value is viewed as an individual's willingness to pay to preserve or maintain a resource in its present state. If total value exceeds use value, the difference is a nonuse value, or as some have called it, an "existence value," an "intrinsic value," or a "preservation value." Several authors have chosen one of these terms to represent the total difference and then identified various possible components of this total, often based on assumed motivations for holding these values. For example, Fisher and Raucher (1984) used the term "intrinsic value" to refer to the aggregate, and stated that the total intrinsic value is the sum of option value, aesthetic value, existence value, and bequest value. Sutherland and Walsh (1985) used "preservation value" to refer to the aggregate and stated that it is the sum of option, existence, and bequest values.

The attention given to questions of classification and motivation is perhaps misdirected. Motivations do not play an important role in the empirical analysis of the demands for market goods. There is little talk of "prestige value" and "speed value" in the literature on the demand for automobiles. So why should motivations be important in the case of nonuse values? Arguments about motivations seem to be offered primarily to persuade the reader of the plausibility of the hypothesis that nonuse values are positive. However, the real test of this hypothesis will come from the data. Rather than further debating definitions and possible motivations, it would be more useful to proceed with a test of the hypothesis that nonuse values (defined in a way that makes testing of the hypothesis feasible) are positive. If the evidence supports this hypothesis, then further research might be devoted to testing hypotheses about the determinants of the size of nonuse values in different cases. At that point, investigation of motivations might be useful in formulating hypotheses for testing. Then the choice of terms and explanatory variables, however, would be governed by what are empirically meaningful distinctions.

One of the major issues in the literature on nonuse values is how to define the use that lies behind use value. The most common approach is to identify some market good or service that is a complement to the resource, as regards consumption, and to define and measure use in terms of the purchased quantity of this complementary good. If the resource is a park, the complementary good is the purchase of travel services to the park, and use of the resource is measured by the number of trips purchased. However, this approach to defining use is clearly a simplification of a more complex reality. The physical proximity that one normally thinks of as being an essential part of use can occur independently of the purchase of a complementary good such as travel. For example, people who live within the natural range of an endangered species such as the California condor or wood stork may be able to view one of these birds and experience the utility and value associated with the sightings as an incidental part of their daily routine. In that case, there is no connection between that kind of use and any market good.

Some have argued that use does not require the physical proximity of the user and the resource. In their view, use can be defined to include the purchase of a complementary market good that embodies some visual or literary representation of the resource. Randall and Stoll (1983), for example, argued that there can be off-site uses, which they label as "vicarious consumption." Examples of vicarious consumption could include watching a nature video, observing a live webcam of an eagle's nest, or watching a YouTube post of a forest fire. This is essentially what Boyle and Bishop (1987) chose to call indirect use value.

Defining use in this way creates some problems for the measurement of value. One problem with such vicarious or indirect forms of use is that the observable market transaction, such as the purchase of a nature magazine, often entails the simultaneous use of many environmental resources, so that allocation of the market transaction to specific resources is not possible. Furthermore, where vicarious uses involve information conveyed by photographs, films, and the like, the public-good dimension of information seems likely to virtually destroy any meaningful relationship between observed market behavior and underlying values. In addition, vicarious use has the odd feature that it can occur even though the resource itself no longer exists (as in the viewing of films and photographs).

This discussion of the possible definitions of, and motivations for, various types of use and nonuse values highlights the difficulty that authors have had in making clear distinctions among the concepts. However, definitions can be considered in part a matter of taste. A set of definitions can be considered useful if it furthers research objectives and leads to useful answers to meaningful questions, and if the definitions are based on operationally meaningful distinctions. If use values are limited by definition to those associated with *in situ* use, as measured by the purchase of a complementary good, then the definition has the virtue of distinguishing between cases where use of a resource generates observable market data and cases where no meaningful data can be obtained by observing market transactions. However, this definition leaves out vicarious or indirect uses as well as what one might call incidental uses. This argues for dropping the distinction between use and nonuse values and instead distinguishing between values that can be estimated with revealed preference data versus those that cannot. In the next section, a theoretical framework is provided that distinguishes between those cases where changes in the level of q lead to changes in the behavior of individuals and those where they do not. This framework is then used to examine whether total value can be usefully decomposed into component parts.

A Theoretical Framework

Assume that an individual has a preference ordering over a vector of market goods, $\boldsymbol{X}$, and some nonmarketed resource, q. The individual has no control over the level of q, but takes it as given. Here, q is taken to be a scalar measure of some characteristic of the environment, such as the population of some species or the value of some parameter of water quality. In the abstract, q can represent a measure of either a quantity or a quality. The choice of a unit for measuring q

has important implications for measurement in practice, but that question is not addressed here. The assumption that the environmental resource can be described by a single attribute is clearly a simplification. A more realistic model would allow for simultaneous changes in two or more quantitative and/or qualitative characteristics of the resource.

The notion that the total value of the resource q is not solely tied to the purchase of a complementary market good can be captured by assuming that the individual's direct utility function takes the form

$$U = T\left[u\left(\boldsymbol{X},q\right),q\right], \tag{4.75}$$

where $T\left(u,q\right)$ is increasing in u and q, and $u\left(\boldsymbol{X},q\right)$ is increasing and quasi-concave in $\boldsymbol{X}$ and q. In this case, the nonmarketed resource q enters utility in two distinct places, first by interacting with the marketed commodities, $\boldsymbol{X}$, and second on its own. The key feature of this functional form is that the marginal rates of substitution among the marketed goods, which dictate their observed consumption levels, are independent of this second role of q. Thus, information on $\boldsymbol{X}$ provides no information on this portion of preferences.

The functional form in (4.75) has been used by a number of authors to decompose the total value associated with a change in q into a series of component elements (e.g., Hanemann 1988, Freeman 2003, and Herriges, Kling and Phaneuf 2004, among others). The indirect utility function implied by (4.75) is given by

$$\begin{aligned} V\left(\boldsymbol{P},q,M\right) &= \max_{\boldsymbol{X}}\left\{T\left[u\left(\boldsymbol{X},q\right),q\right] \middle| \sum_j p_j x_j = M\right\} \\ &= T\left[v\left(\boldsymbol{P},q,M\right),q\right] \end{aligned} \tag{4.76}$$

where

$$v\left(\boldsymbol{P},q,M\right) = \max_{\boldsymbol{X}}\left\{u\left(\boldsymbol{X},q\right) \middle| \sum_j p_j x_j = M\right\} \tag{4.77}$$

captures that portion of utility that derives from the interaction of $\boldsymbol{X}$ and q in the marketplace. The total compensating variation CS^T for a change in the resource from q^0 to q^1 can then be implicitly defined by

$$T\left[v\left(\boldsymbol{P},q^0,M\right),q^0\right] = T\left[v\left(\boldsymbol{P},q^1,M-CS^T\right),q^1\right]. \tag{4.78}$$

Hanemann (1988) suggested identifying that portion of CS^T *revealed* by market transactions, as CS^R. This is defined implicitly by

$$T\left[v\left(\boldsymbol{P},q^0,M\right),q^0\right] = T\left[v\left(\boldsymbol{P},q^1,M-CS^R\right),q^0\right]. \tag{4.79}$$

The residual component of value (i.e., $CS^E \equiv CS^T - CS^R$), which is called *existence value* here, is then implicitly defined by

$$\begin{aligned} T\left[v\left(\boldsymbol{P},q^0,M\right),q^0\right] &= T\left[v\left(\boldsymbol{P},q^1,M-CS^R\right),q^0\right] \\ &= T\left[v\left(\boldsymbol{P},q^1,M-CS^R-CS^E\right),q^1\right]. \end{aligned} \tag{4.80}$$

The first line captures the impact that the change in q has on the portion of individual preferences revealed through the market (CS^R). The second line then indicates the *additional* compensation required to make the individual whole. However, as Mäler, Gren, and Folke (1994) and others have noted, this decomposition of total value is arbitrary. One could, for example, reverse the order of changes in (4.80), and decompose total value using $CS^T = \widetilde{CS}^E + \widetilde{CS}^R$, where the components of this decomposition are implicitly defined by

$$\begin{aligned} T\left[v\left(\boldsymbol{P},q^0,M\right),q^0\right] &= T\left[v\left(\boldsymbol{P},q^0,M-\widetilde{CS}^E\right),q^1\right] \\ &= T\left[v\left(\boldsymbol{P},q^1,M-\widetilde{CS}^R-\widetilde{CS}^E\right),q^1\right]. \end{aligned} \tag{4.81}$$

The problem is analogous to the one that arises when multiple price changes are sequenced in calculating compensating variations. By path independence, the sum, CS^T, is independent of the sequence of evaluation. However, as in the case of multiple price changes, the individual components of the sum are not.

Herriges, Kling, and Phaneuf (2004) further decomposed CS^R when q is tied to a specific market commodity or subset of market commodities, as is the case, for example, when q represents an environmental attribute associated with a recreational site. Let x_1 denote the number of trips to the associated site. CS^R can then be decomposed into two pieces, direct use value, and indirect use value as follows:

$$CS^R = CS^U + CS^{IU} \tag{4.82}$$

where $CS^U \equiv CS^R - CS^{IU}$ and CS^{IU} is implicitly defined by

$$v\left[p_1^*\left(q^0\right),\boldsymbol{P}_{-1},q^0,M\right] = v\left[p_1^*\left(q^1\right),\boldsymbol{P}_{-1},q^1,M-CS^{IU}\right], \tag{4.83}$$

with $p_1^*\left(q^s\right)$ denoting the choke price for x_1 when $q = q^s$ (s = 0, 1) and $\boldsymbol{P}_{-1} = \left(p_2,\ldots,p_J\right)$ denoting the vector of all the remaining market prices. Herriges, Kling, and Phaneuf (2004) suggested referring to CS^{IU} as a measure of "indirect use" value, as it corresponds to welfare changes stemming from changes in q when the corresponding market commodities are not in use, whereas CS^U corresponds to direct use value. The main distinction between CS^{IU} and CS^E is that the former affects market behavior, and therefore may be revealable from observations on market transactions, whereas the latter leaves no such footprint.

Unfortunately, while the above decompositions of total value have some intuitive appeal, they are of little guidance to the practitioner, particularly one faced with only revealed preference data. With only revealed preference data, the best one can hope to identify is the revealable portion of value CS^R. Yet, even in this instance, a fundamental problem remains—as noted above, division of total value between CS^E and CS^R is not unique. Indeed, Herriges, Kling, and Phaneuf (2004) provided an empirical example in which two different specifications for the partial indirect utility function $v(\boldsymbol{P}, q, M)$ that yield *identical* Marshallian demand equations result in estimates of CS^R that differ by more than a factor of three.

One tempting approach to this problem would be to take a "conservative" approach acknowledging the possibility of existence values, but to report CS^R as a conservative estimate of total value. Unfortunately, CS^R is not necessarily a lower bound on the overall compensating variation. Indeed, Herriges, Kling, and Phaneuf (2004) provided several counter-examples to show that there is no unique decomposition and that specific results are analogous to the path dependence problem in integrating demand functions over multiple price changes.

These findings are troubling. In simple terms, they imply that if an individual's preferences do not exhibit weak complementarity, then it is not even possible to estimate a lower or upper bound on that individual's value of a change in environmental quality based on market-based information. Intuitively, the lack of weak complementarity means that not all of the information related to use value of a quality change is contained in the demand for the market good. If the lack of weak complementarity arises due to the fact that there is more than one private good that has joint weak complementarity with q, then the Bockstael and Kling model described above can be used to estimate the total value. If this option is not possible, then this leaves analysts with two real alternatives in terms of capturing the total value associated with a change in q.

The first, as Bockstael and McConnell (2007, 66) suggested, is to take the position that, while weak complementarity is not a testable hypothesis (at least on the basis of revealed preference data alone), it "is often a good assumption to maintain" and that "cases where weak complementarity would be a misguided assumption—for example, a travel cost study of the demand for access to the Arctic National Wildlife Refuge—will often be obvious." Assuming weak complementarity eliminates the ambiguity in terms of the resource's value, since, in this case, $CS^E = CS^{IU} = 0$, so that $CS^T = CS^R = CS^U$ (i.e., all the value associated with the resource q is tied to the *use* of the weakly complementary market commodity x_1).

Second, the analyst can seek to eliminate the ambiguity in terms of individual preferences by augmenting available revealed preference data using information from stated preference sources. As Ebert (1998) argued, the fundamental problem when assessing the welfare impact of changes in the level of a nonmarketed commodity is that the ordinary demand equations for the marketed commodities do not contain all of the information needed to recover preferences. In particular, any effort to integrate back from observed demands to underlying preferences must make assumptions about how the constant of integration changes with changes in the public good, q. Weak complementarity provides one such set of assumptions. Ebert argued, however, that stated preference techniques can be used to obtain the missing information in the form of inverse demand equations for the public goods. Eom and Larson (2006) provided an illustration of this approach. The downside, of course, is that the technique relies on the assumption that stated preference elicitation methods provide an accurate representation of individual preferences—an assumption that (in the case of nonuse values at least) is once again untestable using revealed preference data alone. This topic will be explored in more depth in Chapter 12.

The Household Production Framework and Use Values

The theory of household production provides an alternative framework for thinking about how resources affect utility and welfare and what it means to assume that x_1 is a measure of the use of a resource. Recall that in the household production framework, the demand functions for market goods are derived demands. Assume that q is an input in the production of only one final service flow, z_1. If x_1 is also an input in the production of z_1, then (as shown earlier in this chapter) q will be an argument in the ordinary and compensated demand functions for x_1. If x_1 and q are the only inputs in z_1, then the value of q in producing z_1 will be reflected in the derived demand for x_1 (Bockstael and McConnell 1983). If x_1 and q are both essential inputs, then the marginal product of each is zero if the other input is absent. A positive level of input for x_1 is required to utilize q in household production. This is equivalent to saying that the use of q, and hence its use value, is reflected in the level of x_1.

Another advantage of thinking about resource values in the context of the household production framework is the light that it sheds on the nature of what have been called indirect, or vicarious, use values. It appears that there are two categories of indirect or vicarious values—those that derive from household production and those that derive from market production. If the vicarious use involves the viewing of home-produced films and photos, then the theory suggests that the place to look for measures of indirect use values is in the influence of q on the demands for other inputs in the production of films and photos. If a market good is an essential input to this indirect use, then in principle the use value of changes in q can be measured by areas between the demand curves for the input in home production.

On the other hand, if indirect, or vicarious, use is the viewing of commercially produced films and television programs or the reading of books and magazines, the source of value for q is in its influence on either the production function or the quality-differentiated demand functions for these marketed goods. The place to look for measures of this type of value is either in changes in costs and prices or in the hedonic price functions for these goods. Thus, the role of q is analytically no different from the role of, say, air quality in the production of agricultural crops. There is a well-developed methodology for measuring values in this case, as shown in Chapter 8. However, it seems doubtful that a reduction in the availability of a specific resource would increase the cost of producing the editorial content of films, books, and magazines, so the empirical significance of this approach to measuring values of changes is in question.

It seems reasonable to assume that the demand for a specific magazine or television program will depend in part on the quality and the information content of the material presented. However, the link between changes in q and changes in the demand for information about a specific resource is complex and may not meet the requirements of the standard models for estimating welfare values from changes in demand. For example, a major pollution event that damages a unique resource system may increase the demand for information about that resource; and producers are likely to respond with more articles and programs about that resource. This would be an increase in vicarious use, but it could hardly be called a benefit resulting from damage to the resource.

Altruism and the Relevance of Nonuse Values

As noted in the first part of this section, one possible reason that nonusers might value the preservation of a resource is an altruistic feeling toward those who do use it. There is literature on altruism and how the presence of altruistic feelings toward others affects the Pareto optimum or efficient allocations of resources. One conclusion in this literature is that if those feelings take a form known as nonpaternalistic altruism, they have no effect on the Pareto optimum allocation—that is, they are irrelevant for benefit-cost analysis (see Lazo, McClelland, and Schulze 1997, and references cited therein). If nonusers' willingness to pay for the preservation or improvement of a resource is due to nonpaternalistic altruism, this means that the nonuse values should not be counted in a benefit-cost analysis (Milgrom 1993).

Nonpaternalistic altruism refers to a case where one individual cares about the general level of well-being of others but does not have any preferences regarding the composition of consumption bundles of others. If individual A had such preferences, they could be represented by

$$u^A = u^A\left[\boldsymbol{X}^A, u^B\left(\boldsymbol{X}^B, q^B\right)\right]. \tag{4.84}$$

An increase in q would result in an increase in u^B; and therefore, A would also be better off. A's willingness to pay for this improvement could be defined in the usual fashion; but it would not be proper to add this to B's willingness to pay in an economic assessment of the policy, at least if the increase in q was not costless. This is because someone has to bear the cost of the increase in q. If B bears the cost, this decreases B's utility, and therefore A's willingness to pay for the increase. It can be shown that as long as altruism takes the form shown in equation (4.84) the terms representing altruism cancel out of the conditions for Pareto optimality. However, if A's altruism stems from a concern for the level of q that B experiences, then A's willingness to pay for B's improvement in

q is relevant for economic assessments (Lazo, McClelland, and Schulze 1997; McConnell 1998).

In summary, if the sole source of nonuse values is nonpaternalistic altruism, these values are irrelevant for policy purposes. However, if altruism takes the paternalistic form, then the resulting nonuse values are relevant for policy analysis; and, nonuse values can arise for other reasons besides, or in addition to, altruism.

Toward Measurement

If one accepts the argument that for practical purposes nonuse values should be defined as those values not revealed by market behavior, then by definition other strategies for valuation must be sought. Stated preference methods are an obvious approach, but, before leaving behavioral observations altogether, it is useful to ask whether there is any other behavior, outside of market behavior, where information that might lead to measurement of nonuse values could appear. Since it is concluded that these are unlikely to be fruitful avenues of research, the remainder of the subsection focuses on how stated preference methods might be used as part of a research strategy for measuring nonuse values as a separate component of total value. The subsection concludes with a discussion of other issues that must be resolved in any effort to measure and/or isolate nonuse values.

Because preservation of a resource for those who do not make *in situ* use of it has the properties of nonexcludability and nondepletability, one would expect markets to fail to provide these preservation services, or at least to provide them in suboptimal quantities. Unless there is a market for preservation of a resource or for enhancing its quality for nonusers, there will be no market transactions to reflect the preservation values of individuals, and only stated preference methods of estimating nonuse values will be feasible.

Environmental organizations are observed undertaking a variety of activities to protect and preserve natural environments, and people support these activities through voluntary contributions of time and money. The question is whether this revealed behavior provides an adequate basis for measuring willingness to pay for preservation. For several reasons, it seems unlikely that economic data on either the activities of these organizations or individuals' contributions to them can be relied upon as measures of the value of preservation for policymaking purposes.

The activities of environmental organizations can be placed in one of two categories: (a) direct provision of preservation through acquisition, and (b) advocacy in an effort to influence public-sector provision of preservation. Organizations such as The Nature Conservancy accept private donations and use the funds to purchase lands with special ecological, geological, or scenic characteristics for the purpose of protection and preservation. Individuals' donations and dues paid to such organizations are manifestations of willingness to pay for preservation. However, if "free-rider" behavior were significant, these donations would be only a lower bound of true aggregate willingness to pay. Furthermore, at least in some instances, the lands acquired by such organizations are accessible to individual use; therefore, individual donations could reflect a combination of use as well as nonuse values.

Many environmental organizations devote a substantial portion of their budgets to advocacy activities on behalf of environmental preservation in general and for policy actions to protect specific natural resources. Again, individuals' membership dues and donations in support of these activities reflect individuals' willingness to pay for preservation. However, because of free-rider behavior, aggregate donations are likely to represent less than total willingness to pay. In addition, the organizations undertaking these activities frequently have multiple purposes—they provide services to members such as magazines and other publications, and field trips. This means that only the portion of dues and donations supporting the incremental cost of advocacy to the organization is relevant for estimating preservation values.

In the policymaking arena, another factor weakens the relationship between individual donations to support advocacy and preservation values. This factor is the uncertainty concerning the outcome of the policy process and the contribution of advocacy activities to the desired outcome. A rational organization with limited resources would estimate the probabilities of successful advocacy as a function of the resource commitment for each specific issue, and would allocate resources so as to maximize the expected value of the outcome. The observed allocation of advocacy resources across specific issues would reflect the interaction of the probabilities of a successful advocacy, the marginal productivity of advocacy in increasing the probability of success, and the value of success to the organization (Freeman 1969). It would probably prove difficult, if not impossible, to model this complex policy process so as to identify individuals' willingness to pay on the basis of observed contributions to advocacy on specific issues.

It is reasonable to conclude that individuals' donations to environmental organizations involved in acquisition and advocacy reflect willingness to pay for preservation and nonuse values. These activities provide evidence in support of the hypothesis of significant nonuse values. However, for the reasons outlined here, observed acquisition and advocacy expenditures are likely to be an underestimate of the values held by people who are not *in situ* users of the resource.

The stated preference valuation methods described in Chapter 12 are likely to offer the only feasible approaches to estimating nonuse values. Of course, what matters for policy purposes is total value regardless of how it is divided between use and nonuse value. For individuals who are nonusers (or who do not alter observable behavior in response to changes in q), stated preference questions provide estimates of total value, which consists entirely of nonuse value. For individuals who are users and have separable preferences such as those represented by equation (4.75), stated preference methods still yield estimates of total value; and, revealed preference methods yield underestimates of total value.

However, for research purposes there is some interest in learning more about the magnitudes of nonuse values and factors determining their size. Broadly speaking, there are two approaches to estimating nonuse values. The first is to use a stated preference question to obtain an estimate of total value and then to deduct from this a separately obtained estimate of use value. The latter could be obtained, for example, by the travel cost method for estimating the demand for visits to a recreation site. As should be apparent, one problem with this approach to imputing nonuse values is that the imputation is based on the difference between two other values, both of which are measured with some unknown error. Therefore, without some understanding of the error properties of the other measures, one cannot know whether the imputed value is simply the result of measurement error or is a true nonuse value. Moreover, as demonstrated by Herriges, Kling, and Phaneuf (2004), the division of total value between use and existence value is itself an arbitrary allocation.

The second approach to estimating nonuse values is to ask people explicitly about their nonuse values, either by questioning people who are known to be nonusers or by asking people to assume that their own use is zero. Any value revealed by a nonuser is, by definition, a nonuse value; but if present users are asked to assume that they are not using the resource, their responses to stated preference questions may not be valid indicators of nonuse value. This is because the situation they are being asked to see themselves in is unfamiliar, and they might misunderstand it or reject it outright.

Two additional issues concerning research study design deserve at least brief mention. The first issue concerns the definition and description of the resource to be valued. Resources have both quantitative and qualitative dimensions that can be affected by policy decisions or damaged by pollution events. Yet, in the theoretical discussion above, it was assumed that a resource was measured in a single dimension; therefore, respondents must be given a clear description of the changes in all relevant dimensions of resource quantity and quality. The determination of what is relevant must come in part from the judgment of experienced researchers in this field and in part from research specifically designed to determine what characteristics of the resource are important to people.

The second issue concerns the relevant population for sampling when nonuse values are involved. If the resource to be valued is in California, for example, should the sample include East Coast residents, or should it be limited only to westerners or to California residents? Casting the sampling net too wide wastes scarce research resources, but important values may be missed if the geographic scope of the sample is too narrow. Even small per capita values can loom large when aggregated over a large population. Again, experience and the results of research designed specifically to shed light on this set of issues can help to guide research study design. It is unclear what more can be said about this issue until we know a lot more about what characteristics of resources are likely to give rise to significant nonuse values. Some resources such as the Amazonian rain forest and African elephants may have worldwide significance, implying that the relevant population is the world population. In addition, one cannot rule out the possibility that where there are regionally significant resources, important nonuse values are held only by people within that region. This is an important research question.

Summary

This chapter has provided and reviewed the conceptual basis for empirical techniques that attempt to take advantage of market information. Where q is an input into the production of marketed goods and services, the techniques focus on the effects of changes in q on output and factor markets, not on the utility of q per se. It was also shown that there is a valid theoretical basis for examining changes in factor incomes such as land rents, costs savings in production, and changes in consumer surplus associated with the private good outputs.

When q is a consumption good and enters directly into the utility function, the problems are more severe. One of the more promising techniques is based on the concept of weak complementarity. Benefits are measured in terms of shifts in the demand curve for the private complementary good. The approach may be applicable to estimating recreation benefits due to the water quality changes.

Where the public good is either implicitly or explicitly an input, along with one or more market goods, in the household production of a final service flow, benefit estimation is straightforward if the substitution relationship or marginal rate of technical substitution is known, and the problem has a sufficiently simple structure so that unobservable utility terms cancel out. Defensive expenditures and measures of additional costs (for example, for household cleaning or for medical care and drugs in the case of health) are examples of estimates that are based on approximations of this approach.

Another promising case is where q varies across space, as in air pollution, or as a characteristic embodied in some private good. Then individuals can choose different quantities of q by varying residential locations or by choosing different private good models. Finally, the hedonic price approach can be used to measure the implicit price of q; and, under some circumstances, the demand curve for q can be identified.

In most of the models described in the first part of this chapter, it makes intuitive sense to speak of the individual as using the environmental or resource service in question. For example, in the models of substitution and averting behavior, the individual can be interpreted as using clean air to produce a clean household or good health. Thus, the values estimated with the models described here are sometimes referred to as use values. It is also possible for individuals to value environmental services independently of any use they might make of those services. The concept of so-called nonuse values is the topic of the last section of this chapter. It was shown there that, when nonuse values are defined and modeled in a plausible way, the result is that transactions in market goods reveal nothing about these nonuse values. Therefore, the revealed preference methods described in this chapter can shed no light on the possible magnitude of these values; and instead, we must rely on stated preference methods, which are the topic of Chapter 12.

References

Bartik, Timothy J. 1988. "Evaluating the Benefits of Non-marginal Reductions in Pollution Using Information on Defensive Expenditures." *Journal of Environmental Economics and Management* 15(1): 111–127.

Blackorby, Charles, Daniel Primont, and R. Robert Russell. 1978. *Duality, Separability, and Functional Structure: Theory and Economic Applications*. New York: North Holland.

Bockstael, Nancy E., and Catherine L. Kling. 1988. "Valuing Environmental Quality: Weak Complementarity with Sets of Goods." *American Journal of Agricultural Economics* 70(3): 654–662.

Bockstael, Nancy E., and Kenneth E. McConnell. 1983. "Welfare Measurement in the Household Production Framework." *American Economic Review* 73(4): 806–814.

Bockstael, Nancy E., and Kenneth E. McConnell. 1993. "Public Goods as Characteristics of Non-Market Commodities." *Economic Journal* 103(3): 1244–1257.

Bockstael, Nancy E., and Kenneth E. McConnell. 2007. *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models.* Dordrecht: Springer.

Bockstael, Nancy E., Kenneth E. McConnell, and Ivar Strand. 1991. "Recreation." In *Measuring the Demand for Environmental Quality*, edited by John B. Braden and Charles D. Kolstad. Amsterdam: North Holland.

Boyle, Kevin J., and Richard C. Bishop. 1987. "Valuing Wildlife in Benefit-Cost Analyses: A Case Study Involving Endangered Species." *Water Resources Research* 23(5): 943–950.

Bradford, David F., and Gregory G. Hildebrandt. 1977. "Observable Preferences for Public Goods." *Journal of Public Economics* 8(2): 111–121.

Brown, James N., and Harvey S. Rosen. 1982. "On the Estimation of Structural Hedonic Price Models." *Econometrica* 50(3): 765–768.

Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. 2003. "Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez Oil Spill." *Environmental and Resource Economics* 25(3): 257–286.

Colwell, Peter F., and Gene Dilmore. 1999. "Who Was First? An Examination of an Early Hedonic Study." *Land Economics* 75(4): 620–626.

Courant, Paul N., and Richard Porter. 1981. "Averting Expenditure and the Cost of Pollution." *Journal of Environmental Economics and Management* 8(4): 321–329.

Deaton, Angus, and John Muellbauer. 1980. *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.

DesJardins, Joseph R. 1999. *Environmental Ethics: Concepts, Policy, and Theory*. London: Mayfield.

Ebert, Udo. 1998. "Evaluation of Nonmarket Goods: Recovering Unconditional Preferences." *American Journal of Agricultural Economics* 80(2): 241–254.

Eom, Young-Sook, and Douglas M. Larson. 2006. "Improving Environmental Valuation Estimates through Consistent Use of Multiple Information." *Journal of Environmental Economics and Management* 52(1): 501–516.

Fisher, Ann, and Robert Raucher. 1984. "Intrinsic Benefits of Improved Water Quality: Conceptual and Empirical Perspectives." In *Advances in Applied Microeconomics*, edited by V. Kerry Smith and Ann Dryden Witte. Greenwich: JAI Press.

Freeman, A. Myrick, III. 1969. "Advocacy and Resource Allocation Decisions in the Public Sector." *Natural Resources Journal* 9(2): 166–175.

Freeman, A. Myrick, III. 1974. "On Estimating Air Pollution Control Benefits from Land Value Studies." *Journal of Environmental Economics and Management* 1(1): 74–83.

Freeman, A. Myrick, III. 2003. *The Measurement of Environmental and Resource Values: Theory and Methods*. 2d ed. Washington, D.C.: Resources for the Future.

Garcia, Philip., Bruce L. Dixon, James W. Mjelde, and Richard Adams. 1986. "Measuring the Benefits of Environmental-Change Using a Duality Approach – The Case of Ozone and Illinois Cash Grain Farms" *Journal of Environmental Economics and Management* 13(1): 69–80.

Goldman, S.M., and Hirofumi Uzawa. 1964. "A Note on Separability in Demand Analysis." *Econometrica* 32(3): 387–398.

Griliches, Zvi. 1961. "Hedonic Prices for Automobiles: An Econometric Analysis of Quality Change." In *The Price Statistics of the Federal Government*. New York: General Series No. 73, Columbia University Press for the National Bureau of Economic Research.

Hahn, Robert W., Paul C. Tetlock, and Jason K. Burnett. 2000. "Should You Be Allowed to Use Your Cellular Phone While Driving?" *Regulation* 23(3): 46–55.

Hanemann, W. Michael. 1988. "Three Approaches to Defining 'Existence' or 'Non-use' Value Under Certainty." Unpublished paper. Available at: www.escholarship.org/uc/item/5nb5n99r.

Harford, Jon D. 1984. "Averting Behavior and the Benefits of Reduced Soiling." *Journal of Environmental Economics and Management* 11(3): 296–302.

Harrington, Winston, and Paul R. Portney. 1987. "Valuing the Benefits of Health and Safety Regulations." *Journal of Urban Economics* 22(1): 101–112.

Herriges, Joseph A., Catherine L. Kling, and Daniel J. Phaneuf. 2004. "What's the Use? Welfare Estimates from Revealed Preference Models when Weak Complementarity Does not Hold." *Journal of Environmental Economics and Management* 47(1): 55–70.

Jamieson, Dale. 2000. *A Companion to Environmental Philosophy*. Maldon: Blackwell.

Katzner, Donald W. 1970. *Static Demand Theory*. New York: Macmillan.

Kopp, Raymond J. 1992. "Why Existence Value *Should* Be Used in Cost-Benefit Analysis." *Journal of Policy Analysis and Management* 11(1): 123–130.

Krutilla, John V. 1967. "Conservation Reconsidered." *American Economic Review* 57(4): 777–786.

Krutilla, John V., and Anthony C. Fisher. 1975. *The Economics of Natural Environments: Studies in the Valuation of Commodity and Amenity Resources*. Baltimore: The Johns Hopkins University Press.

Larson, Douglas M. 1991. "Recovering Weakly Separable Preferences." *Journal of Environmental Economics and Management* 21(2): 97–108.

Lazo, Jeffrey K., Gary H. McClelland, and William D. Schulze. 1997. "Economic Theory and Psychology of Non-Use Values." *Land Economics* 73(3): 358–371.

Mäler, Karl-Göran. 1974. *Environmental Economics: A Theoretical Inquiry*. Baltimore: The Johns Hopkins University Press for Resources for the Future.

Mäler, K-G., Inge Marie Gren, and Carl Folke. 1994. "Multiple Use of Environmental Resources: A Household Production Function Approach to Valuing Natural Capital." In *Investing in Natural Capital: The Ecological Economics Approach to Sustainability*, edited by AnnMari Jansson, Monica Hammer, Carl Folke, and Robert Costanza. Washington, D.C.: Island Press.

McConnell, Kenneth E. 1983. "Existence and Bequest Value." In *Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas*, edited by Robert D. Rowe and Lauraine G. Chestnut. Boulder: Westview Press.

McConnell, Kenneth E. 1998. "Does Altruism Undermine Existence Value?" *Journal of Environmental Economics and Management* 32(1): 22–37.

Milgrom, Paul. 1993. "Is Sympathy an Economic Value? Philosophy, Economics, and the Contingent Valuation Method." In *Contingent Valuation: A Critical Assessment*, edited by Jerry A. Hausman. Amsterdam: North Holland.

Mjelde, James W., Richard M. Adams, Bruce L. Dixon, and Phillip Garcia. 1984. "Using Farmers' Actions to Measure Crop Loss Due to Air Pollution." *Journal of the Air Pollution Control Association* 34(4): 360–364.

Neeliah, H., and B. Shankar 2010. "Ozone pollution and farm profits in England and Wales." *Applied Economics* 42(19): 2449–2458.

Neidell, Mathew. 2009. "Information, Avoidance Behavior, and Health: The Effect of Ozone on Asthma Hospitalizations." *Journal of Human Resources* 44(2): 450–478.

Palmquist, Raymond B. 1984. "Estimating the Demand for Characteristics of Housing." *Review of Economics and Statistics* 64(3): 394–404.

Palmquist, Raymond B. 2005. "Weak Complementarity, Path Independence, and the Intuition of the Willig Condition." *Journal of Environmental Economics and Management* 49: 103–115.

Randall, Alan, and John R. Stoll. 1983. "Existence Value in a Total Valuation Framework." In *Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas*, edited by Robert D. Rowe and Lauraine G. Chestnut. Boulder: Westview Press.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Perfect Competition." *Journal of Political Economy* 82(1): 34–55.

Shibata, Hirofumi, and J. Steven Winrich. 1983. "Control of Pollution when the Offended Defend Themselves." *Economica* 50(200): 425–438.

Smith, V. Kerry. 1991. "Household Production Functions and Environmental Benefit Estimation." In *Measuring the Demand for Environmental Quality*, edited by John B. Braden and Charles D. Kolstad. Amsterdam: North Holland.

Smith, V. Kerry, and H. Spencer Banzhaf. 2004. "A Diagramatic Exposition of Weak Complementarity and the Willing Condition." *American Journal of Agricultural Economics* 86(2): 455–466.

Sutherland, Ronald J., and Richard G. Walsh. 1985. "Effect of Distance on the Preservation Value of Water Quality." *Land Economics* 61(3): 281–291.

Van De Veer, Donald, and Christine Pierce. 1994. *The Environmental Ethics and Policy Book: Philosophy, Ecology Economics*. Belmont: Wadsworth.

Willig, Robert D. 1976. "Consumers' Surplus Without Apology." *American Economic Review* 66(4): 589–597.

Willig, Robert D. 1978. "Incremental Consumer's Surplus and Hedonic Price Adjustment." *Journal of Economic Theory* 17: 227–253.

Zivin, Joshua G., Matthew Neidell, and Wolfram Schlenker. 2011. "Water Quality Violations and Avoidance Behavior: Evidence from Bottled Water Consumption." *American Economic Review: Papers and Proceedings* 101(3): 448–453.

Chapter 5

Valuing Changes in Risk

Up to this point, the discussion in this book has been based, at least implicitly, on the assumption of perfect certainty. For example, individuals have been assumed to know what prices, income, and environmental quality will prevail over the relevant planning horizon. This clearly cannot be literally true; and for some questions, the essence of the matter is uncertainty about some event or condition, such as the probability of an accident or a natural disaster (say, a flood) occurring or the probability of contracting a serious illness following exposure to a chemical. The topic of this chapter is how to define and measure changes in economic welfare when uncertainty is an important characteristic of the economic world in which people are living and making choices.

The uncertainty faced by individuals can take many forms. For example, people could be uncertain about future prices of consumer goods, future income, or the returns on financial assets. However, since there is already an extensive literature on the welfare effects of price and income variability, those problems are not treated here. See, for example, Just, Hueth, and Schmitz (2004) and Rothschild and Stiglitz (1976). The uncertainty that is important in this chapter is that which individuals who are users or potential users of some environmental service flow or resource face. Individuals might be uncertain as to whether a specific environmental service flow or resource will be available for use in the future; they might be uncertain as to whether they will actually want to use some resource in the future; or they could be uncertain as to whether their exposures to environmental hazards and pollutants will actually result in illness or death.

These uncertainties raise two kinds of questions for policymakers. First, what form of welfare criterion is appropriate under conditions of uncertainty, or more specifically, what modifications to the Hicks–Kaldor potential compensation test are required? Second, when public policies change the uncertainties facing people, how are the resulting welfare changes to be measured? These two questions are connected, as we shall see, in that the choice of a welfare criterion has important implications for how welfare changes are to be measured.

Most of the analysis in this chapter is based on the now standard model of individual preferences under uncertainty, which in turn is based on the hypothesis that individuals maximize their expected utility. The theory is very attractive as a normative prescription for behavior and choice. However, there is now substantial evidence that individuals' choices frequently violate expected utility theory. For examples of this evidence, and discussions of its implications for the theory of preferences under uncertainty, see Kahneman and Tversky (1979), Arrow (1982), Thaler (1987), Machina (1990), and Kahneman (2011). Despite this evidence on the descriptive inadequacies of expected utility theory, its simplicity and elegance still make it attractive for expositional purposes.

In the next section, the standard theory is developed and used to define welfare measures in several contexts. Several alternative welfare indicators that can be derived from this model of preferences are discussed. In the second section, alternative aggregate welfare criteria and their implications for the way individuals' welfares change are described. In the third section, dynamic welfare measures that reflect the ability of individuals to gather information, learn, and plan their purchase decisions accordingly are introduced. In the fourth section, some models for the measurement of individuals' willingness to pay for risk-reduction and risk-prevention policies using revealed preference methods are presented. The implications of some forms of nonexpected utility preferences for welfare measurement are also examined. In the fifth section, the concept of option value is introduced and related to the welfare measures developed in the chapter.

Individual Uncertainty and Welfare

The term individual uncertainty refers to situations in which an individual is uncertain as to which of two or more alternative states of nature will be realized and is not indifferent as to which state actually occurs. In this discussion, individuals are assumed to assign probabilities to these alternative states of nature and these probabilities are assumed to be correct in the sense of summing to one and incorporating all available information. Thus, no distinction is made between risk (where probabilities are known) and uncertainty (where probabilities are unknowable). For a discussion of this often-cited distinction, which is often attributed to Frank Knight, see LeRoy and Singell (1987). This discussion also abstracts from questions of risk perception, how individuals make assessments of probabilities, and how these assessments are revised in light of new information. For introductions to some of the issues concerning risk perceptions and probability assessments, and for references to the relevant literature in psychology and economics, see DellaVigna (2009), Arrow (1982), Viscusi and Zeckhauser (2006), and Machina (1989).

Risk and Environmental Policy

Consider some environmental risk such as a chemical spill or pollution event, or some natural hazard such as a flood or an earthquake. Other examples of such risks include tornadoes and hurricanes, accidents at nuclear reactors, explosions in chemical plants, releases of toxic materials from hazardous waste storage sites, and air pollution episodes associated with unusual meteorological conditions. Risks such as these can be described in terms of two characteristics: the range of possible adverse consequences, and the probability distribution across consequences. In this analysis, adverse consequences are measured in units that reflect the consequences for people (for example, the number of buildings damaged and degree of destruction, or the number of days of illness) rather than in terms of measures such as height of flood stage, Richter scale reading, or maximum atmospheric concentration of a pollutant. To keep the exposition simple, consider only one possible adverse event and two states of nature: the event occurs with a given set of consequences, or the event does not occur.

Suppose there is some public policy action that has the effect of reducing the probability of the adverse event, reducing the severity of its consequences, or both. Examples of policy measures that reduce the magnitude of an adverse event include regulations requiring earthquake-resistant construction techniques and the building of public shelters for protection against tornadoes and hurricanes. These are often referred to as *risk-reduction* measures. Examples of policy measures that reduce the probability of an adverse event include regulations for nuclear reactor safety, standards for durable containment techniques at hazardous waste storage sites, and the construction of dams to control streamflow to reduce the probability of flooding. These are often referred to as *risk-prevention* measures.

The distinction between risk reduction and risk prevention may sometimes be arbitrary. For example, the benefits of wearing a seatbelt could be modeled in terms of the belt's effect on the severity of injury associated with accidents of unchanged probability, or in terms of its effects on the probabilities of experiencing injuries of various severities per mile driven. The choice of a modeling strategy may depend upon such pragmatic considerations as availability of data.

Many risk-reduction and risk-prevention measures are public goods in that they have the characteristics of nonrivalry and nonexcludability. The public good character of these measures means that a private market system will fail to provide them in efficient quantities. Thus, there is a case for government intervention to improve the efficiency of resource allocation. In order to determine whether risk-reduction and risk-prevention measures result in improvements in welfare, it is necessary to define and measure the benefits and costs of changes in risks.

Individual Preferences and Expected Utility

Turning to the characterization of individual preferences in situations involving risk, assume that an individual has a well-behaved preference ordering over bundles of goods $\boldsymbol{X}$, and that there is some adverse environmental event A over which that person has no control. The variable A measures the severity of the adverse event. Let $A = A^*$ represent the occurrence of the adverse event, and $A = 0$ represent the absence of the adverse event. A takes the value A^* with probability π and 0 with the probability $1 - \pi$. For any given state of nature, this preference ordering can be represented from an ex post perspective by $u(\boldsymbol{X},A)$ with $u(\boldsymbol{X},0) > u(\boldsymbol{X},A^*)$. Assume $u_A < 0$ and $u_{\boldsymbol{X}} > 0$, where subscripts indicate partial derivatives. To avoid unnecessary complexity, income M and prices $\boldsymbol{P}$ are assumed to be certain. Finally, assume that $u(\cdot)$ is the same in both states, but that preferences over $\boldsymbol{X}$ are state-dependent. The consumption bundle actually chosen can depend on the state of nature—that is, on whether $A = 0$ or A^*. This means that the marginal utility of $\boldsymbol{X}$ can vary across states of nature. The model can be easily generalized to many states and to make $u(\cdot)$ depend on the state of nature (for examples see, Graham 1981 and Freeman 1984b, 1985, 1989).

Given utility maximization, there is an ex post indirect utility function $v = v(\boldsymbol{P}, M, A)$ that shows the maximum attainable utility given $\boldsymbol{P}$, M, and A, where $\boldsymbol{P}$ is the vector of prices and M is income. It has the properties

$$\begin{aligned} &v_M > 0, \quad v_{MM} < 0, \\ &v_A < 0, \quad v_{AA} < 0. \end{aligned} \tag{5.1}$$

From here on the $\boldsymbol{P}$ term is suppressed, since prices are assumed to be unchanged through the analysis.

Assume that individuals know the magnitude of the adverse event A^*, the probability of its occurrence, and that there are no opportunities for individual protective activities—that is, no ways of spending money to reduce either π or A^*. Ehrlich and Becker (1972) referred to the latter activities as self-protection (for reducing π) and self-insurance (for reducing A^*) and analyzed individual behavior in the case of purely monetary losses. To the extent that self-protection and self-insurance activities are possible in the case of risks of nonmonetary losses, they may provide a basis for empirical estimation of the values of risk changes. This possibility is investigated in the section on revealed preference methods in this chapter.

Let D represent the monetary value of the damages caused by the event, given that the adverse event has occurred. D is the maximum sum of money the individual would give up to experience $A = 0$ rather than A^*, and is the solution to

$$v(M, A^*) = v(M - D, 0). \tag{5.2}$$

Thus, D is a form of compensating surplus (CS) measure of welfare change where the reference level of utility is the utility realized if A^* occurs.

People often have to make choices before the state of nature is revealed, in other words, ex ante choices. Let us assume that individuals make these choices so as to maximize their expected utility, where expected utility is defined as

$$E[u] \equiv \pi \cdot u(X, A^*) + (1 - \pi) \cdot u(X, 0). \tag{5.3}$$

The expected utility expression provides a basis for an alternative measure of the value of avoiding A^*, namely, the WTP ex ante. This state-independent payment is sometimes referred to as option price, or OP. Option price is defined as the maximum payment the individual would make to change from the status quo risk to a situation in which A^* would not occur. It is also a form of compensating surplus, but one where the reference point is defined in terms of expected utility. OP is the solution to

$$\pi \cdot v\left(M, A^*\right) + (1-\pi) \cdot v\left(M, 0\right) = v(M - OP, 0) . \tag{5.4}$$

OP and $\pi \cdot D$ will not, in general, be equal. This is because they are measuring two different things, or more precisely, measuring the monetary equivalents of two different forms of utility change: the ex post change in $v(\cdot)$ and the ex ante change in $E[v]$. The relationship between these two measures is the basis of the extensive literature on option value. A review of this literature is contained in the fifth section of this chapter. Also, these two measures are not the only possible ways of expressing a compensating measure of the welfare value of preventing the adverse event. This can best be seen with the aid of Graham's (1981) WTP locus.

The Willingness-to-Pay Locus

Given the assumption that individuals maximize expected utility, their behavior can be described as the solution to the following problem:

$$\text{Max} : E\left[u\right] \equiv \pi \cdot u\left(X, A^*\right) + (1-\pi) \cdot u\left(X, 0\right) , \tag{5.5}$$

subject to the usual budget constraint that expenditure equal income. The solution to this problem is denoted as E^1.

Now consider a policy that would reduce the magnitude of the adverse event from A^* to 0. If the individual were required to pay D to avoid the adverse effects given that the event occurs, and would pay nothing otherwise, the realized utility in each of the two states would be unchanged. Thus, one set of payments that would not change expected utility is (D, 0). Alternatively, if the individual were to pay OP before the uncertainty were resolved, by definition expected utility would also be unchanged. Thus, another pair of payments that leaves expected utility unchanged is (OP, OP). However, these are only two of an infinite number of possible pairs of payments that would leave the individual indifferent, in expected utility terms, between the status quo and a situation in which the consequences of the adverse event were avoided with certainty. The set of these payments is denoted as $\left(t^*, t^0\right)$:

t^* = payment given the state in which the adverse event occurs, and

t^0 = payment given the state in which it does not occur.

These payments satisfy the following condition:

$$E^1 = \pi \cdot v\left(M - t^*, A^*\right) + \left(1-\pi\right) \cdot v\left(M - t^0, 0\right) . \tag{5.6}$$

This equation defines Graham's WTP locus, which is shown in Figure 5.1.

Similarly, for a risk-prevention policy the WTP locus is given by the solution to

$$E^1 = E^2 = \pi' \cdot v\left(M - t^*, A^*\right) + \left(1-\pi'\right) \cdot v\left(M - t^0, 0\right) , \tag{5.7}$$

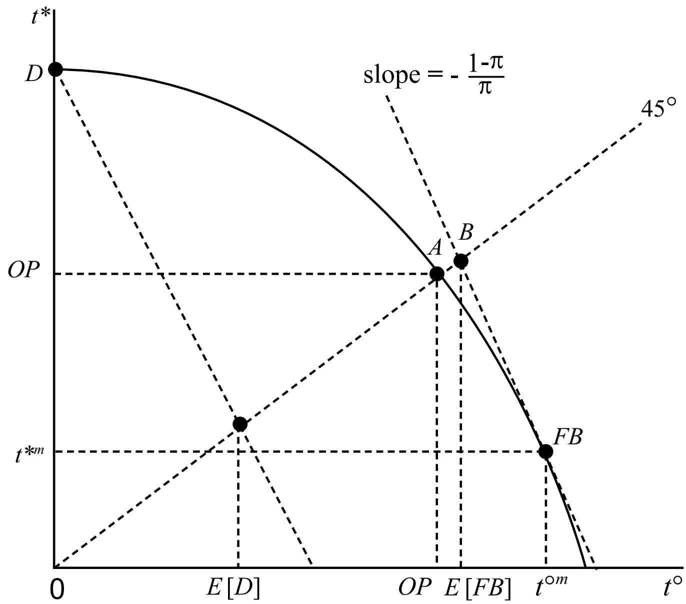

Figure 5.1 The willingness-to-pay locus

where π' is the probability of the adverse event with the risk-prevention policy in place. All of the subsequent analysis pertains to risk-reduction policies, but with suitable modification it can also be applied to the benefit-cost analysis of policies that change probabilities.

By setting the total differential of equation (5.6) equal to zero, we obtain an expression for the slope of the WTP locus:

$$\frac{dt^*}{dt^0} = \frac{(1-\pi)v_M^0}{\pi v_M^*}, \tag{5.8}$$

where v_M^* is the marginal utility of income evaluated at $M - t^*$ given the event has occurred, and v_M^0 is evaluated at $M - t^0$ given no adverse event. If the individual is risk-averse as assumed above ($v_{MM} < 0$), then the WTP locus will be concave to the origin—increasing t^* (for example) raises v_M^* relative to v_M^0. Risk neutrality, with its constant marginal utility of income, would result in a linear WTP locus.

Three points on the WTP locus are of particular interest as possible welfare measures. The first is the vertical intercept in Figure 5.1. This point involves a payment of zero if the event does not occur and there are no realized damages, and a payment of $t^* = D$ if the event does occur. This sum represents the maximum payment the individual would make to experience the consequence 0 rather than A^* given that the event occurs. This sum is also equal to the damages (D) from the event were it to occur. There is a line from this intercept with a slope of $-(1-\pi)/\pi$ representing the locus of all pairs of monetary values having the same expected value as D. Similar loci can be constructed for any t^*. These loci satisfy

$$E[t] = \pi \cdot t^* + (1-\pi)t^0 . \tag{5.9}$$

Differentiating, this gives

$$\frac{dt^*}{dt^0} = -\frac{(1-\pi)}{\pi} . \tag{5.10}$$

This line is one of a family of iso-expected payment lines. The point where the iso-expected payment line through point *D* intersects the 45° line can be used to identify the expected value of damages (*E*[*D*]). This is the ex post expected *CS* measure of the welfare change of preventing A^*.

A second point of interest in Figure 5.1 is *A*, which shows an alternative state-independent payment scheme in which $t^* = t^0 = OP$. In this example, the state-independent payment is greater than the expected value of damages, but this will not always be the case. The difference between *OP* and *E*[*D*] can be either positive or negative, and in some circumstances it can be quite large (see, for example, Graham 1981 and Freeman 1984b, 1985, 1989).

Finally, there is an iso-expected payment line tangent to the WTP locus at point *FB* in Figure 5.1. This point represents that state-dependent payment scheme, $\left(t^{*m}, t^{0m}\right)$, which maximizes the expected value of the individual's payments (*E*[*FB*]). The tangency of the WTP locus and the iso-expected payment line at point *FB* implies that the marginal utilities of income are equal in the two states. Since the slopes of the WTP locus and iso-payment lines are equal, making use of equation (5.7) yields

$$-\frac{(1-\pi)}{\pi} = -\frac{(1-\pi)v_M^*}{\pi \cdot v_M^0} , \tag{5.11}$$

and $v_M^* / v_M^0 = 1$. This equality is the condition for the efficient distribution of risk and for the optimum purchase of contingent claims at actuarially fair prices.[1] In this sense, point *FB* represents an optimum contingent payment scheme. In Graham's (1981) terminology, this is the fair bet point.

As illustrated in Figure 5.1, these three alternative payment schemes can give quite different summary measures of maximum willingness to pay (WTP) for the policy. This is because they differ with respect to the implied opportunities for insurance—that is, opportunities to alter the payments in different states of nature to reflect attitudes toward risk. In the example of Figure 5.1, the fair bet point allows the individual to make a larger payment in the state of nature in which no adverse event occurs and in which, therefore, the individual is otherwise better off.

In order to see the relationship between state-dependent payments of this sort and insurance, suppose that actuarially fair insurance is available to the individual. Specifically, suppose that the individual could make a payment *I* before the uncertainty was resolved to purchase an insurance policy that would

1 A contingent claim is a contract specifying in advance a set of payments or receipts, or both, in which the amounts depend on the state of nature. An insurance policy is a contingent claim.

pay R if the adverse event occurred. This means that the individual can exchange income across states of nature according to $I/(R-I) = k$, the price of insurance. If the price of insurance is actuarially fair, then $k = \pi/(1-\pi)$. In Figure 5.1, the insurance price line has the same slope as the iso-expected payment lines. The condition for the optimal purchase of insurance is, in this case, the equality of the marginal utilities of income in the two states of nature. This is because the individual chooses I (and therefore R) so as to maximize

$$E[u] = \pi \cdot v(M + R - I, A^*) + (1-\pi)v(M - I, 0), \tag{5.12}$$

subject to the constraint that $I/(R-I) = k$. Substituting the constraint into the objective function and differentiating gives the first-order condition

$$\frac{\pi}{(1-\pi)}\frac{v_M^*}{v_M^0} = k\ . \tag{5.13}$$

If the insurance is fair, then $k = \pi(1-\pi)$ and $v_M^* = v_M^0$.

Suppose now that the individual is required to make the payment represented by point D if the adverse event occurs and pays nothing otherwise. Paying D is equivalent to experiencing damages equal to D in this state of nature. Suppose also that the individual can purchase insurance at fair prices. The iso-expected payment line through point D shows that the individual could purchase an insurance contract that would require the individual to pay $E[D]$ if the adverse event did not occur and would reimburse him so that his net payment would be $E[D]$ if the adverse event did occur. Since this point lies inside the WTP locus, it results in a higher expected utility than could be realized in the absence of the policy that would reduce A^* to zero. This means that if the individual can also purchase insurance so as to adjust his state-dependent payments, he would be willing to pay more than D for the policy that reduces A^* to zero. By failing to take into account the individual's opportunities for diversifying risk through insurance, the expected damage measure of welfare change underestimates the true willingness to pay of the individual.

Now suppose that there is some institutional barrier to imposing a payment scheme that varies across states of nature, but that the individual can still purchase contingent claims at fair prices. The maximum state-independent payment that could be extracted from the individual ex ante is actually represented by point B in Figure 5.1. Since point B lies outside of the WTP locus, the individual would be left worse off with this payment scheme in the absence of fair insurance. But with fair insurance, the individual can move down to the right and reach the fair bet point through the optimal purchase of insurance. Thus, the availability of fair insurance makes the maximum expected payment of the fair bet point feasible even when varying payments across states of nature are not possible.

The existence of the WTP locus and points such as B (in Figure 5.1) that can be reached by purchasing contingent claims raises an important question. If welfare change is to be measured by the maximum payment that holds expected utility constant, which pair of state-dependent payments is the best welfare measure? We are not yet ready to answer this question, because how welfare is to be measured depends in part on the form of the social welfare criterion. The Potential Pareto Improvement (PPI) form of social welfare criterion requires a comparison of the welfare changes of those who gain and those who lose by the project. When losers face uncertainty about costs, the potential for state-dependent payments and compensation to redistribute risks between gainers and losers must be taken into account, along with the ability of each group to insure itself through the purchase of contingent claims. A full consideration of these issues must be left to the next section. However, first we must show how uncertainty about costs can be represented in Graham's state-dependent payment model.

The Required Compensation Locus

The costs of government risk-management activities come ultimately in the form of reductions in the utilities of individuals. This is true whether the activity involves direct government spending or regulations imposing requirements on the private sector. Furthermore, these costs, as correctly measured, may be quite different from the direct expenditures of the government or regulated firms. For an example, see Hazilla and Kopp (1990). The measurement of these costs from an ex ante perspective entails first determining the reductions in expected utility for all of those who would bear the costs and then finding the minimum compensation required to restore each person to her original level of expected utility in the absence of the government activity.

Any individual who would bear part of the cost of a project has a reference level of expected utility in the absence of the project, say

$$E^2 = \pi \cdot v\left(M, A^*\right) + \left(1 - \pi\right) v\left(M, 0\right). \tag{5.14}$$

The project would impose costs on the individual that could vary across states of nature. These costs could take several forms: a reduction in money income through taxation, the direct disutility of restrictions on nonmarket activities, or changes in product and factor prices.

If these costs vary across states, then compensation must be calculated so as to restore each cost-bearer's expected utility to the status quo level. Let c^* and c^0 represent all of the nonmonetary dimensions of costs and let M_c^* and M_c^0 represent money income inclusive of project-induced income changes. So that the problem can be analyzed graphically, suppose the probabilities affecting costs and benefits depend on the same random process in nature, so that π and $1 - \pi$ represent the relevant probabilities. This assumption will be relaxed later.

Given these assumptions, the project would result in a level of expected utility in the absence of compensation of

$$E^{2'} = \pi \cdot v\left(M_c^*, A^*, c^*\right) + \left(1 - \pi\right) v\left(M_c^0, 0, c^0\right), \tag{5.15}$$

where M_c^* and M_c^0 represent net income and c^* and c^0 represent the direct negative effects of the policy on utility in each state. There is an infinite number

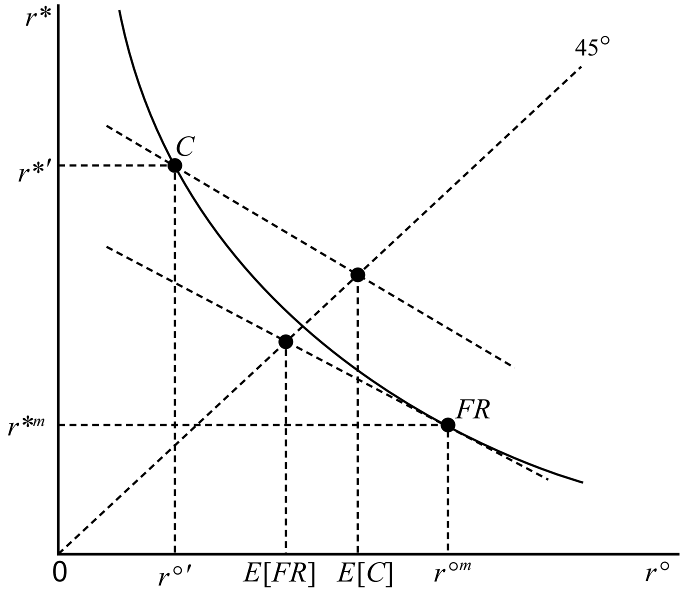

Figure 5.2 The required compensation locus

of pairs of state-dependent compensation payments, r^* and r^0, that can be made to restore expected utility to E^2. These pairs are solutions to

$$E^2 = \pi \cdot v\left(M_c^* + r^*, A^*, c^*\right) + \left(1-\pi\right)v\left(M_c^0 + r^0, 0, c^0\right). \tag{5.16}$$

The solution to this expression gives the required compensation (RC) locus for an individual who is harmed by the policy, shown in Figure 5.2. This is the locus of all possible compensating payments that leave those who bear the costs no worse off in terms of expected utility than if the policy had not been undertaken. The slope of the RC locus is found by setting the total differential of equation (5.16) equal to zero:

$$\frac{dr^*}{dr^0} = -\frac{\left(1-\pi\right)}{\pi}\frac{v_M^0}{v_M^*}. \tag{5.17}$$

Diminishing marginal utility of income assures that the locus is convex to the origin. The probabilities of the two states also define a family of iso-expected payment lines. They can be used to find the set of compensations that optimally distributes the risk to the cost-bearer across states—that is, that combination of payments that equates the marginal utilities of income in the two states. This is shown as $\left(r^{*m}, r^{0m}\right)$ at point *FR* in Figure 5.2. The intersection of the iso-expected payment line through this point with the 45° line gives the expected value of the fair compensation point, *E* [*FR*].

It is also possible to define an expected value of the required compensation from an ex post perspective. Given the occurrence of the event, there is an $r^{*\prime}$ that satisfies

$$v\left(M+r^{*'},A^{*},c^{*}\right)=v\left(M,A^{*}\right). \tag{5.18}$$

Similarly, $r^{0'}$ satisfies

$$v\left(M+r^{0'},0,c^{0}\right)=v\left(M,0\right). \tag{5.19}$$

These solutions define one point on the *RC* locus, labeled as *C* in Figure 5.2. Expected cost is

$$E\left[C\right]=\pi\cdot r^{*'}+\left(1-\pi\right)r^{0'}. \tag{5.20}$$

Finally, we can define a state-independent compensation analogous to *OP* (say, *OR*). Since it also holds expected utility constant, it is the solution to

$$E^{0}=\pi\cdot v\left(M+OR,A^{*},c^{*}\right)+\left(1-\pi\right)v\left(M+OR,0,c^{0}\right). \tag{5.21}$$

But, as will become clear in the next section, neither the expected value of cost nor *OR* will in general be useful as welfare measures for benefit-cost analysis.

Aggregation and the Welfare Criterion[2]

A major issue in the literature on benefit-cost analysis under uncertainty, at least since Weisbrod's classic article (1964), has been finding the appropriate measure of the welfare change for an individual who benefits from a project. The early option-value literature focused on comparing the expected value of consumer surplus with option price. Graham (1981) redirected the discussion by showing that there are an infinite number of alternative state-dependent payment vectors, and that the choice of a vector depends at least in part on the availability of contingent claims and individuals' opportunities to redistribute risk. Graham added a third candidate to the list of potential welfare measures, the expected value of the fair bet point; and he showed under what conditions the third candidate would be preferable to expected surplus and option price.

The literature since Graham has continued to focus on the question of the availability of contingent claims and whether risks are individual or collective. For examples of works dealing with this issue, see Mendelsohn and Strang (1984), Graham (1984), Cory and Saliba (1987), Colby and Cory (1989), and Smith (1990). For the most part this literature has ignored the cost side of the problem or treated it in a simplified fashion. In particular, most of this literature has neglected explicit treatment of uncertainty in costs, the possibility for changing the nature of risk through state-dependent compensation schemes, and the implications of the latter possibility for the form of the PPI criterion. An exception is the work of Meier and Randall (1991).

Also, for the most part this literature has dealt with only the simplest form of uncertainty (i.e., that where only two alternative states of nature are possible). Exceptions include Graham's 1981 and 1992 articles. The latter paper presented

2 This section is adapted, with permission, from Freeman (1991b).

a much more formal, abstract treatment of some of the problems considered here. The simple two-state model and its graphic treatment can be maintained only if both cost and benefit uncertainty result from the same random process and there are only two possible outcomes. If there are N possible benefit outcomes and J possible cost outcomes, the total number of alternative states of nature is $N \times J$.

In this section, alternative forms of the PPI criterion under uncertainty are discussed. A simple model with one gainer and one loser (so that the problems of aggregation and collective risk as opposed to individual risk are not important) is presented to show that none of the three candidates for welfare measurement is a reliable screen for potential Pareto improvements. This is because all three candidates ignore the opportunity for redistribution of risk between the gainer and the loser through state-dependent payments and compensation. The analysis is extended to a many-person economy. In those cases where the collective nature of risk precludes contingent claims markets, the importance of opportunities for redistributing risk between gainers and losers through state-dependent payment and compensation remains. Neither state-independent payments and compensation nor expected values of fair bet points are reliable indicators of potentially Pareto improving projects.

Potential Pareto Improvements under Uncertainty

Recall that the foundation of the PPI criterion is that a project passes only if there is some way to make redistributive payments so that no one is made worse off by the project and some people are made better off. It is generally agreed that in the context of risk and uncertainty "better off" and "worse off" are defined in terms of expected utility. However, there are several ways of formulating potential compensation tests in terms of expected utility. As shown above, there is an infinite number of potential payments and compensations for each individual that will leave expected utility unchanged. However, not all of these payment and compensation vectors are feasible in the aggregate, in the sense that whatever state of nature occurs, the potential payment equals or exceeds the required compensation. In other words, not all of the payment and compensation vectors pass an ex post balanced budget test. Here the PPI criterion is developed in a general form that is consistent with the balanced budget test. Then in the next subsection, the implications of this criterion for project selection and for welfare measurement in the simple two-state model are examined. The generalization to many states is straightforward. Finally, the aggregation of individual welfare measures across many gainers and losers is discussed.

A proposed policy can be deemed an improvement in PPI terms only if there is at least one set of state-dependent payments that will finance (in the sense of balancing the budget in all states of nature) a set of state-dependent compensations without making anyone worse off and leaving at least one person better off in expected utility terms. More formally, let $i = 1, \ldots, N$ index alternative outcomes of the random process affecting benefits;

let $j = 1, \ldots, J$ index alternative outcomes of the random process affecting costs; and

if $\boldsymbol{T}$ and $\boldsymbol{R}$ represent points on the N- and J-dimensional WTP and RC loci respectively, a project passes the PPI criterion if (and only if) there is at least one $\boldsymbol{T}$ and $\boldsymbol{R}$ such that $t_i \geq r_j$ for all i, j with the strict inequality holding for at least one pair.

Graphically, a project represents a PPI only if the WTP locus in the $(N \times J)$-dimensional space intersects the RC locus. Notice that since this expression of the PPI criterion compares t_i and r_j independently of the probabilities associated with alternative outcomes, expected values are irrelevant in screening for PPI projects. An equivalent procedure is to aggregate the positive and negative payments into a net WTP locus. If this locus has a positive segment in the $(N \times J)$-dimensional space, then the project passes the PPI test.

This version of the PPI criterion can pass projects that fail more restrictive forms of the criterion. This is because the criterion takes into account the possible benefits of redistributing risks between gainers and losers through state-dependent payments and compensation (Cook and Graham 1977). With this form of the criterion, project evaluation requires consideration of both the real effects of the project and project financing, because it is the latter that determines the benefits, if any, of redistributing risks.

Consider a project that would affect only two people. One person would gain, at least in one state of nature, and the other would bear the costs. In this simplification, the problem of how to aggregate benefits and costs within the beneficiary and cost-bearer groups is avoided. Assume that it is not possible to buy contingent claims on income across the two states of nature that are relevant. Should the project be built? The answer depends on which version of the PPI compensation test one adopts.

One form of compensation test was offered by Bishop (1986), who called it an Ex Ante Compensation Test. For the two-person economy considered here, this test asks whether the option price of the gainer exceeds the required compensation of the loser in every state of nature. Ready (1988) also discussed this criterion, calling it a "weak" form of potential Pareto improvement. The implications of this form of compensation test can be seen in Figure 5.3. Point A on the WTP locus shows the option price of the gainer. The ex ante compensation, or weak PPI criterion, will pass only those projects whose costs lie below and to the left of point A. Points B and C represent projects with uncertain costs. Projects B and C would not pass the test because if the adverse event occurred, the person bearing the cost would require compensation that exceeded the amount collected in the form of option price.

Ready (1988) also proposed an alternative, which he called the "strong" potential Pareto criterion. A project passes the strong form of this criterion if

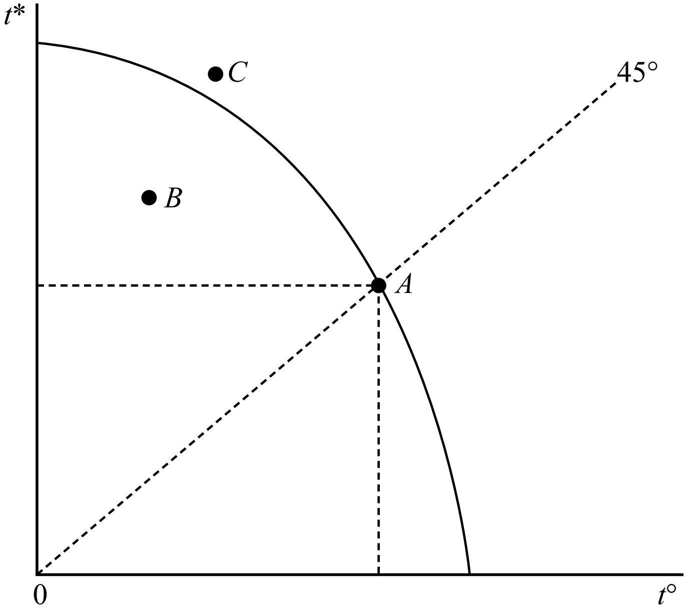

Figure 5.3 The willingness-to-pay locus and project costs

there exists some set of state-dependent payments that exceed the statedependent project costs in every state. The strong form of PPI criterion would pass a project with uncertain costs represented by point *B* in Figure 5.3, but it would reject a project whose costs were represented by point *C*. The strong form of the PPI criterion justifies the use of the expected value of the fair bet point as a benefit measure. If $E[FB]$ exceeds $E[FR]$, then the project passes.

Note that although both weak and strong forms of the criterion allow for cost uncertainty, they both treat costs as a point in the payment and compensation space. Both also ignore opportunities to redistribute risk among cost bearers or between gainers and losers through state-dependent payments and compensation.

Recall that according to the general form of the criterion stated above, a proposed policy is an improvement only if there is at least one set of balanced-budget, state-dependent payments that will finance a set of state-dependent compensations without making anyone worse off and leaving at least one person better off. In terms of Figure 5.3, the general form of PPI criterion would pass a project with costs represented by point *C* if the required compensation locus through point *C* intersected the WTP locus. As noted above, this form of the criterion takes account of both the real effects of the project and the redistribution of risks through project financing.

Ready (1993) has raised an interesting issue regarding the benefits of redistributing risk. The analysis here counted the benefits of redistributing risk whether or not the compensation is actually paid. Ready argued that they should be counted only if compensation is actually paid, since they result from creating a more efficient distribution of risk, and this is qualitatively different from project evaluation under certainty where project finance and compensation only involve the redistribution of wealth. He showed that it would be possible for a project that passes this form of PPI test because of its redistribution of risk benefits to make all parties worse off if the compensation were not in fact paid.

Applying the Criterion in a Two-Person World

Again consider the simplest case of a project that would affect only two people and where it is not possible to buy contingent claims on income across the two states of nature. To provide a specific example, suppose that the project is to divert water from a river through an existing generator to produce hydroelectric power for a beneficiary. The cost would be borne by a farmer who would otherwise use the water for irrigation. In a rainy year, the benefit would be large because of the large flow through the generator. The cost to the farmer would be small because the rainfall would substitute for the irrigation water. Similarly, in a dry year the benefit would be small and the cost to the farmer large.

Should the project be built? Since the benefits and the costs are subject to the same random process, the WTP and RC loci can be combined in a single diagram, which is shown in Figure 5.4. Points *B* and *C* show the state-dependent

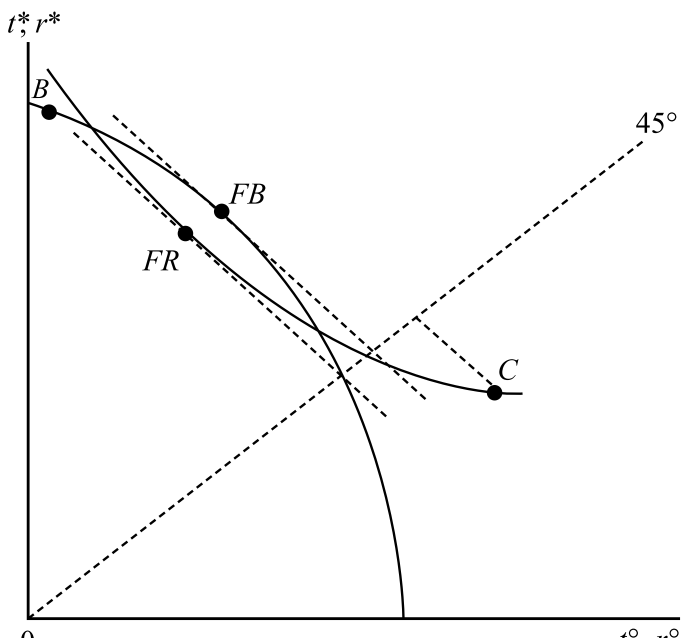

Figure 5.4 A potential Pareto improving project with expected costs exceeding expected benefits

benefits and costs, and the WTP and RC loci pass, respectively, through these two points. Another way of stating the PPI criterion is that it asks whether there is a set of state-dependent transfers from the beneficiary to the cost-bearer such that neither party is made worse off and at least one of the parties gains. Any state-dependent transfer is represented by one point in the WTP-RC space. Any such point on or below the WTP locus leaves the beneficiary no worse off; if the point is on or above the RC locus, the costbearer is no worse off. Thus, any set of transfers that lies in the lens-shaped area between the two loci will successfully finance this project, and therefore, the project passes the PPI test.

There are several important points to note about this conclusion. First, neither the expected value of cost nor the expected value of benefit is relevant. In the example of Figure 5.4, the expected value of cost substantially exceeds the expected value of benefit. But the project still passes the PPI test because there are feasible payment schemes that redistribute the risks associated with the states of nature in a way that can make both parties better off.

The irrelevance of expected benefit and expected cost holds even if benefits and costs are certain, as long as individuals are risk-averse. For example, suppose that points *C* and *B* in Figure 5.4 are both on the 45° line, with *C* above *B*. It is still possible that the aggregate WTP and RC loci through these points could intersect over some range. This demonstrates the importance of considering the manner in which the payment and compensation scheme redistributes risk in project evaluation. For example, suppose that both individuals faced substantial uncertainty about their incomes. The beneficiary would prefer a payment scheme with the larger payment in the state of nature in which income was high. The cost-bearer would prefer to receive more compensation in the low-income state. Project financing provides both individuals with a means of hedging the income uncertainty. The risk of the project also has to be evaluated in the context of both individuals' total risk portfolios.

The second point about the conclusion above is that option prices are not reliable indicators of PPI projects. Again, the example in Figure 5.4 shows that it is possible for the state-independent compensation to exceed the state-independent willingness to pay and yet have the project pass the PPI test because of the ability of state-dependent payments and compensation to redistribute risk in a more favorable manner.

The third point is that neither the expected value of the fair bet point nor the expected value of fair compensation is relevant to project evaluation. Although the potential Pareto improving project in Figure 5.4 has an expected value of the fair bet point exceeding the expected value of compensation, this condition is not sufficient to assure that the project is potentially Pareto improving. A case in which the expected value of the fair bet point exceeds the expected value of fair compensation is shown in Figure 5.5. Yet, because the WTP locus lies below the RC locus everywhere, there is no state-dependent transfer scheme that would make this project potentially Pareto improving. Although on average (that is, in an expected value sense) WTP exceeds RC, there are some states of nature in

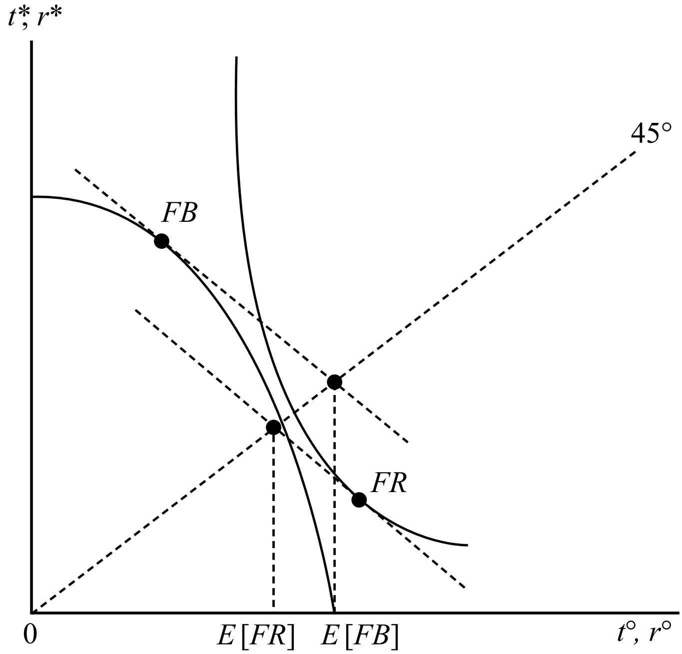

Figure 5.5 A nonpotential Pareto improving project

which the funds for the required compensation could not be collected from the beneficiary.

These results can be summarized in the form of three statements about necessary and sufficient conditions for identifying PPI projects.

1 The expected value of benefits being greater than the expected value of costs is neither necessary nor sufficient for a PPI project.
2 The option price of beneficiaries being greater than the option price of cost-bearers is sufficient but not necessary for a PPI project.
3 The expected value of the fair bet point exceeding the expected value of the fair compensation point is necessary but not sufficient to identify a PPI project.

Aggregation in the Many-Person Economy

In the many-person economy the principal issue to be addressed is how the WTP and RC loci of individuals should be aggregated for the purposes of social welfare analysis. The answer depends on the nature of the risks faced by gainers and losers—that is, whether risks are collective or individual in nature. Collective risk refers to the case in which, if the event affecting benefits and costs occurs, all of the potentially affected individuals experience the event. Individual risk refers to a case in which the risks facing different individuals are independent.

Consider the case of individual risk first. With individual risks, the analysis of aggregate benefits and costs is much simplified. Let one individual who would be a beneficiary have a probability of π^b of experiencing the event. Then if N is the number of potential beneficiaries, $\pi^b \cdot N$ individuals will experience the event with virtual certainty, with sufficiently large N. Similarly, $\pi^c \cdot K$ of the K bearers of the costs will experience the cost state associated with the event with virtual certainty, with sufficiently large K. Although individuals are uncertain as to their own outcomes, aggregate payments and compensation can be calculated with virtual certainty.

Assume for the moment that all people are identical. If those cost-bearers who experience the event are compensated by $r^{*\prime}$ and the rest of the cost-bearers are compensated by $r^{0\prime}$, then the aggregate compensation is

$$\pi^c \cdot K \cdot r^{*\prime} + \left(1-\pi^c\right)K \cdot r^{0\prime} = K\left[\pi^c \cdot r^{*\prime} + \left(1-\pi^c\right)r^{0\prime}\right]. \tag{5.22}$$

The aggregate compensation is minimized by finding the expected fair compensation for an individual and multiplying by K. The same procedure is used to find the maximum aggregate willingness to pay. If

$$N\left[\pi^b \cdot t^{*M} + \left(1-\pi^b\right)t^{0m}\right] > K\left[\pi^c \cdot r^{*\prime} + \left(1-\pi^c\right)r^{0\prime}\right], \tag{5.23}$$

then the project passes the PPI test. If individuals are different, all individuals' fair bets and compensations must be found and summed. In general, if each gainer makes the state-dependent payments represented by her fair bet point, the aggregate payment is the sum of the expected values of the fair bet points. Results are similar for the losers. Also, there is a virtual certainty that the gainers can compensate the losers and still be better off.

As long as a social welfare judgment is made not to require actual payment and compensation, this is all that needs to be said on the matter. However, if a social welfare judgment is made to require payment and compensation, then one must address the question of the feasibility of the required state-dependent payment and compensation schemes. If it should turn out that state-dependent payment and compensation is not feasible, then individual option prices are relevant, unless contingent claims could be purchased at actuarially fair prices. In the case of individual risks, it is more likely that such contingent claims will in fact be available.

In the case of collective risk, the full aggregate WTP and RC loci are required for welfare analysis. The procedure for aggregating individuals' WTP loci is explained in Graham (1981, 718–719). For each possible marginal rate of substitution (MRS) between contingent payments in different states as measured by dt^*/dt^0, each individual's payment vector is determined and all of these payment vectors are summed to obtain

$$T^{*m} = \sum_i^N t_i^{*m} \text{ and } T^{*0} = \sum_i^N t_i^{*0}, \tag{5.24}$$

where i indexes the N beneficiaries, and the superscript m indicates that all individuals' payments are at their fair bet points for a given common MRS. This procedure gives one point on the aggregate WTP locus for each MRS. Each point also represents an efficient distribution of risk within the group of beneficiaries since all individuals have the same MRS for contingent payments. A similar procedure of aggregating individuals' required compensation at different marginal rates of substitution will yield the aggregate RC locus. Each point on the RC locus corresponds to an efficient distribution of cost risk among cost-bearers.

A project passes the potential Pareto improvement test if there exist payment and compensation points, $T^* = R^*$ and $T^0 = R^0$, that lie both on or above the aggregate RC locus and on or below the aggregate WTP locus. The graphic analysis is similar to that in the two-person world shown in Figures 5.4 and 5.5. Aggregate option price is sufficient but not necessary as a screen for PPI projects; and the aggregate fair bet points are necessary but not sufficient as indicators of PPI projects.

There are two exceptions to these conclusions. First, if contingent claims could be purchased at actuarially fair prices, the expected values of the fair bet and fair compensation points do provide a sufficient test for potential Pareto improvement. This is because individuals' actual payments and compensations can be made at the expected values of the fair bet and fair compensation points. Individuals can then adjust their own positions through contingent claims purchases to reach their individual WTP or RC locus at their fair bet point. However, the nature of collective risks is likely to preclude a market for contingent claims at actuarially fair prices.

Second, if a social welfare judgment is made to require compensation, and if state-dependent payments and compensations are not feasible, then as is the case in the two-person world, the welfare criterion must be based on option price measures. If individuals are identical, the appropriate test is based on the aggregate OP^B and OR^C because they are equal to the sums of the individual *OP*s and *OR*s. But if individuals are different, then the sums of the individuals' option price measures will be less than the aggregate measures. This is because the former do not provide for the efficient distribution of risk across individuals within each group. The appropriate welfare test becomes

$$\sum_{i}^{N} OP_i > \sum_{j}^{J} OR_j \ . \tag{5.25}$$

In this section, some of the consequences of extending the model of benefit-cost analysis for uncertain projects to reflect the symmetry of benefits and costs as changes in the expected utilities of affected individuals have been studied. One of the implications of this analysis is that there is much work still to be done on the development of measures of benefits and costs in order to implement the appropriate welfare criteria. Specifically, since presently available methods of benefit and cost estimation under uncertainty generally give only expected cost and benefit and option price measures, it is necessary to develop new methods for estimating the WTP and RC loci.

Uncertainty and Welfare in a Dynamic Setting

The derivation of option price and Graham's WTP locus assumes that individuals are faced with some type of uncertainty, but that they do not have the opportunity to learn and reduce that level of uncertainty over time and then decide whether to purchase the good. Likewise, these models do not consider the possibility that individuals might have the opportunity to reverse their decisions if, after receiving more information, they decide that the good was not worth the price paid. However, learning and the ability to delay or reverse a decision once made, are part of many real world decisions with respect to both private goods and public goods. If through the acquisition of information, an individual can learn more about a good and reduce or eliminate the uncertainty associated with making a decision, then that individual might choose to delay the decision. The value of delaying a decision until additional information becomes available is the motivation for the concept of quasi-option value. Quasi-option value is a term coined by Arrow and Fisher (1974) to describe the welfare gain associated with delaying a decision when there is uncertainty about the payoffs of alternative choices, and when at least one of the choices involves an irreversible commitment of resources (or more generally, positive costs associated with reversal).

Much of the early literature concluded that consideration of quasi-option value would lead to relatively less irreversible development and relatively more preservation of natural environments (Arrow and Fisher 1974; Conrad 1980; Miller and Lad 1984; Fisher and Hanemann 1987; Mäler and Fisher 2005). In these models quasi-option value stems only from the value of the information gained by delaying an irreversible decision (such as developing a historically pristine natural area). But it is not difficult to imagine situations where the relevant information to guide future decisions can be gained only by undertaking at least a little development now. For example, suppose there is uncertainty about the magnitude of a mineral deposit underlying a wilderness area. Perhaps the only way that the uncertainty about the magnitude of the benefits of development relative to preservation can be resolved is through exploratory drilling—that is, through a little bit of development. In such cases there can be positive quasi-option value to development, or equivalently, a negative quasi-option value to preservation (Miller and Lad 1984; Freeman 1984a; Kolstad 1996). Thus, depending on the sources of learning and degrees of irreversibility, it can make sense to delay or hurry a project (Pindyck 2000, 2002; Balikcioglu, Fackler, and Pindyck 2011); it can also make sense to adjust the size of the project in response to the magnitude of the uncertainty (Zhao 2001, 2003).

While the quasi-option value literature suggests that there are times when the presence of uncertainty can mean that the timing or size of a project should be altered for efficiency decisions under uncertainty, Zhao and Kling (2004, 2009) demonstrated that these same factors can alter the magnitude of the willingness to pay or sell for a good at any given point of time. Rather than considering the effect of uncertainty and learning opportunities on the optimal timing of a decision, they considered the maximum price an individual would be willing to pay *now* when they know that by purchasing today they forgo the opportunity to learn and reduce the risk of making a "bad" purchase decision. An individual faced with uncertainty about the value of a good might still be willing to make a purchase if the price is low enough to compensate for the loss of the opportunity to acquire more information and make the decision later. This intuition forms the basis for the concept of dynamic willingness-to-pay measures.

As an example, consider an individual who is considering purchasing an environmentally-friendly vehicle (perhaps a dual-fueled or electric car). Suppose that the individual is willing to pay a premium to be recognized as a green-conscious consumer, but only if the fuel and maintenance costs of the vehicle are not too high. Given the newness of the vehicle, the lifetime fuel and maintenance costs may be highly uncertain. This consumer faces a dynamic decision problem. She can decide to purchase the vehicle in the current period and thereby enjoy the benefits of being a green consumer now and for all future periods, or she can delay the decision and wait until more information is available on the variable costs of maintaining and running the vehicle. A delay would mean that the consumer does not get the benefits of being a green consumer in the current period. In forming her maximum willingness to pay for the green vehicle, the consumer will consider both the risk of making a decision she will later regret with the lost benefits of not obtaining the good for consumption this period. A low enough price in the current period will make the tradeoff acceptable.

Consider another example: that of greenhouse gas emissions and climate change. Much of the political debate about the need for addressing global warming relates to the uncertainties associated with the extent of climate change that is likely to occur under a range of human behaviors, and what the associated damages from those changes are likely to be. This uncertainty has led some to argue that society should delay investments in climate change mitigation until this uncertainty is resolved, or at least reduced, thereby lowering the risk of incurring costs now that later turn out to have been unnecessary. Implicit in this argument is that by waiting to obtain more information, the costs from investing in greenhouse gas reducing technologies can be avoided without any increased damages from climate change. Alternatively, a number of authors have argued that delay could ultimately lead to *higher* damages in the future. Invoking the concept of the "precautionary principle," they argue it may make sense to proceed more quickly to address climate concerns than if no uncertainty were present (an argument more akin to the early quasi-option value examples). Thus, depending on the source and magnitude of uncertainty one considers, climate change is a case in which quasi-option values may lead to either a delay, or a speeding up, of the decision to implement a policy of project. Either way, the maximum willingness to pay to reduce greenhouse gas emissions in the current period may be affected by the possibility of obtaining additional information about future damages.

To study the conditions under which dynamic welfare values will differ from their static counterparts (i.e., option prices), consider a simple model based on Zhao and Kling (2009) with three time periods (today and two future periods). The individual's per period utility function is $v(x, M, \theta)$, where x is a public good, M is a composite good (equivalent to income in this situation), and θ is a parameter that affects the value of an improvement in the public good from an initial level, x_0, to an improved level, x_1. In the climate change case, θ might represent the degree of warming or variability in storm events and x might represent a flood prevention structure (which would prevent flood damages associated with higher rainfall and more severe storm events associated with climate change). Suppose that the project takes one period to build and that while the value of θ is uncertain today (period 1), its value will be known with complete certainty at the beginning of period 2 (imagine that the definitive climate change study will be completed and released precisely at the beginning of the period). For simplicity, assume that there are two possible outcomes for $\theta_i = (\theta_L, \theta_H)$. If θ_H is realized, then the flood prevention project would have been highly valuable, but if θ_L is realized then there is little or no value to the project. In this situation, how much would the consumer be willing to pay for the flood prevention project if she must commit today to an annualized payment in all future periods once the project is finished?

To answer this question, it will be instructive to consider two cases, one in which the consumer cannot take advantage of the opportunity to delay and learn (the no-learning case), and the other in which it is possible to delay and wait until the uncertainty is resolved before making a decision (the learning case). In the no-learning case, the most an individual would be willing to pay is the amount that equates the expected value of purchasing the good in the first period and having it available for consumption in the second and all future periods with the utility of not having the good in those periods, and paying nothing. This results in a per period payment (starting in period 2) of p_{nl}, where p_{nl} is defined implicitly by

$$\begin{aligned}&\tfrac{1}{r}\left[\pi v(x_0, M, \theta_H) + (1-\pi) v(x_0, M, \theta_L)\right] = \\ &\tfrac{1}{r}\left[\pi v(x_1, M - p_{nl}, \theta_H) + (1-\pi) v(x_1, M - p_{nl}, \theta_L)\right],\end{aligned} \tag{5.26}$$

and both sides are divided by r, the discount rate, to reflect the fact that this stream of utility occurs in perpetuity. The reader will quickly recognize that the WTP value, p_{nl}, is simply the option price. Thus, when delay and/or learning is not possible, there is an equivalence between the option price and the dynamic WTP.

To consider how the WTP might change when delay and learning is possible, it is useful to realize that by waiting, the individual can eliminate the downside risk of making the "wrong" decision (paying a positive price for x_1 in the event that θ_L is realized, or not having x_1 when θ_H is realized). For simplicity, assume that if $\theta_i = \theta_L$ there is no value from the project and therefore the ex post WTP would be zero. Likewise, assume that if $\theta_i = \theta_H$ there is a range of positive prices that would generate positive surplus.

The expected value of purchasing the good in the current period is the sum of the expected values of the current period (without the good) plus the expected value of the utility of receiving the good in the second and all future periods:

$$\begin{aligned}v_0(p) = {} &\pi v(x_0, M, \theta_H) + (1-\pi) v(x_0, M, \theta_L) + \\ &\tfrac{1}{r}\left[\pi v(x_1, M - p, \theta_H) + (1-\pi) v(x_1, M - p, \theta_L)\right].\end{aligned} \tag{5.27}$$

If instead of purchasing the good in the first period, the individual delays and waits to observe the realized value of θ_i, he will be able to purchase the increase in x only if its value is positive (i.e., if $\theta_i = \theta_H$). Essentially, he eliminates the second term in brackets in (5.27), which is the lower utility associated with paying a positive price x_1 when $\theta_i = \theta_L$, and is instead guaranteed that he will only purchase x_1 when it generates positive surplus. But, this means that he will not be able to enjoy x_1 in the second period if its value is high. The expected utility of delaying until the uncertainty is resolved can be written as

$$\begin{aligned} v_1(p) &= \left(1+\frac{1}{1+r}\right)\left[\pi v(x_0,M,\theta_H)+(1-\pi)v(x_0,M,\theta_L)\right]+ \\ &\quad \left(\frac{1}{r(1+r)}\right)\left[\pi v(x_1,M-p,\theta_H)+(1-\pi)v(x_0,M,\theta_L)\right] \\ &= \left(1+\frac{1}{r}\right)\left[\pi v(x_0,M,\theta_H)+(1-\pi)v(x_0,M,\theta_L)\right]+ \\ &\quad \left(\frac{1}{r(1+r)}\right)W(p) \end{aligned} \tag{5.28}$$

where $W(p)=\pi\left[v\left(x_1,M-p,\theta_H\right)-v\left(x_0,M,\theta_H\right)\right]>0$ is the value associated with being able to wait and obtain the new information. The first line of the equation represents the expected utility from the first and second period when the project has not been built. The second line represents the expected utility from building the project only when it makes sense to build it (i.e., when $\theta_i = \theta_H$), and not building it when it has no value (i.e., when $\theta_i = \theta_L$). The second term in brackets incorporates the value of waiting $W(p)=\pi U(x_1,M-p,\theta_H)+(1-\pi)U(x_0,M,\theta_L)$, and it is enjoyed in perpetuity, which is why the term is divided by r in equation (5.28).

To determine the maximum WTP of an individual who has the option of waiting, solve for the price that equates the right-hand side of equations (5.27) and (5.28) which, after some rearranging, yields the expression

$$\begin{aligned} &\tfrac{1}{r}\left[\pi v(x_0,M,\theta_H)+(1-\pi)v(x_0,M,\theta_L)\right]= \\ &\tfrac{1}{r}\left[\pi v(x_1,M-p_l,\theta_H)+(1-\pi)v(x_1,M-p_l,\theta_L)\right]-\left(\frac{1}{r(1+r)}\right)W(p_l). \end{aligned} \tag{5.29}$$

The value of p_1 (the price when learning is an option) represents the WTP for the improvement in x that an individual would be willing to commit to in the current period when they know that by not doing so they can wait until the uncertainty is resolved and only purchase the good if it is valuable. The expression for p_1 as written in (5.29) has the same left-hand side expression as does the expression for p_{nl} in (5.26) making comparison straightforward. In forming the WTP when learning is not available (p_{nl}), the individual will set the expected utility of not purchasing the good (left-hand side of both expressions) with the expected utility of purchasing it as expressed in the right-hand side of (5.26). But, in forming the dynamic WTP, the individual will equate the expected utility of not purchasing the good with the expected utility of waiting and deciding whether to purchase the good in the second period. The two components of the right-hand side of (5.29) represent this value. Since $W(p)$ is positive, the right-hand side of (5.29) is lower than the right-hand side of (5.26). This, in turn, implies that the solution for p_{nl} in (5.26) must be greater than p_1 in (5.29); that is, $p_{nl} > p_1$. By committing to purchase the good now, a consumer gives up the option to learn and delay until better information is available. To be induced to do so, the consumer will need to be "compensated" by paying a lower price today than otherwise. It is important to emphasize that the lower price refers to the price the consumer would be willing to commit to today, but could be either higher or lower in the second period since it will be conditional on the realized value of θ.

For simplicity, the expressions derived in (5.26) to (5.29) were restricted to state-independent payments, but as Zhao and Kling (2009) showed, this can be generalized to a set of state-dependent payments so that a full locus of payments can be derived just as in Graham's locus. They demonstrated that in general, the dynamic WTP locus will lie below the static (no-learning) locus of Graham, except for sharing a common value at the "certainty" point.

There are a number of implications of understanding that welfare measures may be dynamic and conditional on expectations about being able to delay a decision until more information is available. These relate to both the use of welfare measures in benefit-cost analysis and the interpretation of those measures, particularly those generated by stated preference methods. The appropriate welfare measure for use in deciding whether a project should be built when learning and delay are possible is the dynamic WTP measure. However, if the project is not built it will be necessary to reconsider the questions when additional information has become available and has reduced the uncertainty. Just because a project does not pass a benefit-cost analysis with the current information set does not mean that it will not necessarily pass such a test in the future. In essence, benefit-cost analysis needs to become a dynamic process in this situation, where the decision-maker may need to revisit the decision once increased information becomes available over time. On a related point, it is important to recognize that dynamic WTP measures incorporate considerations about delay and uncertainty, and it would be a form of double counting if quasi-option values were calculated in addition to using dynamic WTPs in evaluating the efficiency of a project.

It is also important for analysts to understand whether the value they have elicited from a welfare measurement study contains considerations of delay and learning so that it can be interpreted and used appropriately. For example, if an analyst designs a stated preference survey to elicit respondents' WTP for a dam project that could be delayed until additional information is available concerning environmental effects of the project and the analyst has carefully communicated this information to respondents, then a comparison between the costs and estimated benefits should yield the correct efficiency decision. However, if respondents answer the question implicitly assuming that the dam can be built in future years once more information is available, when in fact the opportunity to purchase the land for the dam is possible only in the near term, then the analyst might end up comparing a welfare measure that is based incorrectly on the assumption that delay and learning is possible. Such a measure would be an underestimate of the true WTP.

Zhao and Kling (2001) also argued that one source of the often-observed disparity between WTP and WTA measures of value can be explained by the dynamic nature of these values. Specifically, suppose consumers perceive that the loss of an environmental good (e.g., a pristine wilderness area or large tract of open space) is difficult to reverse, but that a decision to allow development is always possible in the future. This may lead to large values of WTA because, in this case, consumers would need an extra large payment to be compensated for the loss of the chance to learn more about the value of the pristine wilderness and allow development. More discussion of the disparity between WTP and WTA and possible explanations for it appears in Chapter 3.

Revealed Preference Methods for Measuring Values

Stated preference methods will often be the method of choice for estimating ex ante WTP or WTA values under uncertainty, but there are some cases in which revealed preference methods may prove useful (e.g., Desvousges, Smith, and Fisher 1987; Cameron and Englin 1997; Cameron 2005; Nguyen et al. 2007). Specifically, when individuals have opportunities to adjust to risky positions through transactions in related private goods markets, it may be possible to use one of the revealed preference methods described in Chapter 4 to infer individuals' values for risk changes. In the situation where learning and delay are possible, the values that are estimated will correspond to the dynamic measures described in the previous section. Otherwise, the values will correspond to conventional option prices.

In this section, models of averting behavior and hedonic prices that can be used to obtain measures of value based on observable behavior are described. The models developed in this section are initially based on the hypothesis of expected utility maximization. Then the models are extended to take account of other assumptions about individuals' preferences. Since most observable transactions involve ex ante state-independent payments, the models described here are useful only in estimating option price measures of value.

Expressions for Option Price and Marginal Willingness to Pay

Suppose public policies are under consideration that will reduce either the magnitude of an adverse event or its probability. For simplicity, assume that learning and delay are not possible so that the policymaker is interested in the option price for a policy that reduces the magnitude of an adverse event—that is, the set of state-independent payments that results in the same level of expected utility as would have occurred in the original uncertain situation. Denote this payment as OP^A to indicate that it refers to a risk-reduction policy. Making use of the indirect utility function, OP^A for a policy that reduces A from A^* to zero is the solution to

$$\begin{aligned}
&\pi \cdot v\left(M, A^*\right) + (1-\pi) v(M, 0) \\
&= \pi \cdot v\left(M - OP^A, A^*\right) + \ 1 - \pi \cdot v\left(M - OP^A, 0\right) \\
&= v\left(M - OP^A, 0\right).
\end{aligned} \tag{5.30}$$

Similarly, the value of reducing A from A^* to $A' > 0$ is the solution to

$$\begin{aligned}
&\pi \cdot v\left(M, A^*\right) + (1-\pi) v(M, 0) \\
&\quad = \pi \cdot v\left(M - OP^{A'}, A'\right) + (1-\pi) v\left(M - OP^{A'}, 0\right).
\end{aligned} \tag{5.31}$$

For some purposes, it will be more useful to deal with marginal values. The option price for a marginal reduction from A^* is the marginal change in income that holds expected utility constant. This can be found by taking the total differential of the expression for expected utility in its indirect utility function form, setting it equal to zero, letting $dw = 0$, letting $dA = 0$ for $A = 0$, and rearranging terms. The individual's marginal WTP ex ante is given by

$$w^{A*} = \frac{dM}{dA^*} = -\frac{\pi \cdot v_A^*}{\left[\pi \cdot v_M^* + (1-\pi) v_M^0\right]}, \tag{5.32}$$

where * and ^{0} indicate the values of A (i.e., A^* versus $A^0 = 0$) at which the partial derivatives are evaluated. Recalling that $v_A < 0$, this expression is positive for reductions in A^* ($dA^* < 0$), indicating a positive willingness to pay to reduce the magnitude of the uncertain event.

Equation (5.32) shows a variant of a standard result from welfare theory. The marginal willingness to pay for a change in A^* is equal to the marginal disutility of A^* converted to a money measure by using the marginal utility of income. But since this is an ex ante willingness to pay, the marginal utility terms are the expected values of the relevant marginal utilities—that is, the weighted averages of the marginal utility terms in the two states of nature where the weights are the probabilities of the two states.

The option price for reducing the probability of the event to zero OP_π^0 is found by solving the following expression for

$$\begin{aligned}
&\pi \cdot v\left(M, A^*\right) + (1-\pi) v(M, 0) \\
&\quad = 0 \cdot v\left(M - OP_\pi^0, A^*\right) + \ 1 \cdot v\left(M - OP_\pi^0, 0\right) \\
&\quad = v\left(M - OP_\pi^0, 0\right).
\end{aligned} \tag{5.33}$$

Comparison of this expression with equation (5.30) shows that the option price for reducing the probability of an event to zero is equal to the option price for reducing its severity or magnitude to zero. Thus, in this case there is no difference between the value of risk reduction and the value of risk prevention, since in both cases the policy being valued eliminates the risk entirely.

The option price for a reduction in the probability of the event from π to π' is found by solving the following expression for OP'_π:

$$\begin{aligned}&\pi \cdot v\left(M, A^*\right)+(1-\pi) v(M, 0) \\ &\quad=\pi' \cdot v\left(M-OP'_\pi, A^*\right)+\left(1 \cdot \pi'\right) v\left(M-OP'_\pi, 0\right).\end{aligned} \tag{5.34}$$

The marginal value for a change in π can be derived by taking the total differential of equation (5.34), setting it equal to zero, and holding dA^* at zero. The result is a standard result in the analysis of the value of risk prevention (Jones-Lee 1974; Cook and Graham 1977; Machina 1983; Smith and Desvousges 1988):

$$w^\pi=\frac{dM}{d\pi}=\frac{v(M, 0)-v\left(M, A^*\right)}{\left[\pi \cdot v_M^*+(1-\pi) v_M^0\right]}, \tag{5.35}$$

where w^π is the WTP ex ante for a change in the probability of A^*. This expression is positive, indicating a positive willingness to pay for reductions in π ($d\pi < 0$).

A comparison of equations (5.35) and (5.32) is instructive. Whereas equation (5.32) is based on the marginal disutility, as given by the term v_A^*, equation (5.35) is based on the nonmarginal difference in utility between the two states of nature, converted to monetary units by a weighted average of the marginal utilities of income in the two states of nature.

Averting Behavior

Observations of an individual's averting behavior might provide a basis for inferring the values of changes in risk. Suppose that, ex ante, an individual can select a level of private spending R, that will reduce the magnitude of A^* given that the event occurs according to the relationship $A^* = A(R, G)$, where G is the level of government protective spending. An example would be purchasing and wearing a seat belt to reduce the severity of injury given the occurrence of an accident. Assume that this function has the following properties:

$$A(0,0)=A^* \quad \text{and} \quad A_R^*<0 \quad \text{and} \quad A_G^*<0.$$

The individual chooses R, given G so as to maximize expected utility:

$$E[u]=\pi \cdot v[M-R, A(R, G)]+(1-\pi) v[M-R, 0]. \tag{5.36}$$

The first-order conditions include

$$\frac{1}{A_R^*}=\frac{\pi \cdot v_A}{\pi \cdot v_{M^*}+(1-\pi) v_{M^\circ}}, \tag{5.37}$$

where v_{M^*} is the marginal utility of income evaluated at the level of A associated with the given level of G. The term $\frac{1}{A_R^*}$ is the reciprocal of the marginal productivity of expenditure on risk reduction, or equivalently, the marginal private cost of reducing A^*. The right-hand side is the marginal value of reducing A^* that was derived above.

The relevant value for policymakers is the value to the individual of an increase in public spending on risk reduction, dG. To find the option price measure, take the total differential of equation (5.36), set it equal to zero, and substitute the first-order condition for the choice of private protective spending into this expression. After some simplification, we have

$$\frac{dM}{dG} = -\frac{A_G^*}{A_R^*} = \frac{\partial R}{\partial G}. \tag{5.38}$$

This means that the individual's marginal willingness to pay for a small increase in government spending is the ratio of the marginal productivities of private spending and public spending in reducing A^*, or the marginal rate of technical substitution between R and G in reducing A^*. This measure can be calculated if the technical relationship $A(R, G)$ is known. The relationship is observable in principle. The welfare change is also given by the marginal rate of substitution between private and public spending, holding expected utility constant. This is not the same thing as the observed change in private spending. For example, if G increases, the individual will reduce R but will also attain a higher level of expected utility. If R enters the utility function directly, then the welfare measure will include unobservable marginal utility terms, and the measure derived here will be an underestimate (overestimate) if R provides positive (negative) utility.

Similar results can be obtained for the case where individual ex ante spending has the effect of reducing the probability of the adverse event. Now let the production function relating private and public expenditures to the probability of the adverse event be

$$\pi = \pi(R,G), \tag{5.39}$$

where $\pi(0, 0) = \pi^*$ and $\pi_R < 0$, $\pi_G < 0$. The individual chooses R so as to maximize expected utility given by

$$E[u] = \pi(R,G)v(M-R,A^*) + [1-\pi(R,G)]v(M-R,0). \tag{5.40}$$

The first-order conditions include

$$\frac{1}{\pi_R} = -\frac{v(M-R,0)-v(M-R,A^*)}{\pi(R,G)v_M^* + [1-\pi(R,G)]v_M^0}. \tag{5.41}$$

Again, the left-hand side is the reciprocal of the marginal productivity of private expenditure on reducing the probability, or, equivalently, the marginal private cost of reducing π. The right-hand side is the marginal value of reducing π as given by equation (5.35).

The marginal value of an increase in public spending to reduce the probability of the event is found by totally differentiating equation (5.40), setting the result equal to zero, and substituting the first-order condition where appropriate. The result is

$$\frac{dM}{dG} = -\frac{\pi_G}{\pi_R} = \frac{\partial R}{\partial G}. \tag{5.42}$$

This result is similar to the case of public spending to reduce A^*. The individual's marginal willingness to pay for public spending is equal to the ratio of the marginal productivities of private and public spending to reduce π, or to the marginal rate of technical substitution between R and G, holding A^* constant. Again, this result is analogous to those derived in the existing literature on protective spending in the absence of uncertainty.

Unfortunately, these results do not carry over to the case where the averting activity jointly produces reductions in π and A^*. Repeating the steps described above but making both A^* and π functions of R and G leads to the following expression (Shogren and Crocker 1991):

$$\frac{dM}{dG} = -\frac{\pi \cdot v_A^* \cdot A_G + \pi_G \cdot \left(v^* - v^0\right)}{\pi \cdot v_A^* \cdot A_R + \pi_R \cdot \left(v^* - v^0\right)}. \tag{5.43}$$

The unobservable utility terms do not cancel out of this expression, so marginal willingness to pay cannot be inferred from information on the averting technology. The inability to use the averting behavior model in this case is due to the jointness of the implicit production technology. It is not a consequence of introducing risk into the analysis.

Hedonic Prices

If either the probability or the magnitude of a risk (or both) is a characteristic of heterogeneous goods such as housing, hedonic price estimation can be used to obtain the relevant ex ante marginal values for risk changes. Suppose that the magnitude of the adverse event varies across the space used for residential housing. For example, the dose of a toxic chemical from an accidental release would depend on the distance from the source of the release. If people are aware of this spatial variation, then they should be willing to pay more for houses in those areas with lower-magnitude risks. Competition for these more attractive houses would result in a systematic inverse relationship between the price of housing, P_h and A_i^*, where i indexes the spatial location of the house. For simplicity, suppose that the magnitude of the event is the only relevant characteristic of housing. Then the price of a house at location i can be found from the hedonic price function $P_h\left(A_i^*\right)$. Given income, the probability of the event, and the magnitude of the event, the individual chooses a location so as to maximize expected utility:

$$E[u] = \pi \cdot v\left[M - P_h\left(A_i^*\right), A_i^*\right] + (1-\pi)v\left[M - P_h\left(A_i^*\right), 0\right]. \tag{5.44}$$

The first-order condition is

$$-\frac{\partial P_h}{\partial A_i^*} = \frac{\pi \cdot v_{A_i}^*}{\pi \cdot v_M^* + (1-\pi)v_M^0}. \tag{5.45}$$

Because the righthand side of this condition is the ex ante marginal value of a reduction in A_i^* (see equation (5.32) again), this condition says that expected utility maximization calls for setting the marginal value of risk reduction equal to its marginal implicit price, the slope of the hedonic price function. Thus, if individuals and the housing market are in equilibrium, the estimated marginal implicit price of risk reduction for each individual reveals each individual's marginal ex ante valuation for risk reduction. However, since a house is a longlived asset, and P_h is an asset price, equation (5.45) yields a compensating wealth measure of the lifetime welfare change associated with a permanent change in π.

If the relevant housing characteristic that varies across space is the probability of the adverse event, and housing prices reflect differences in π_i, the results are similar. The expression for expected utility is

$$E[u] = \pi_i \cdot v\left(M - P_h(\pi_i), A^*\right) + (1 - \pi_i)\, v\left(M - P_h(\pi_i), 0\right) \tag{5.46}$$

and the first-order condition is

$$\frac{\partial P_h}{\partial \pi_i} = -\frac{v\left[M - P_h(\pi_i), 0\right] - v\left[M - P_h(\pi_i), A^*\right]}{\pi_i \cdot v_M^* + (1 - \pi_i)\, v_M^0}\,. \tag{5.47}$$

Again, the right-hand side is the ex ante marginal value of the probability change in equation (5.35). Thus, the observed implicit price of probability reduction also reveals the individual's marginal ex ante value of risk reduction (Smith 1985). If both the probability and the magnitude of the event vary independently across space, housing prices will be a function of both characteristics. Both equations (5.45) and (5.47) must be satisfied in equilibrium. Hedonic price functions that do not include both characteristics as explanatory variables will be misspecified.

Welfare Change with Non-Expected Utility Preferences[3]

The models described above for revealed preference benefit measurement are based on expected utility as a representation of individuals' preferences under uncertainty. As mentioned above, there is quite a bit of empirical evidence against expected utility maximization as a description of behavior. An important question, therefore, is whether models of value of the sort described here can be modified for use with nonexpected utility preferences.

Revising the models in this way is straightforward and involves no additional complications, at least in certain circumstances. This follows from the key features of revealed preference methods for estimating individuals' values from data on behavior. Generally speaking, the models involve first deriving the expression for welfare change, finding the first-order conditions for optimization, and substituting them into the expression for welfare change. Given the assumptions of the models described here, the substitution allows for the canceling out of any observable

3 This section is adapted, with permission, from Freeman (1991a).

utility terms. It turns out that at least for many nonexpected utility representations of preferences, the same result occurs, so that the derived observable welfare measures are independent of the particular form of preferences. This is a straightforward consequence of the envelope theorem. In what follows, this is demonstrated first for the general case and then for two specific forms of nonexpected utility preferences.

Let I be some general index of preferences where the preferences depend on income, prices (implicitly), the probabilities of different states of nature and the magnitudes of the adverse event in different states. Thus,

$$I = f\left(M, A, \pi\right). \tag{5.48}$$

This function is assumed to be convex and twice differentiable. This expression could be nonlinear in the probabilities or incorporate regret and rejoice terms or other deviations from the standard expected utility function, or both. Expected utility preferences also fit this general formulation.

Consider the averting behavior model where $A^* = A(R, G)$. The first-order condition for the optimum R is

$$\frac{\partial I}{\partial R} = -f_{M^*} + f_{A^*} \cdot A_R^* = 0 \tag{5.49}$$

or

$$f_M = f_{A^*} \cdot A_R^*. \tag{5.50}$$

To find the marginal welfare measure for a policy that reduces A^*, totally differentiate equation (5.48), rearrange terms, and substitute in the first-order condition to obtain

$$dI = f_{M^*} \cdot dM + \left(f_{A^*} \cdot A_R^* - f_M\right) dR + f_{A^*} \cdot A_G^* \cdot dG = 0 \tag{5.51}$$

$$\frac{dM}{dG} = -\frac{f_{A^*} \cdot A_G^*}{f_{M^*}} = -\frac{A_G^*}{A_R^*}. \tag{5.52}$$

Thus, the marginal willingness to pay for publicly supplied risk reduction is equal to the marginal rate of technical substitution between public and private risk reduction.

Suppose that in the hedonic model it is probabilities that vary across space. Then the general index of preferences would be

$$I = f\left[M - P_h\left(\pi_i\right), A, \pi_i\right]. \tag{5.53}$$

The first-order condition for the selection of the risk characteristic of housing is

$$\frac{\partial I}{\partial \pi_i} = -f_M \cdot \frac{\partial P_H}{\partial \pi_i} + f_{\pi_i} = 0 \tag{5.54}$$

or

$$f_{\pi_i} = f_M \cdot \frac{\partial P_h}{\partial \pi_i}. \tag{5.55}$$

Totally differentiating equation (5.53) to obtain the welfare measure for the change in π and substituting the first-order condition gives

$$dI = f_M \cdot dM + \left(f_{\pi_i} - f_M \cdot \frac{\partial P_h}{\partial \pi_i} \right) d\pi_i + f_{\pi_i} \cdot dG = 0 \tag{5.56}$$

$$\frac{dM}{dG} = -\frac{f_{\pi_i}}{f_M} = \frac{\partial P_h}{\partial \pi_i}. \tag{5.57}$$

The marginal willingness to pay for publicly supplied risk prevention is equal to the observable marginal implicit price of the risk characteristic of housing.

To illustrate this general result, consider the prospect theory model of Kahneman and Tversky (1979). In the two-state model based on the indirect utility function, the index of preferences takes the following form:

$$I = g(\pi) \cdot v\left(M, A^*\right) + g(1-\pi) \cdot v(M, 0), \tag{5.58}$$

where

$$g(0) = 0, \;\; g(1) = 1 \tag{5.59}$$

and

$$g(\pi) + g(1-\pi) < 1 \text{ for } 0 < \pi < 1. \tag{5.60}$$

Assume that the magnitude of the adverse event depends on both the level of expenditure on a private averting activity R and public expenditure G. The value of reducing A^* is

$$\frac{dM^*}{dA^*} = \frac{g(\pi) v_A^*}{g(\pi) v_M^* + g(1-\pi) v_M^0}, \tag{5.61}$$

which is not directly observable because of the utility and probability weighting terms, but it can be inferred. Given the level of G, the individual's optimal level of the private averting activity is given by

$$\frac{\partial I}{\partial R} = -g(\pi) v_M^* + g(\pi) v_A^* A_R^* - g(1-\pi) v_M^0 = 0. \tag{5.62}$$

Thus,

$$\frac{1}{A_R^*} = \frac{g(\pi)}{g(\pi) v_{M^*} + g(1-\pi) v_{M^\circ}} \cdot v_{A^*}. \tag{5.63}$$

The marginal value to the individual of a change in G is found by totally differentiating equation (5.51), setting it equal to zero, and solving for

$$\frac{dM}{dG} = \frac{dR}{dG} - \frac{g(\pi)}{g(\pi) v_M^* + g(1-\pi) \cdot v_M^0} \cdot v_A^* \cdot A_R^* \frac{dR}{dG} - \frac{g(\pi)}{g(\pi) \cdot v_M^* + g(1-\pi) \cdot v_M^0} \cdot v_A^* \cdot A_G. \tag{5.64}$$

After substituting in equation (5.63), this becomes

$$\frac{dM}{dG} = -\frac{A_G^*}{A_R^*}. \tag{5.65}$$

Similarly, if it is π that can be reduced by private and public expenditure, the value of a reduction in π is

$$\frac{dM}{d\pi} = \frac{g(1-\pi)\cdot v(M-R,0) - g(\pi)\cdot v\left(M-R,A^*\right)}{g(\pi)\cdot v_M^* + g(1-\pi)\cdot v_M^0}. \tag{5.66}$$

The first-order condition for private averting expenditure is

$$-\frac{1}{\pi_R} = \frac{v(M-R,0) - v\left(M-R,A^*\right)}{g(\pi)\cdot v_M^* + g(1-\pi)\cdot v_M^0}. \tag{5.67}$$

After substitution, the value of the public risk-prevention expenditure is

$$\frac{dM}{dG} = -\frac{\pi_G}{\pi_R}. \tag{5.68}$$

Similar results can be derived for other forms of preferences, for example the regret theory of Loomes and Sugden (1982). See Jindapon and Shaw (2008) for a derivation of a complete WTP locus in a model based on rank-dependent expected utility (Quiggin 1982).

Extending these results to the case of weak complementarity is straightforward. Suppose that the consumption of some market good x_i increases the probability of occurrence of some adverse event for the purchaser. Suppose further that there is some public policy action represented by G that can reduce the risk associated with consuming x_i for all consumers. So

$$\pi = \pi\left(x_i, G\right), \tag{5.69}$$

with

$$\partial\pi/\partial x_i > 0$$

and

$$\partial\pi/\partial G < 0.$$

This policy will increase the general preference index for all consumers of x_i and will cause the demand curve for x_i to shift outward. Formally, the general preference index can be written as

$$I = f\left(M, A, \pi, p_i\right), \tag{5.70}$$

where p_i is the price of the complementary good. The general form of the expenditure function is

$$e = g\left(p_i, A, \pi, I^*\right). \tag{5.71}$$

A sufficient condition for weak complementarity requires that there be a choke price for x_i, and that at the choke price

$$\frac{\partial e}{\partial \pi} = 0 . \tag{5.72}$$

As shown in Chapter 4, when the conditions for weak complementarity are satisfied, the monetary equivalent of the increase in well-being is the area between the compensated demand curves for the good before and after the public policy change. Similarly, if the market good affects the severity of the adverse event, weak complementarity requires that $\partial e / \partial A^*$ or $\partial I / \partial A^*$ be zero when the good is not purchased. Since these results are independent of any particular specification of the preference function (other than the conditions of weak complementarity), they will hold for expected utility and nonexpected utility.

The three broad classes of models that have been developed to measure the benefits of environmental change from revealed preferences under certainty can be easily generalized to apply to valuing changes in risk. This generalization of the models does not require that individual preferences take the expected utility form. The principal requirement is that individuals be maximizing some objective function. Then by the envelope theorem, welfare measures that contain unobservable preference terms can be reduced to functions of observable relationships by substitution of the first-order conditions for preference maximization. Thus, if the conditions for utilizing these models are satisfied, there is no particular need to be concerned with how people make their choices under uncertainty.

Option Price, Option Value, and Expected Damages

There is a long history of using ex post expected damage measures in benefit-cost analysis. Perhaps the earliest example of this approach to measuring the value of risk changes is the U.S. Army Corps of Engineers' method for estimating the benefits of flood control projects. For many years, the Corps calculated the reduction in the costs of replacement, repair, and cleaning up after a flood for each possible flood stage, multiplied the monetary damages by the probability associated with that level of flooding, and summed across all possible flood stages (Eckstein 1961). Of course, we would now add to the repair and cleaning costs some monetary measure of the loss of utility associated with the flooding events.

Expected damage measures have the virtue of being relatively easy to calculate from experience with risky events. For example, for most river basins in the United States there is a large body of data on hydrology, land use, and the distribution of structures that can be used to calculate expected flood damages under alternative proposed flood control projects. But even if these losses (including lost utility) were accurately measured, as shown in earlier sections, they will not in general correctly measure welfare change from the ex ante perspective.

Since at least in some circumstances expected damage measures may be relatively easy to obtain, an important question is whether they can be taken as useful approximations of the other forms of ex ante values when policymakers would prefer to have the latter. Are the differences between the two types of measure likely to be small or large? Is the expected damage measure likely to be larger or smaller than the option price measure? What factors determine the sign of the difference and its magnitude?

The derivation of an expected damage measure for risk reduction is straightforward. The first task is to establish a measure of the value of avoiding an event given that an event will occur with certainty. The willingness to pay to avoid the consequences of the event given its occurrence is a *CS* measure of damage, *D*, and is found as the solution to

$$v\left(M, A^*\right) = v(M - D, 0). \tag{5.73}$$

The expected damage measure of the value of avoiding the event that has probability π of occurring is the mathematical expectation of the willingness to pay given by equation (5.73); that is, $\pi \cdot D$. The marginal value of reducing A^* given that the event has occurred can be found by totally differentiating the left-hand side of (5.73) setting it equal to zero, and solving for

$$\frac{dM}{dA^*} = -\frac{v_A^*}{v_M^*}. \tag{5.74}$$

Let this be denoted by d^A to indicate that it is a marginal value and that it refers to changes in *A*. Equation (5.74) is a standard result showing that, ex post, the marginal value is the marginal utility of A^* converted to a money measure by using the marginal utility of income. Taking the expectation yields the ex post value of a marginal reduction in A^*:

$$E\left[d^A\right] = -\pi \frac{v_{A^*}}{v_{M^*}}. \tag{5.75}$$

The percentage difference between the option price and expected damage measures is defined according to the following expression, using equations (5.32) and (5.75):

$$\%\text{ difference} \equiv 100 \cdot \frac{w^* - E\left[d^A\right]}{w^*} = 100 \cdot \left\{(1-\pi) - \left[(1-\pi)\ \frac{v_M^0}{v_M^*}\right]\right\}. \tag{5.76}$$

This expression shows that if the marginal utility of income is independent of the event, so that $v_M^0 = v_M^*$, the difference in the measures is zero. This would be the case, for example, if the individual were risk-neutral or could purchase actuarially fair insurance.

If the adverse event reduces the marginal utility of income so that v_M^* is less than v_M^0, then the expected damage measure will exceed the option price measure and the difference will be negative. On the other hand, if the adverse event increases the marginal utility of income, the option price measure will be larger than the expected damage measure. In both cases the magnitude of the difference will be large when the difference in the marginal utilities of income is large and when the probability of the adverse event is low. To summarize, the sign of the difference between the option price and expected damage measures depends on the way in which the marginal utility of income varies with A; and the magnitude of the difference depends on how much it varies and on the probability of the event.

Freeman (1989) computed these differences for several explicit forms of von Neuman–Morgenstern cardinal utility functions with different degrees of risk aversion. Examples included

$$v = (M - A^*)^b \text{ with } 0 < b < 1 \tag{5.77}$$

$$v = -e^{-b(M - A^*)} \text{ with } b > 0 \tag{5.78}$$

$$v = -e^{-bM(1 - A^*)} \tag{5.79}$$

$$v = \ln[(1 - A^*)M] \ . \tag{5.80}$$

For equations (5.77) and (5.78), the price flexibility of income for A^* is zero, meaning that the marginal value of a change in A is independent of M. Risks of this sort are essentially equivalent to financial risks. An arbitrary choice of units for measuring A^* can be made such that A^* is equal to the ex post D. For some of these functional forms, the marginal utility of income is higher given the adverse event. One way of interpreting this last characteristic is that income and the absence of A^* are substitutes in consumption such that if A^* does not occur, the marginal utility of the consumption that M allows is diminished. This might be the case if the event increased the marginal utilities of ameliorating activities such as cleaning up and repairing damages after a flood. Option price values for reducing A^* will be greater than expected damages, and more so for larger losses occurring with lower probabilities.

The findings for utility functions with this characteristic can be summarized by the following statements:

1 Since the differences are positive as predicted by equation (5.76), empirical valuation measures based on changes in expected damages will understate the option prices.
2 The difference is an increasing function of CS, since a larger CS means a larger difference between v_M^0 and v_M^*.
3 The differences are for the most part trivially small for small losses (equivalent to up to 1 percent of income). Only for high degrees of risk aversion does the difference approach 10 percent for low probability losses.
4 For relatively large losses (equivalent to, say, 10 percent of income), the differences range between 5 percent and 20 percent, depending on the probability and for moderate degrees of risk aversion. For relatively high degrees of risk aversion, the differences range above 60 percent.
5 For catastrophic loss (equivalent to 50 percent of income), the differences are large for all but the least risk-averse form of utility function and most likely event. For the more risk-averse forms of utility function, this means

that option prices for risk reduction can exceed expected damages by factors of 2.5 to more than 100.

This last conclusion might be a possible explanation for a phenomenon that some observers have noted—that is, that the public seems unwilling to accept some very low-probability, high-consequence risks that experts have judged to be acceptable. If the experts' judgments are based on ex post valuation measures and if people are relatively risk-averse and their preferences take this form, the experts could be vastly underestimating the true potential welfare costs that these risks would impose on people.

For other forms of utility functions, v^0_M is greater than v_{M^*}. Examples include

$$v = (1 - A^*)^c M^b \text{ with } 0 < b, c < 1, \tag{5.81}$$

and

$$v = (1 - A^*) \ln M. \tag{5.82}$$

One way of interpreting this characteristic is that income and the absence of A^* are complements in consumption such that if the event does not occur, the marginal utility of the consumption that M allows is high. But if the event does occur, the capacity for consumption to generate utility is diminished. For example, the event "broken leg" decreases the marginal utility of expenditures for hiking and skiing trips. For these functions, expected damages exceed option prices.

Calculations with these functions show that differences are negative as predicted and are less than 10 percent except for low probabilities of very large losses (50 percent of income). For most functional forms, the magnitude of the difference is larger for the most risk-averse utility functions. But for one functional form the difference is largest when the degree of risk aversion is zero; and for one functional form, even the sign of the difference depends on the degree of risk aversion.

As these results demonstrate, option price and expected damage measures of the value of risk reduction are likely to be different. Expected damage measures are very unreliable proxies for option price measures. The sign of the difference depends on specific features of the functions chosen to represent preferences. The degree of risk aversion is not a reliable predictor of the size or even the sign of the differences. Values for risk reduction based on expected damages are not likely to be useful, and could be seriously misleading as guides for risk management decisions.

Next, we turn to the value of risk prevention. The expected damage measure of the value of reducing the probability of a loss is the reduction in the expected value of the compensating surplus associated with the loss. As explained above, the expected loss, $E[D]$, is $\pi \cdot D$. By differentiation, we have

$$dE[D] = d(\pi \cdot D) = D \cdot d\pi. \tag{5.83}$$

In order to examine the relationship between the option price and expected damage measures of the value of a probability change, the right-hand sides of equations (5.35) and (5.83) must be compared, however, they are not easily comparable. The option price measure starts with the difference between two utility levels and converts it to a money measure by using the weighted average of the marginal utilities of income in the two states. The expected damage measure is already in monetary units; specifically, it is the difference in money expenditure necessary to achieve the same level of utility under two different sets of conditions.

Again, Freeman (1989) assumed specific functional forms and parameters and calculated differences. The results are difficult to summarize except to say that differences can be positive or negative, large or small, and can be large even in the absence of risk aversion.

In estimating the values of reducing the probability of adverse events, expected damage measures are unreliable indicators of the desired option prices. Expected damage measures can be either underestimates or overestimates, and the errors involved in using them as proxies for option prices can be large. However, they cannot be predicted without detailed knowledge of the specific characteristics of individuals' preferences.

Is Option Value a Value?

The concept of option value was introduced by Weisbrod in a much-cited paper nearly 50 years ago (Weisbrod 1964). This concept has had an interesting history since its birth, both in terms of its development as a theoretical construct and because of its relationship to, and role in, policy discussions concerning environmental resources. Weisbrod argued that an individual who was unsure of whether he would visit a site such as a national park would be willing to pay a sum over and above his expected consumer surplus to guarantee that the site would be available should he wish to visit it. Weisbrod called this extra sum the option value of the site. Option value was seen to arise when an individual was uncertain as to whether he would demand a good in some future period and was faced with uncertainty about the availability of that good. If option price (OP) is defined as the maximum sum the individual would be willing to pay to preserve the option to visit the site before his own demand uncertainty is resolved, then the excess of option price over expected consumer surplus can be called option value (OV). It was thought that option value should be measured, if it were possible, and added to expected consumer surplus in order to obtain the full measure of the value of providing an environmental service.

Weisbrod apparently viewed the existence of positive option value as being intuitively obvious. Indeed, there is no formal mathematical or logical proof of the existence of option value in Weisbrod's paper. But as the subsequent literature has shown, there are a number of subtleties and complications to the concept, as well as traps for the unwary investigator. For example, Cicchetti and Freeman thought they had proved that option value was positive for risk-averse individuals (Cicchetti and Freeman 1971), but other modelers demonstrated that option value could be negative in this case (see Schmalensee 1972; Anderson 1981; Bishop

1982). Schmalensee showed that even for a risk-averse individual, option value could be greater than, equal to, or less than zero depending upon the particular circumstances. This is essentially a matter of the relationship between the expected damage (or surplus) measure and the option price measure discussed in a preceding section of this chapter. As Bohm (1975) pointed out in a comment on Schmalensee, what matters is the relationship between the marginal utilities of income in the different states of nature.

The matter seemed to rest there until Hartman and Plummer (1987) and Freeman (1984b) provided specific characterizations of the nature of demand uncertainty and examined the implications of different types of demand uncertainty for the relationship between the marginal utilities of income for different states of nature. For example, Hartman and Plummer showed that if an individual was uncertain about future income and the demand for the good in question was a positive function of income, option value is unambiguously negative for risk-averse individuals. Freeman (1984b) showed that in this case risk lovers would have positive option values, and that one plausible form of state-dependent preferences assured positive option values for risk-averse individuals.

At about the same time, Bishop (1982) suggested that useful insights might be obtained by considering the simpler case in which an individual is certain of demand but faces uncertain supply of the good. He showed that option value is greater than zero for a project that eliminates the uncertainty of supply. Thus, option value appeared to be resurrected as a form of benefit associated with a guaranteed future supply of goods, such as national parks. However, Freeman showed that Bishop's conclusion did not hold for all possible forms of reduction in uncertainty of supply. What Freeman called supply-side option value could also be either positive or negative (Freeman 1985). See also Wilman (1987) for more on this topic.

Providing a rationale for preservation of wilderness and scenic beauties seems to have been the motivation for the original investigation of option value. Milton Friedman (1962) in *Capitalism and Freedom* had argued that since there were no externalities associated with uses of national parks, all of the relevant economic values could be captured by the owners of parks through admission fees. Thus, the allocation of land to national parks should be subjected to a market test. Weisbrod's 1964 paper was clearly, at least in part, a response to this argument; and Krutilla (1967) included option value in his list of reasons why markets might fail to achieve allocative efficiency for unique environmental resources. Option value was presented as an economic value over and above the expected use values; and it could not be captured through admission fees. The preservation of the option to visit the park was a form of public good that might justify the preservation of a natural area even when expected use values were less than opportunity costs.

Schmalensee's (1972) theoretical proof that the sign of option value is ambiguous seems to have reduced interest in option value as a policy-relevant concept, at least temporarily. But, in 1983 the Environmental Protection Agency (EPA) listed option value as one form of intrinsic benefit that could be included in the benefit-cost analyses of proposed regulations required under the terms of Executive Order 12291 (U.S. Environmental Protection Agency 1983). Measurement of option value was also part of the work to be done in several EPA-sponsored research projects during the 1980s and early 1990s. In light of the state of the theory, this continuing effort to measure option value and to use it in policymaking is an interesting comment on the power of the idea.

The idea of option value may have been most useful through stimulating the more rigorous analysis of the theory of welfare measurement under uncertainty. Because of the theoretical contributions of authors such as Schmalensee (1972), Graham (1981), Bishop (1982), and Smith (1987a, 1987b), we can now see that what has been called an option value is really just the algebraic difference between the expected values of two different points on a WTP locus. Specifically, it is the algebraic difference between the expected value of the consumer surplus and the state-independent willingness to pay (option price). Since these two points represent alternative ways of measuring the same welfare change, the difference between their expected values cannot be a separate component of value. Furthermore, option value cannot be measured separately—it can only be calculated if we have enough information on preferences to calculate both option price and expected surplus. Finally, as has been shown in this chapter, neither of these points on the WTP locus has any particular claim as a superior welfare measure. Perhaps in recognition of this, option value is not mentioned in EPA's most recent set of guidelines for economic assessment (U.S. Environmental Protection Agency 2000).

Summary

In this chapter, the various ways in which welfare changes can be defined and measured for changes in risks have been described. Graham's WTP locus, option prices, and dynamic WTP measures were all defined and related to each other. The relationship between option price and option value was discussed, as was the fact that option value does not constitute a unique component of value. These measures were related to the PPI criterion when it is employed in benefit-cost analyses of policies dealing with risk. In addition, dynamic welfare measures were described and related to their static counterparts.

The methods presented here are based on an important maintained assumption, which is that individuals are capable of assigning probabilities to alternative outcomes that are consistent with probability theory and contain all relevant information. However, there is a substantial body of evidence that people find it hard to think in probabilistic terms and often have serious misperceptions about the magnitude of important environmental risks. For references to some of this evidence and discussions of its implications for economic analysis, see Machina (1990) and Arrow (1982). Other important references include Slovic, Fischhoff, and Lichtenstein (1979, 1980, 1982). As noted above, people often make choices that are not consistent with expected utility theory.

The observation that our models of behavior may not be very good descriptions of how people actually behave toward risk raises some interesting and difficult questions concerning the evaluation of public policies dealing with risks. From a theoretical perspective it is appropriate to use the preferences and probability assessment of individuals who are facing the risks to form welfare measures for use in a cost-benefit analysis. However, when an individual's perceptions of the risk is significantly out of line with an objective values of that risk, a case can be made that society would be better off if the more accurate representation were used.

References

Anderson, Robert J. 1981. "A Note on Option Value and the Expected Value of Consumer's Surplus." *Journal of Environmental Economics and Management* 8(2): 187–191.

Arrow, Kenneth J. 1982. "Risk Perception in Psychology and Economics." *Economic Inquiry* 20(1): 1–9.

Arrow, Kenneth J., and Anthony C. Fisher. 1974. "Environmental Preservation, Uncertainty, and Irreversibility." *Quarterly Journal of Economics* 88(1): 312–319.

Balikcioglu, Metin, Paul L. Fackler, and Robert S. Pindyck. 2011. "Solving Optimal Timing Problems in Environmental Economics." *Resource and Energy Economics* 33(3): 761–768.

Bishop, Richard C. 1982. "Option Value: An Exposition and Extension." *Land Economics* 58(1): 1–15.

Bishop, Richard C. 1986. "Resource Evaluation Under Uncertainty: Theoretical Principles for Empirical Research." In *Advances in Applied Microeconomics, vol. 4*, edited by V. Kerry Smith. Greenwich: JAI Press.

Bohm, Peter. 1975. "Option Demand and Consumer's Surplus: Comment." *American Economic Review* 65(4): 733–736.

Cameron, Trudy. 2005. "Individual Option Prices for Climate Change Mitigation." *Journal of Public Economics* 89: 283–301.

Cameron, Trudy, and Jeffrey Englin. 1997. "Welfare Effects of Changes in Environmental Quality Under Individual Uncertainty About Use." *Rand Journal of Economics* 28: S45–S70.

Cicchetti, Charles J., and A. Myrick Freeman III. 1971. "Option Demand and Consumer's Surplus: Further Comment." *Quarterly Journal of Economics* 85(3): 528–539.

Colby, Bonnie E., and Dennis C. Cory. 1989. "Valuing Amenity Resources Under Uncertainty: Does the Existence of Fair Contingent Claims Markets Matter?" *Journal of Environmental Economics and Management* 16(1): 149–155.

Conrad, Jon M. 1980. "Quasi-Option Value and the Expected Value of Information." *Quarterly Journal of Economics* 94(2): 813–820.

Cook, Philip J., and Daniel A. Graham. 1977. "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities." *Quarterly Journal of Economics* 91(1): 143–156.

Cory, Dennis C., and Bonnie Colby Saliba. 1987. "Requiem for Option Value." *Land Economics* 63(1): 1–10.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47(2): 315–372.

Desvousges, William H., V. Kerry Smith, and Ann Fisher (1987). "Option Price Estimates for Water-Quality Improvements – A Contingent Valuation Study for the Monongahela River." *Journal of Environmental Economics and Management* 14(3): 248–267.

Eckstein, Otto. 1961. *Water Resources Development: The Economics of Project Evaluation*. Cambridge: Harvard University Press.

Ehrlich, Isaac, and Gary S. Becker. 1972. "Market Insurance, Self Insurance and Self Protection." *Journal of Political Economy* 80(4): 623–648.

Fisher, Anthony C., and W. Michael Hanemann. 1987. "Quasi-Option Value: Some Misconceptions Dispelled." *Journal of Environmental Economics and Management* 14(2): 183–190.

Freeman, A. Myrick, III. 1984a. "The Quasi-Option Value of Irreversible Development." *Journal of Environmental Economics and Management* 11(3): 292–295.

Freeman, A. Myrick, III. 1984b. "The Sign and Size of Option Value." *Land Economics* 60(1): 1–13.

Freeman, A. Myrick, III. 1985. "Supply Uncertainty, Option Price, and Option Value." *Land Economics* 61(2): 176–181.

Freeman, A. Myrick, III. 1989. "Ex Ante and Ex Post Values for Changes in Risks." *Risk Analysis* 9(3): 309–317.

Freeman, A. Myrick, III. 1991a. "Indirect Methods for Valuing Changes in Environmental Risks with Non-expected Utility Preferences." *Journal of Risk and Uncertainty* 4(2): 153–165.

Freeman, A. Myrick, III. 1991b. "Welfare Measurement and the Benefit-Cost Analysis of Projects Affecting Risks." *Southern Economic Journal* 58(1): 65–76.

Friedman, Milton. 1962. *Capitalism and Freedom*. Chicago: University of Chicago Press.

Graham, Daniel A. 1981. "Cost-Benefit Analysis Under Uncertainty." *American Economic Review* 71(4): 715–725.

Graham, Daniel A. 1984. "Cost-Benefit Analysis Under Uncertainty: Reply." *American Economic Review* 74(5): 1100–1102.

Graham, Daniel A. 1992. "Public Expenditure Under Uncertainty: The Net Benefit Criteria." *American Economic Review* 82(4): 882–846.

Hartman, Richard, and Mark Plummer. 1987. "Option Value Under Income and Price Uncertainty." *Journal of Environmental Economics and Management* 14(3): 212–225.

Hazilla, Michael, and Raymond J. Kopp. 1990. "The Social Cost of Environmental Quality Regulations: A General Equilibrium Analysis." *Journal of Political Economy* 98(4): 853–873.

Jindapon, Paan, and W. Douglass Shaw. 2008. "Option Price Without Expected Utility." *Economics Letters* 100(3): 408–410.

Jones-Lee, Michael W. 1974. "The Value of Changes in the Probability of Death or Injury." *Journal of Political Economy* 99(4): 835–849.

Just, Richard E., Darrell L. Hueth, and Andrew Schmitz. 2004. *The Welfare Economics of Public Policy: A Practical Approach to Project and Policy Evaluation*. Northampton: Edward Elgar.

Kahneman, Daniel. 2011. *Thinking Fast and Slow*. New York: Farrar, Straus, and Giroux.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decisions Under Risk." *Econometrica* 47(2): 263–292.

Kolstad, Charles D. 1996. "Learning and Stock Effects in Environmental Regulation: The Case of Greenhouse Gas Emissions." *Journal of Environmental Economics and Management* 31(1): 1–18.

Krutilla, John V. 1967. "Conservation Reconsidered." *American Economic Review* 57(4): 777–786.

LeRoy, Stephen F., and Larry D. Singell, Jr. 1987. "Knight on Risk and Uncertainty." *Journal of Political Economy* 95(2): 394–406.

Loomes, Graham, and Robert Sugden. 1982. "Regret Theory: An Alternative Theory of Rationale Choice Under Uncertainty." *Economic Journal* 92(368): 805–824.

Machina, Mark J. 1983. "Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom." In *New Foundations of Utility and Risk Theory with Applications*, edited by Bernt. P. Stigum and Fred Wenstop. Dordrecht: Reidel.

Machina, Mark J. 1989. "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty." *Journal of Economic Literature* 27(4): 1622–1668.

Machina, Mark J. 1990. "Choice Under Uncertainty: Problems Solved and Unsolved." In *Valuing Health Risks, Costs, and Benefits for Environmental Decision Making*, edited by P. Brett Hammond, and Rob Coppock. Washington D.C.: National Academy Press.

Mäler, Karl-Göran, and Anthony Fisher. 2005. "Environment, Uncertainty, and Option Values." In *Handbook of Environmental Economics, vol. 2,* edited by Karl-Göran Mäler and Jeffrey R. Vincent. Amsterdam: North-Holland.

Meier, Charles E., and Alan Randall. 1991. "Use Value Under Uncertainty." *Land Economics* 67(4): 379–389.

Mendelsohn, Robert, and William J. Strang. 1984. "Cost-Benefit Analysis Under Uncertainty: Comment." *American Economic Review* 74(5): 1096–1099.

Miller, Jon R., and Frank Lad. 1984. "Flexibility, Learning and Irreversibility in Environmental Decisions: A Bayesian Approach." *Journal of Environmental Economics and Management* 11(2): 161–172.

Nguyen, To N., W. Douglas Shaw, Richard T. Woodward, Robert Paterson, and Kevin Boyle. 2007. "An Empirical Study of Option Prices for Hunting Permits." *Ecological Economics* 63(2–3): 476–484.

Pindyck, Robert S. 2000. "Irreversibilities and the Timing of Environmental Policy." *Resource and Energy Economics* 22(3): 233–259.

Pindyck, Robert S. 2002. "Optimal Timing Problems in Environmental Economics." *Journal of Economic Dynamics & Control* 26(9–10): 1677–1697.

Quiggin, J. 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior and Organization* 3: 323–343.

Ready, Richard C. 1988. "Cost-Benefit Analysis Under Temporal Uncertainty." Doctoral dissertation, University of Wisconsin, Madison.

Ready, Richard C. 1993. "The Choice of a Welfare Measure Under Uncertainty." *American Journal of Agricultural Economics* 75(54): 896–904.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets." *Quarterly Journal of Economics* 90(4): 629–649.

Schmalensee, Richard. 1972. "Option Demand and Consumer's Surplus: Valuing Price Changes Under Uncertainty." *American Economic Review* 62(5): 813–824.

Shogren, Jason F., and Thomas D. Crocker. 1991. "Risk, Self-Protection, and Ex Ante Economic Value." *Journal of Environmental Economics and Management* 20(1): 1–15.

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein. 1979. "Rating the Risks." *Environment* 21: 14–20, 36–39.

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein. 1980. "Facts and Fears: Understanding Perceived Risk." In *Societal Risk Assessment: How Safe is Safe Enough*? edited by Richard C. Schwing and Walter A. Albers, Jr. New York: Plenum.

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein. 1982. "Response Mode, Framing, and Information-Processing Effects in Risk Assessment." In *Question Framing and Response Consistency*, edited by Robin M. Hogarth. San Francisco: Jossey-Bass.

Smith, V. Kerry. 1985. "Supply Uncertainty, Option Price, and Indirect Benefit Estimation." *Land Economics* 61(3): 303–307.

Smith, V. Kerry. 1987a. "Non-Use Values in Benefit-Cost Analysis." *Southern Economic Journal* 54(1): 19–26.

Smith, V. Kerry. 1987b. "Uncertainty, Benefit-Cost Analysis, and the Treatment of Option Value." *Journal of Environmental Economics and Management* 14(3): 283–292.

Smith, V. Kerry. 1990. "Valuing Amenity Resources Under Uncertainty: A Skeptical View of Recent Resolutions." *Journal of Environmental Economics and Management* 19(2): 193–202.

Smith, V. Kerry, and William H. Desvousges. 1988. "The Valuation of Environmental Risks and Hazardous Waste Policy." *Land Economics* 64(3): 211–219.

Thaler, Richard H. 1987. "The Psychology of Choice and the Assumptions of Economics." In *Laboratory Experimentation in Economics: Six Points of View*, edited by Alvin E. Roth. Cambridge: Cambridge University Press.

U.S. Environmental Protection Agency. 1983. *Guidelines for Performing Regulatory Impact Analyses*. Washington, D.C.: U.S. EPA.

U.S. Environmental Protection Agency. 2000. *Guidelines for Preparing Economic Analyses.* Washington, D.C.: U.S. EPA.

Viscusi, W. Kip, and Richard J. Zeckhauser. 2006. "The Perception and Valuation of the Risks of Climate Change: A Rational and Behavioral Blend." *Climatic Change* 77(1–2): 151–177.

Weisbrod, Burton A. 1964. "Collective Consumption Services of Individual Consumption Goods." *Quarterly Journal of Economics* 77(3): 71–77.

Wilman, Elizabeth A. 1987. "Note on Supply-Side Option Value." *Land Economics* 63(3): 284–289.

Zhao, Jinhua. 2001. "Uncertainty, Irreversibility and Water Project Assessment." *Water Resources Update* 121: 51–57. Available at: http://ucowr.org/files/Achieved_Journal_Issues/V121_A8Uncertainty, Irreversibility, and Water Project Assessment.pdf

Zhao, Jinhua. 2003. "Irreversible Abatement Investment under Cost Uncertainties: Tradable Emission Permits and Emissions Charges." *Journal of Public Economics* 87: 2765–2789.

Zhao, Jinhua, and Catherine L. Kling. 2001. "A New Explanation for the WTP/WTA Disparity." *Economics Letters* 73(3): 293–300.

Zhao, Jinhua, and Catherine L. Kling. 2004. "Willingness to Pay, Compensating Variation, and the Cost of Commitment." *Economic Inquiry* 42(3): 503–517.

Zhao, Jinhua, and Catherine L. Kling. 2009. "Welfare Measures When Agents Can Learn: A Unifying Theory." *Economic Journal* 119(540): 1560–1585.

Chapter 6

Aggregation of Values across Time

For purposes of economic planning, management, and policymaking, it may often be useful to view environmental and natural resources as assets that yield flows of services over time. The theoretical framework necessary for defining and measuring the values of these service flows to the individuals who receive them has now been developed. However, this theory of value has been developed without explicit consideration of the temporal dimension of these service flows. Specifically, the theory has been applied to the values of services only at a single point in time or over some relatively short period of time (a day, week, or year) during which it is reasonable to treat other things such as income, prices, and the consumption of other goods and services as fixed. The theory also has defined values only at that moment in time at which the service flow is received; and thus, we are left with the question of how to aggregate values that are realized at different moments in time and that might vary from one moment to another.

This chapter deals explicitly with the temporal dimension of value and welfare theory. In the first section, the standard theory of individual preferences and choice in an intertemporal setting is reviewed. In the absence of taxes and other capital market imperfections, utility maximizing individuals borrow or lend so as to equate their marginal rate of substitution between present and future consumption with the market rate of interest. Thus, the interest rate can be taken as a revealed preference indicator of individuals' intertemporal marginal rates of substitution.

In the second section, an individual's marginal rate of substitution is used to convert a marginal welfare measure for one moment in time into its equivalent at any other point in time. Measures for several different points in time can be made commensurate by converting them all to one period—for example, the present. Then these measures can be added to obtain an intertemporal aggregate welfare measure.

After reviewing the theory of intertemporal choice and using this theory to develop measures of intertemporal welfare change, the third section addresses the question of choosing an interest rate in a world with many different market interest rates and other intertemporal prices. The multitude of market interest rates arises because of inflation, taxation, and risk, among other things. Consequently, there is no single number that can be identified as the "correct" interest rate for intertemporal welfare measurement in all circumstances; but it is possible to suggest a reasonable range that will be appropriate in many situations.

One consequence of the multiplicity of interest rates is that environmental policies that divert resources from other investments with high rates of return may have intertemporal opportunity costs at the margin that are different from individuals' intertemporal marginal rates of substitution. One approach to calculating these opportunity costs and making them commensurate with the measure of welfare gains is developed in the fourth section.

The theory of intertemporal welfare measurement presented here has been developed to examine the welfare implications of changes that affect one individual at different points in time. However, there are environmental issues that have long horizons where policy decisions made now will affect future generations. Global climate change is an area in which this concern has become a major point of discussion. The question of interpersonal welfare comparisons across time and the choice of discount rate when multiple generations are affected is taken up the final section of this chapter.

Individual Preferences and Intertemporal Choice

The theory of preferences used up to this point models only single-period choices and makes single-period utility a function of the single-period consumption of marketed goods and services and environmental quality. With time introduced explicitly, the single-period utility in time period t is given by

$$u_t = u_t\left(X_t, q_t\right), \tag{6.1}$$

where X_t measures the expenditure on consumption in t (X_t is a numeraire good with a unit price), and q_t is the level of the environmental or resource service flow at t. This model has obvious limitations in that it cannot be used to analyze intertemporal choices and such things as saving and lending.

The alternative is a model of lifetime utility. This general representation of preferences would have an individual at any point in time experiencing a level of "lifetime" utility from the goods and services and environmental quality that person anticipates over his or her remaining lifetime:

$$u^* = u^*\left(X_1, \ldots, X_t, \ldots, X_T, q_1, \ldots, q_t, \ldots, q_T\right), \tag{6.2}$$

where T is the number of years of remaining life. For simplicity, this formulation implies that the individual gains no satisfaction from making a bequest to others at the end of his or her life; but it is a straightforward matter to add a bequest motivation if that is an important part of the problem being analyzed (see, for example, Cropper and Sussman 1988). This analysis abstracts from all considerations of uncertainty concerning the future.

The existence of single-period preferences as described by equation (6.1) implies that the lifetime utility function (6.2) is separable. Brekke (1997, 93, 108–113) argued that this is a dubious assumption since the utility of X_t and q_t could plausibly depend on the past or future levels of either variable for a variety of reasons. If Brekke's argument is accepted, this means that the expressions for the marginal values of q_t derived later in this chapter would have to include as arguments all past and future levels of X_t and q_t. This would vastly complicate the empirical task of estimation, so as a pragmatic simplification lifetime utility is written as

$$u^* = u^*\left[u_1\left(X_1, q_1\right), \ldots, u_t\left(X_t, q_t\right), \ldots, u_T\left(X_T, q_T\right)\right]. \tag{6.3}$$

It is common to assume that this expression is additively separable and that the single-period utility function is the same for all periods. Thus, we can write

$$u^* = \sum_{t=1}^{T} D_t \cdot u\left(X_t, q_t\right), \tag{6.4}$$

where $u(\cdot)$ is invariant in t, $\partial u^*/\partial X_t$ is positive and decreasing in X_t, and where D_t is meant to capture the individual's time preference.[1] For example, if an individual intrinsically preferred present consumption over future consumption, D_t would decrease with t and the marginal lifetime utility of X_t would be a decreasing function of t.

A useful way to characterize time preference is to let $D_t = 1/(1 + d)^t$ where d can be interpreted as a subjective rate of time preference analogous to an interest rate. This form of constant d and geometrically decreasing D_t over time is referred to as exponential discounting and has been shown to be necessary for intertemporal consistency of individual choice (Strotz 1956; Page 1977). Given exponential discounting, the relationship between d and time preference is as follows:

- $d > 0$ represents positive time preference with the marginal utility of X_t decreasing with t;
- $d = 0$ represents neutral time preference; and
- $d < 0$ represents negative time preference with the marginal utility of X_t increasing with t.

Now assume that the individual knows with certainty that he or she will receive a stream of income payments, M_t, and a stream of environmental services, q_t, and that the individual can borrow or lend in a perfect capital market at an interest rate of r% per period. Assume that wealth is measured as of the beginning of the first period and that income and expenditures are made at the end of each period. Then, if the individual does not wish to leave a bequest and assuming that all accounts must be settled in year T, the individual's lifetime budget constraint is

$$\sum_{t=1}^{T} X_t \cdot \left(1+r\right)^{-t} = W^* = W_0 + \sum_{t=1}^{T} M_t \cdot \left(1+r\right)^{-t}, \tag{6.5}$$

1 See Trostel and Taylor (2001) for an interesting model of time preference that makes $u(\cdot)$ a function of age.

where W^* is lifetime wealth and W_0 is initial wealth. Knowledge of the initial wealth, the streams of income and environmental services, and the interest rate is sufficient to determine the lifetime pattern of consumption of X that maximizes lifetime utility. In each period, if X_t is greater than (less than) M_t, the individual must borrow (lend) the difference.

The first-order conditions for maximizing equation (6.4) subject to the wealth constraint (6.5) take the form

$$\frac{\partial u^*/\partial X_t}{\partial u^*/\partial X_{t+1}} = 1 + r , \tag{6.6}$$

or, since $\partial u^*/\partial X_t = D_t \cdot \partial u/\partial X_t = (1+d)^{-t}\, \partial u/\partial X_t$,

$$(1+d)\frac{\partial u/\partial X_t}{\partial u/\partial X_{t+1}} = 1 + r \quad \text{for all } t = 1,\ldots,t,\ldots,T-1 . \tag{6.7}$$

The individual's intertemporal marginal rate of substitution must be equal to one plus the interest rate. Thus, the individual's intertemporal marginal rate of substitution can be inferred by observing the interest rate that governs these intertemporal tradeoffs. The first-order conditions imply that if both r and d equal 0, the individual will equate the single-period marginal utility of consumption across all periods, thus choosing a constant stream of consumption over time. If $r = 0$ and $d > 0$, the single-period marginal utility of consumption in t must be less than that in $t + 1$, implying that the individual chooses to consume relatively more in period t. An increase in r, other things being equal, requires adjustments to increase the marginal utility of consumption in earlier periods, thus implying a deferral of consumption to future time periods.

Measures of Welfare Change

Next, the development of measures of intertemporal welfare change are considered. First, consider a marginal increase in q_t. The individual would be willing to accept a reduced quantity of the numeraire in that period to hold utility constant. This compensating surplus measure of welfare change is the marginal rate of substitution between q_t and X_t, or

$$w_{qt}^t = \frac{\partial u/\partial q_t}{\partial u/\partial X_t} . \tag{6.8}$$

From equation (6.6) or (6.7), in intertemporal equilibrium the individual would be indifferent between paying w_{qt}^t in period t and paying

$$w_{qt}^0 = \frac{w_{qt}^t}{(1+r)^t} \tag{6.9}$$

now. If the whole stream of future $\boldsymbol{Q}$ increases, the marginal willingness to pay now is the sum of the willingness to pay for each of the components of the increase, or

$$w_{\boldsymbol{Q}}^{0} = \sum_{t=1}^{T} \frac{w_{qt}^{t}}{\left(1+r\right)^{t}}. \tag{6.10}$$

If the changes in $\boldsymbol{Q}$ are nonmarginal, aggregation over time is not so simple. As in Chapter 3, we can define single-period compensating surplus (*CS*) and equivalent surplus (*ES*) welfare measures for changes in any q_t. These are changes in single-period income based on single-period utility. In an analogous fashion, we can use equation (6.2) or (6.3) combined with equation (6.5) to define compensating and equivalent changes in initial wealth that are based on lifetime utility. What is of interest is the relationship between these two alternative perspectives on the measurement of welfare change.

To keep it simple, consider a policy that would result in a change in q during only one period. Specifically, suppose that the policy results in $q_t'' > q_t'$. What is the welfare gain to the individual of this change? The single-period compensating surplus measure is the solution to

$$u\left(X_t - CS_t, q_t''\right) = u\left(X_t, q_t'\right). \tag{6.11}$$

The present value of the single-period measure of gain is $\left(1+r\right)^{-t} CS_t$.

In order to derive the compensating and equivalent wealth measures of lifetime welfare change, the first step is to use the solution to the lifetime utility maximization problem to obtain the lifetime indirect utility function:

$$v^*\left(W^*, q_1, \ldots, q_t, \ldots, q_T\right). \tag{6.12}$$

Then from the perspective of lifetime utility, the individual is willing to make a payment at $t = 0$ that will equate the lifetime utilities with and without the change in q_t. Call this the compensating wealth, or *CW*. It is the solution to

$$v^*\left(W^* - CW, q_1, \ldots, q_t'', \ldots, q_T\right) = v^*\left(W^*, q_1, \ldots, q_t', \ldots, q_T\right). \tag{6.13}$$

Following the analysis by Blackorby, Donaldson, and Moloney (1984), it can be shown that the present value welfare measures based on equations (6.11) and (6.13) are different except under certain special conditions. Specifically:

$$\left|CW\right| \geq \left|\left(1+r\right)^{-t} CS_t\right|. \tag{6.14}$$

This result can be easily explained. If the individual pays *CS* in period t, then according to equation (6.11) single-period utility is restored to its initial level, and so is lifetime utility. However, if the single-period marginal utility of consumption in t is affected by the change in q_t, then the first-order conditions for an optimum allocation of consumption over time will not be satisfied. The individual will wish to reallocate consumption across time through some pattern of additional borrowing and lending. Such reallocation of consumption will increase utility. This means that the individual could have made a larger single-period payment to keep lifetime utility at its original level. This establishes the inequality of equation (6.14).

The equality will hold only if the marginal utility of consumption is invariant to the combined changes in q_t and its offsetting CS_t. For example, suppose that the increase in q_t causes a decrease in the marginal utility of consumption in that period. The payment of CS_t will increase the marginal utility of consumption in period t. These two forces might exactly offset each other, leaving the marginal utility of consumption and the optimal intertemporal pattern of consumption unchanged. However, if the single-period utility function is separable in q_t, such that the marginal utility of consumption is unchanged by the change in q_t, or if the increase in q_t leads to an increase in the marginal utility of consumption, then the net effect of the change in q_t and the offsetting compensation payment will be to increase the marginal utility of consumption in period t, and to cause a reallocation of consumption across time.

Similar inequalities can be derived for equivalent surplus and equivalent wealth (EW) measures of welfare gain and for both types of measures for welfare losses associated with a decrease in q_t. For example, suppose that q_t decreased. If the individual instead had to pay ES_t as the equivalent measure of welfare loss, and if this resulted in an increase in the marginal utility of consumption in t, then the individual could restore part of the loss of lifetime utility by reallocating spending toward period t, and ES_t would be an overestimate of the lifetime welfare loss. These results can be summarized as shown in Table 6.1. All of these expressions can be easily generalized to the case where there are changes in $\boldsymbol{Q}$ in many periods.

These results have somewhat disturbing implications for applied welfare analysis. The typical practice in evaluating environmental policies that affect welfare over many years is to estimate single-period welfare measures and to compute the present value of these single-period changes using some interest rate. But as shown here, if compensating surplus measures of single-period welfare change are used, this procedure will underestimate the true lifetime welfare benefit of the stream of improvements in $\boldsymbol{Q}$ and overstate the loss of decreases in $\boldsymbol{Q}$. If the present value of the single-period CS_t is greater than zero, then the project unambiguously passes a Kaldor potential compensation test; but projects with negative aggregate CS_t could still be efficient. Conversely, if equivalent surplus single-period measures are used, the conventional practice leads to upward-biased estimates of total welfare gains and downward-biased estimates of lifetime losses. Thus, if the sum of the single-period ES_t is less than zero, the project unambiguously fails a Hicks potential compensation test; but projects with positive aggregate present values of single-period ES_t might decrease aggregate welfare (Blackorby, Donaldson, and Moloney 1984).

Table 6.1 Derived inequalities

For gains (q_t increases)	*For losses (q_t decreases)*
$CW \geq (1+r)^{-t} CS_t$	$\lvert CW \rvert \leq \lvert (1+r)^{-t} CS_t \rvert$
$EW \leq (1+r)^{-t} ES_t$	$\lvert EW \rvert \geq \lvert (1+r)^{-t} ES_t \rvert$

Of course, the important question is, how big is this bias? Keen (1990) has shed some light on this question. He showed that the difference between the single-period and the lifetime welfare measures was a term that he called the intertemporal compensating variation, or the maximum amount the individual would be willing to pay to be able to reallocate expenditure and utility across periods. Keen then derived expressions for this term and showed that in most instances the term would be of only second-order importance. In general, the error will be significant only if high values for the elasticity of intertemporal substitution combine with large values of CS_t over time.

Although most of the methods for measuring welfare change produce only single-period measures, there are cases in which it is possible to obtain direct unbiased estimates of CW or EW. If changes in $\boldsymbol{Q}$ affect the market price of an asset, the change in its price might provide a basis for estimating CW or EW. For example, the hedonic price method is based on differences in the prices of assets such as houses. Since what is observed are differences in the market prices for the housing asset, price differentials between low- and high-quality houses reflect the compensating wealth change associated with the difference in the expected stream of environmental quality associated with the house. It is common practice in applied welfare analysis to convert the compensating wealth measure to an annual equivalent by using some interest rate or discount factor. Provided the interest rate properly represents the individual's opportunities for intertemporal substitution, this annual equivalent can be used as an unbiased indicator of lifetime welfare change.

To summarize, if an individual can reallocate his stream of consumption over time through transactions at some interest rate r, that person will equate his intertemporal marginal rate of substitution with a discount factor equal to $1 + r$. This means that r can be interpreted as an intertemporal price or measure of intertemporal value, at least for marginal changes. For nonmarginal changes in $\boldsymbol{Q}$, the correct measure of intertemporal welfare change is the compensating or equivalent change in initial wealth. But many of the available methods for measuring welfare change yield estimates of single-period values rather than lifetime values; and the present value of these single-period welfare measures is, at best, only an approximation (an upper or lower bound) on the lifetime welfare measure. Given the limited availability of data, the approximation may be all that can be obtained.

An important example of computing intertemporal welfare measures is the construction of estimates of the social cost of carbon for use in regulatory analysis for the U.S. government. An interagency working group was formed in 2009 and issued a report in 2010 (Interagency Working Group 2010; Greenstone, Kopits, and Wolverton 2013) that describes the approach the group used to generate estimates of the economic damages generated by additional carbon emissions from 2010 to 2050. To generate this path of emission damages, the authors drew on three integrated assessment models (IAMs) and other data sources to develop a baseline trajectory of emissions, population, world GDP, and consumption levels.

Then, they postulated additional carbon emissions in a given year (say 2020) and used the models to predict changes in the temperature and consumption levels for each year after the postulated change, computed the marginal damages from those changes, and discounted those changes back to the base year. This was done for each year under a range of socioeconomic assumptions. Their estimates of the social cost of carbon rise over time with a "central" social cost of carbon value beginning at \$21 (in 2007 dollars) per ton of CO_2.

Which Interest Rate is the Right Intertemporal Price?

How can we estimate an individual's rate of time preference? Standard economic theory predicts that in a simple economy with no taxes and with perfect capital markets in which all individuals can borrow or lend at the market rate of interest, an individual will arrange her time pattern of consumption by borrowing and lending so that at each point in time her rate of time preference is just equal to the observed market rate of interest (Lind 1982; Varian 1990). In such a simple economy, this is the interest rate that should be used for the intertemporal aggregation of individuals' single-period welfare measures. There are, however, in fact taxes (certainly), inflation (at least sometimes), and various market imperfections; and we observe a multiplicity of market interest rates. The purpose of this section is to discuss briefly the choice of an interest rate for welfare evaluation when factors such as inflation and taxation are present and to suggest some approaches to choosing an intertemporal price when faced with many rates of interest and tax rates.

Consider first the problem of the effects of taxes on income. Taxes can drive wedges between market prices and individuals' marginal rates of substitution. Suppose that an individual's interest income is taxed at the rate of t%. Then money lent at r% will return only $r \cdot (1 - t)$% net of tax. It is this after-tax rate of return to which the optimizing lender will equate his intertemporal marginal rate of substitution. Similarly, if interest expense is treated as deductible in calculating tax liabilities, then the net cost of borrowing at a market rate of r is $r \cdot (1 - t)$%. Again, it is this after-tax net borrowing cost that will govern the borrowing decisions of individuals. Discounting for welfare evaluation should be done at this after-tax interest rate.

A second problem is inflation. If there is price inflation, then the distinction between real and nominal interest rates becomes important. The individual would still arrange her time pattern of consumption so as to equate her rate of time preference with the expected real rate of interest. If inflation of i% is perfectly anticipated, then the observed market interest rate m and the real rate of interest r are related as follows:

$$\frac{(1+m)}{(1+i)}=1+r \quad \text{or} \quad m=r+i+r\cdot i\,. \tag{6.15}$$

Because for reasonable rates of inflation and real interest rates the third term in equation (6.15) will be small and can be ignored, the real rate is (approximately) the market interest rate less the rate of inflation. If the estimates of individuals' single-period CS_t are based on the expectation of actual prices in that period, then discounting should be done at the market rate of interest, which will also reflect the anticipated rate of inflation, *i*. However, the conventional practice in benefit measurement is to abstract from inflation by basing future period estimates of CS_t on the present price level. In that case, discounting should be done at the real rate of interest.

What is a reasonable value for this real rate of interest? One range of estimates comes from Lind (1982, 84). After reviewing historical evidence on rates of return to a variety of investments, Lind concluded that rates of time preference could be as low as 0 percent, if historical real rates of return on U.S. Treasury bills were taken as the benchmark. Over this same period of time, the real after-tax return on a broad portfolio of common stocks was about 4.6 percent (Lind 1982, 83). This could be taken as an upper bound on the appropriate discount rate and would be the correct discount rate if the payments or receipts being discounted had the same risk characteristics as the market portfolio of common stocks.

Given the current openness of financial markets and the ease of moving funds between nations, world average interest rates may be more relevant than rates from any one nation's financial market (Lind 1990). Barro and Martin (1990) presented estimates of an index of real expected and real actual short-term interest rates for nine major industrialized nations, including the United States. For more than two-thirds of the period studied, average pretax real rates of interest were below 2 percent, and they never exceeded 6 percent. At least in the United States, real rates have been in the lower half of this range during most of the first decade of this century**.** These results are broadly consistent with Lind's conclusions (Lind 1982). They also demonstrate the variability of real rates of interest over time. In a recent assessment of this issue, Boardman et al. (2011) examined returns on Moody's AAA rated corporate bonds and after adjusting for inflation concluded that 4.5 percent is a reasonable estimate of the marginal real rate of return on investment.

Rather than look at market rates of interest, an alternative approach is to estimate the implicit interest rate that must be guiding individuals' observed choices over alternative temporal patterns of benefits and costs. For example, suppose an individual who has a choice between receiving \$10 today and \$15 in one year chooses the smaller immediate payment. This implies that the individual discounts future consumption at a rate of at least 50 percent. Empirical studies on choices of this sort often reveal unusually high implicit interest rates—more than 20 percent and up to and above 100 percent.

Frederick, Loewenstein, and O'Donoghue (2002, 389) summarized and discussed over forty empirical studies and estimates of the rate of time preference that had been generated using either revealed preference studies (inferring the rate of time preference from actual economic choices) or stated preference experiments (where inference is based on stated choices). They concluded that this comparison "reveals spectacular disagreement among dozens of studies that all purport to be measuring time preference." They suggested that the disparity is likely due to the fact that each study is measuring both time preference and a combination of other factors such as inflation, habit formation, and uncertainty about the future returns. They also noted, however, that the estimates generally suggest discount rates that are relatively large (many studies generate estimates in excess of 10 to 15 percent). The authors concluded that these "large" estimates may also be due to the confounding factors creating disparate rates in the first place. One striking finding in the review by Frederick, Loewenstein, and O'Donoghue is that the studies that focus on short-term intertemporal tradeoffs generally find much higher implied discount rates than those that focus on longer-term tradeoffs (see Chabris, Laibson, and Schuldt 2008 for additional evidence). In fact, many experiments in which people are offered either hypothetical or real choices between payments of different sizes and dates consistently reveal implicit discount rates that are both high for short periods of time and declining as the interval lengthens (Loewenstein and Thaler 1989; Ainslie 1991; Cropper, Aydede, and Portney 1994). This evidence implies nongeometric discounting, also referred to as hyperbolic discounting. In hyperbolic discounting, each element in a stream of payments or receipts is discounted at a rate that decreases as the period increases. For discussions of the implications of nongeometric or time-inconsistent discounting for individual behavior and the evaluation of public policies, see Harvey (1994), Laibson (1997), O'Donoghue and Rabin (1999), and Cropper and Laibson (1999).

Behavioral economists have advanced a number of possible explanations for this apparent use of hyperbolic discounting relating to limited self-control and immediate gratification (Robinson and Hammitt 2011). Robinson and Hammitt noted these findings pose challenges to the use of standard exponential discounting in benefit-cost analysis. They concluded, however, that since most of the evidence of hyperbolic discounting stems from decisions and behaviors that are short term in nature and most programs being evaluated by benefit-cost analysis are intended to be longer term, it is appropriate to use more traditional market rates. Robinson and Hammitt's recommendation seems appropriate at this juncture; however, as behavioral economics advances, it may be appropriate to reconsider these questions.

Even if we have estimates of market rates of interest or implicit discount rates from revealed or stated behavior, another problem arises in the calculation of welfare measures because of the fact that, for a variety of reasons, people face different interest rates. For example, individuals' real after-tax rates of interest will differ, since individuals face different marginal tax rates and may have different portfolios of investments. Differences in effective interest rates can also arise from market imperfections, transactions costs, and differences in the tax treatment of interest paid and interest received. How should we deal with differences in effective interest rates across individuals? In principle, each individual's single-period welfare changes should be discounted at the interest rate that the person uses to make intertemporal allocation decisions. However, in practice, we typically have single-period welfare measures that are aggregated across individuals who may face quite different rates of interest. Assume for the moment that no individual borrows and lends at the same time and that at any point in time each individual's transactions are entered into at a single interest rate that may vary across individuals. Consider an individual who at period t could borrow for one period at r_{bt} or could lend at another rate r_{st}, where both are real after-tax rates of interest. The individual, in effect, faces a kinked budget line. Whichever direction the individual chooses to move in, the interest rate governing that form of transaction will be the one with which that person equates his or her intertemporal marginal rate of substitution. That interest rate is the one that should be used to discount that individual's welfare measure over that interval of time. So if the individual is a net lender between period t and period $t + 1$, and there is a change in q_{t+1}, the after-tax lending rate should be used to discount CS_{t+1} back to t. If the individual neither borrows nor lends, there is no transaction from which to infer a marginal rate of substitution.

Now consider aggregation across individuals who face different effective interest rates at any point in time because of, for example, differences in marginal income tax rates, or differences in borrowing and lending behavior. Specifically, assume that the ith individual faces an effective real interest rate r_i that is constant over time. Assume then that we wish to calculate the sum of the compensating wealth payments that would leave each individual in his initial lifetime utility position. For individual i, the lifetime CW_i for a change in $\boldsymbol{Q}$ is approximated by the present value of the stream of single-period compensating surpluses:

$$CW_i \simeq \sum_{t=1}^{T} \frac{CS_{it}}{\left(1+r_i\right)^t} . \tag{6.16}$$

The aggregate change for the N affected individuals is

$$\overline{CW} \simeq \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{CS_{it}}{\left(1+r_i\right)^t} . \tag{6.17}$$

There is a simple way of calculating a weighted average discount factor that can be used with equation (6.17). First, each individual's share of each period's aggregate compensating surplus is calculated

$$s_{it} = \frac{CS_{it}}{\sum_{j=1}^{N} CS_{jt}} , \tag{6.18}$$

and then this expression is used to substitute for CS_{it} in equation (6.17) to obtain

$$\overline{CW} \simeq \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it} \frac{\sum_{j=1}^{N} CS_{jt}}{\left(1+r_i\right)^t} . \tag{6.19}$$

The preceding analysis has been based on the assumption that each individual enters into transactions at only one interest rate in any one period. However, introspection and anecdotal evidence suggest that this assumption is often violated in practice. Many people simultaneously save by contributing to pension funds and borrow to buy new houses and consumer durables. Yet, they may earn after-tax rates of return on savings instruments that are only a third or a quarter of their effective borrowing costs on outstanding credit card balances. This behavior raises two questions. First, how can this behavior be reconciled with the assumption of economic rationality? Second, what does this behavior imply for inferring intertemporal marginal rates of substitution and making intertemporal welfare measurements?

Lind (1990) suggested that developments in behavioral economics can help to explain these practices. One key paper in this literature (Shefrin and Thaler 1988) hypothesized that people place different types of income, expenditures, and assets into different mental accounts and that they make different intertemporal allocation decisions for different sets of accounts. If it should prove possible to identify unique marginal rates of substitution or implicit discount rates for each mental account, and if it is possible to discover to which mental accounts people assign changes in environmental quality, then it may prove possible to develop a consistent discounting framework for intertemporal welfare measurement.

In summary, the existence of multiple interest rates in the marketplace does not preclude the consistent intertemporal aggregation of individuals' welfare measures. In principle, each individual's set of welfare measures is aggregated across time, using that individual's intertemporal marginal rate of substitution as revealed by the interest rate at which that person transacts. In practice, a weighted average discount factor reproduces this calculation. The case of individuals who simultaneously transact at different interest rates is more problematic. However, the concept of mental accounts may help to explain this behavior and may also provide a basis for calculating the relevant intertemporal marginal rate of substitution.

Capital Costs of Environmental Policies

A major issue in the economic evaluation of public investment projects has been how to take account of the fact that financing the public investment through some combination of taxes and borrowing is likely to displace some private investment having a higher rate of return than the effective interest rate governing individuals' intertemporal substitutions. The divergence between the rate of return on private investment and the social discount rate can be attributed at least in part to the taxes on corporate and personal income. In this section, the "shadow-cost-of-capital" approach to measuring the cost of public investment projects is reviewed and the implications of using the shadow cost of capital for the economic evaluation of public policies that require private investments that may displace other private investments are discussed.

Consider a public investment project with an initial capital cost of K_0. Assume that no further operating or maintenance costs are associated with the project. Let the project yield benefits of B per year in perpetuity. As before, let r represent the effective after-tax rate of interest governing all individuals' borrowing and lending decisions. Since r will reflect individuals' willingness to trade off present for future consumption, it should be used as the intertemporal price for welfare evaluation, given the individualist welfare perspective adopted in this book. Let s represent the real marginal rate of return on investments in the private sector; but because of taxes on capital income at both the corporate and personal level, $s > r$.

Using r, we can calculate the present value of the perpetual stream of benefits as a lower bound to the aggregate compensating wealth payment that is our desired lifetime welfare indicator. However, if we find $B/r > K_0$, this does not necessarily signal that the project will improve welfare, because the relevant welfare comparison is not between the present value of the stream of benefits and the capital cost of the project. Rather, the relevant comparison is between the present value of the benefits of the project and the opportunity cost of the project as reflected in the stream of consumption forgone. In other words, the capital cost must be converted to its consumption equivalent for comparison with the benefit stream.

If K_0 were invested at the marginal rate of return in the private sector, it would produce a perpetual stream of $s \cdot K_0$ per year of future consumption. This is what is lost by diverting K_0 of capital resources from private investment to public investment, and this is the opportunity cost of the public investment. The present value of the stream of future consumption forgone is

$$\frac{s \cdot K_0}{r}. \tag{6.20}$$

Thus, in this simple example the shadow price or true social opportunity cost of one dollar of capital diverted from private investment is s/r, which is greater than one. In practice, the shadow price of capital will depend on a variety of factors reflecting the extent to which private investment is displaced, on net. If no private investment is displaced, the shadow price is one. More realistic formulations of the shadow price expression would take into account the facts that public investment will not necessarily displace private investment on a dollar-for-dollar basis, that public and private investments have less than infinite lifetimes, and that some portion of the future returns from displaced private investment could be reinvested for future consumption rather than being consumed immediately. For a review of some of these issues and some estimates of the magnitude of the shadow price, see Lind (1982). Earlier important references include Bradford (1975) and Marglin (1963).

To summarize, the shadow price of capital is the present value (discounted at the individual rate of time preference, r) of the stream of future consumption forgone from one dollar of public investment. The proper procedure for project evaluation is to discount benefits (and operating and maintenance costs, if any) at the social rate of discount. The present value of the benefit stream would then be compared with the social cost of the capital investment. The latter would be calculated by multiplying the dollar cost of the investment by the appropriately calculated shadow price of capital.

Now suppose there is a pollution control regulation that, instead of requiring public investment, requires private firms to invest capital in pollution control. As in the case of public investment, the relevant opportunity cost is the present value of displaced private consumption. The private investment can affect consumption through a variety of channels. A simple model will illustrate the basic principles. For more detailed analyses, see Kolb and Scheraga (1990) and Lesser and Zerbe (1994).

Suppose also that this regulation will yield environmental improvements with an aggregate value of B per year in perpetuity. Once again, the present value of the benefit stream from the regulation is B/r. Assume that it is estimated that firms can comply with the regulation by installing equipment with an aggregate capital cost of K_0, that there are operating, maintenance, and repair costs of R per year, and that the equipment will last forever. Suppose further that firms face infinitely elastic factor supply curves for all factor inputs and have constant returns-to-scale production functions. This assures that all of the costs of complying with the regulation will be passed on to consumers in the form of higher prices.

Because under these assumptions the aggregate supply of capital is infinitely elastic, the regulation does not result in any displacement of investment in the economy. The cost of the regulation comes in the form of higher prices to consumers. Prices must be raised sufficiently to amortize the investment of K_0 and to cover the operating costs of R per year. Ignoring the effects of higher prices on demand, the required price increase must generate additional revenues of $s \cdot K_0 + R$. This is the annual cost of the regulation to consumers. To calculate the present value of this stream of costs, the effective consumers' interest rate, r, should be used. The present value of the social cost to consumers is approximated by $(s/r) \cdot K_0 + R/r$. In effect, the capital cost is first annualized using the pretax marginal rate of return on private investment (s) and then converted back to a present value using the consumption rate of interest (r).

In order to take account of the effects of higher prices on demand and output, it is necessary to solve for that K_0 which is required to comply with the regulation after taking account of the effect of higher prices on output and the need for capital. For example, if some firms are expected to exit the industry, then fewer firms would have to make investments. The aggregate capital requirement would be smaller than if output and the number of firms was assumed unchanged. Also, the conceptually correct measure of cost is the CV of the consumers who must pay the higher prices.

Discounting and Aggregation Across Generations

We have seen that the aggregation of an individual's single-period welfare measures over time can be done using a discount factor that represents the individual's marginal rate of intertemporal substitution or rate of time preference. For rational individuals this rate will be equal to the relevant after-tax real rate of interest. The ethical justification for discounting benefits and costs that accrue within one generation's lifetime lies in the observation that this discount rate is a reflection of the individuals' preferences for their own present consumption relative to future consumption.

However, discounting future benefits or costs over the lifetimes of two or more generations is controversial. Applying any discount rate to a future sum can make it look very small in presentvalue terms when the interval is 100 years or more. For example, at 2 percent, the present value of $1 in 100 years is about 14 cents; if the $1 comes in 200 years, its present value is only 2 cents at 2 percent.

The principal argument for discounting for economic welfare analysis also provides some insight into why one might choose not to make discounting calculations across generations. Recall that what is being discounted is individuals' willingness to pay or required compensation, either for a change (compensating measure) or to avoid a change (equivalent measure). The ethical justification for making these welfare calculations is their use in applying either the Kaldor or Hicks form of compensation test to determine whether it is possible to compensate the losers so that no one is made worse off. When the gainers and losers are part of the same generation, actual compensation is feasible, and in any event, both groups can participate in the decision. If a proposed policy would impose costs on a future generation, it is hard to imagine mechanisms for assuring that the compensation would be paid if it were thought desirable or necessary to obtain the consent of future generations—the future generation has no voice in present decisions. Similar problems arise in the case of a policy that would yield benefits to the future but impose costs on the present. The real problem with discounting across generations is not, then, that it results in very small numbers for the present values of future effects; rather, it is that it is hard to imagine a meaningful role for measures of compensation in the context of intergenerational resource reallocations.

So, what is to be done? One place to start is with the recognition that, just as in the case of interpersonal welfare comparisons within a generation, a social welfare function is required in order to make comparisons across generations. One approach to defining a social welfare function has its origins in work by Frank Ramsey nearly 90 years ago (Ramsey 1928). For a more recent restatement and discussion, see Arrow et al. (1996). In this formulation, the social rate of discount, r_s, emerges as the sum of two components. The first is the social rate of time preference, ρ, which reflects the society's judgment regarding the relative intrinsic deservingness of different generations. The most ethically appealing judgment is that ρ is zero meaning that all generations are treated as equally deserving. The second component reflects the expected relative economic position of the present and future generations. It is the product of the elasticity of the marginal utility of consumption, θ, and the expected rate of growth of future per capita consumption, g.

Combining these terms gives

$$r_s = \rho + \theta \cdot g \,. \tag{6.21}$$

The justification for the second term is that if per capita consumption is higher in the future, an extra dollar of consumption in the future will have a lower marginal utility to that generation and therefore should be discounted accordingly. Following this approach, with $\rho = 0$, a growth rate of consumption of 1 percent and an elasticity of 1.5 would yield a social discount rate of 1.5 percent. If growth in consumption were expected to be negative, equation (6.21) would produce a negative discount rate to give increments of consumption to future generations a greater social value than equal increments to the present generation.

Weitzman (1998, 2001) put forth a separate, but important, point concerning the appropriate discount rate to use. He argued for using very low discount rates when evaluating the present value of benefits and or costs that accrue in what he calls the "distant future" (76 to 300 years) and the "far distant future" (more than 300 years). His argument is based solely on the notion that future discount rates are uncertain, largely because of uncertainty about future technological progress. Weitzman defined the "certainty-equivalent" of an uncertain discount factor as its expected value and demonstrated that over a long period this expected value converges to the discount factor evaluated at the lowest discount rate possible. Thus, he provided a coherent justification for the existence of a hyperbolic discount rate schedule.

Newell and Pizer (2003, 53–54) provided a simple numerical example to illustrate what is involved. Consider an investment that is expected to yield benefits of \$1000 in 200 years. Let it be equally likely that the discount rate *r* will be 1 percent or 7 percent at that time. Thus, the expected value of *r* is 4 percent. If this rate is used to discount the future benefit, its present value is 34 cents ($\$1000e^{-0.04 \times 200}$). However, the expected value of the discounted present values at 1 percent and 7 percent is 200 times larger ($0.5(\$1000e^{-0.01 \times 200}) + 0.05(\$1000e^{-0.07 \times 200}) = \$68$). Weitzman (2001, 270) also demonstrated the empirical magnitude of adopting this approach. The appropriate marginal discount rate varies from about 4 percent for the immediate future (out to 5 years), down to about 1 percent for the distant future (75 to 300 years, and 0 percent after 300 years). Portney and Weyant (1999) provided additional discussion of discounting and intergenerational equity. In a recent Policy Forum published in *Science*, a number of economists called for the adoption of a declining discount rate schedule based on Weitzman's logic (Arrow et al. 2013).

The appropriate discount rate to use when costs and benefits with intergenerational stakes are at issue remains challenging and consensus among economists has not been reached. An example of this disagreement—as well as a warning concerning the importance of the choice of discount rate in intergenerational settings—is the controversy surrounding the speed at which nations should address climate change. Using a discount rate of 1.4 percent, Stern (2009) concluded that significant resources should be immediately invested to reduce global warming, calling for about 1 percent of world GDP to be used for this purpose. In contrast, Nordhaus (2008), using an average rate of 4 percent in his model, concluded that only modest investments are called for. While there are other differences between their models that account for differences in findings, the choice of discount rate is critically important.

In recognition of the issues raised by discounting costs and benefits that affect future generations, the U.S. Office of Management and Budget (2003) recommends that analysts examine the sensitivity of their findings using lower discounts than they recommend for short-term projects. Specifically, they recommend discount rates in the range of 1 to 3 percent. In constructing the social cost of carbon estimates discussed earlier, the authors seem to have followed this strategy as they adopted three discount rates for their computations including 2.5, 3, and 5 percent per year (Greenstone, Kopits, and Wolverton 2013).

Conclusions

Taking all of these considerations into account, what interest rate is the right intertemporal price? The U.S. Office of Management and Budget recommends that agencies compute the present value of benefits and costs using both a 3 percent and a 7 percent annual discount rate for the following reasons:

> The 7% rate is an estimate of the average before-tax rate of return to private capital in the U.S. economy, based on historical data. It is a broad measure that reflects the returns to real estate and small business capital as well as corporate capital. It approximates the opportunity cost of capital, and it is the appropriate discount rate whenever the main effect of a regulation is to displace or alter the use of capital in the private sector.
>
> The 3% discount rate is based on a recognition that the effects of regulation do not always fall exclusively or primarily on the allocation of capital. When regulation primarily and directly affects private consumption, a lower discount rate is appropriate. The alternative most often used is sometimes called the "social rate of time preference." This term simply means the rate at which "society" discounts future consumption flows to their present value. If one assumes the rate that the average saver uses to discount future consumption is a measure of the social rate of time preference, the real rate of return on long-term government debt may provide a fair approximation. Over the last 30 years, this rate has averaged around 3% in real annual terms on a pre-tax basis.
>
> (U.S. Office of Management and Budget 2003, 11)

Debates concerning the appropriate rate of discount to apply in benefit-cost analysis will continue, particularly when long time horizons are involved. Practitioners will want to carefully review the recommendations and best practices provided in the literature and by agency directives such as the U.S. EPA's *Guidelines for Preparing Economic Analyses* (2010).

References

Ainslie, George. 1991. "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves." *American Economic Review* 81(2): 334–340.

Arrow, K.J., W.R. Cline, K.-G. Mäler, M. Munasinghe, R. Squitieri, and J. Stiglitz. 1996. "Intertemporal Equity, Discounting, and Economic Efficiency." In *Intergovernmental Panel on Climate Change. Climate Change 1995: Economic and Social Dimensions of Climate Change*, edited by James P. Bruce, Hoesung Lee, and Erik F. Haites. Cambridge: Cambridge University Press.

Arrow, K.J., M. Cropper, C. Gollier, B. Groom, G. Heal, R. Newell, W. Nordhaus, R. Pindyck, W. Pizer, P. Portney, T. Sterner, R.S.J. Pol, and M. Weitzman. 2013. "Determining Benefits and Costs for Future Generations." *Science* 341(6144): 349–350.

Barro, Robert J., and Xavier Sala i Martin. 1990. "World Real Interest Rates." *NBER Macroeconomics Annual* 5: 15–59.

Blackorby, Charles, David Donaldson, and David Moloney. 1984. "Consumer's Surplus and Welfare Change in a Simple Dynamic Model." *Review of Economic Studies* 51(164): 171–176.

Boardman, Anthony E., David H. Greenberg, Aidan R. Vining, and David L. Weimer. 2011. *Cost-Benefit Analysis: Concepts and Practice, 4th ed.* Upper Saddle River: Pearson/Prentice Hall.

Bradford, David F. 1975. "Constraints on Government Investment Opportunities and the Choice of Discount Rate." *American Economic Review* 65(5): 887–899.

Brekke, Kjell Arne. 1997. *Economic Growth and the Environment: On the Measurement of Income and Welfare*. Cheltenham: Edward Elgar.

Chabris, Christopher F., David I. Laibson, and Jonathon P. Schuldt. 2008. "Intertemporal Choice." In *The New Palgrave Dictionary of Economics*. 2d ed., edited by Steven N. Durlauf and Lawrence E. Blume. Basingstoke: Palgrave Macmillan.

Cropper, Maureen L., and David I. Laibson. 1999. "The Implications of Hyperbolic Discounting for Project Evaluation." In *Discounting and Intergenerational Equity*, edited by Paul R. Portney and John P. Weyant. Washington, D.C.: Resources for the Future.

Cropper, Maureen L., and Frances G. Sussman. 1988. "Families and the Economics of Risks to Life." *American Economic Review* 78(1): 255–260.

Cropper, Maureen L., Sema K. Aydede, and Paul R. Portney. 1994. "Preferences for Life Saving Programs: How the Public Discounts Time and Age." *Journal of Risk and Uncertainty* 8(3): 243–265.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* 40(2): 351–401.

Greenstone, Michael, Elizabeth Kopits, and Ann Wolverton. 2013. "Developing a Social Cost of Carbon for US Regulatory Analysis: A Methodology and Interpretation." *Review of Environmental Economics and Policy* 7(1): 23–46.

Harvey, Charles M. 1994. "The Reasonableness of Non-constant Discounting." *Journal of Public Economics* 53(1): 31–51.

Interagency Working Group on Social Cost of Carbon, United States Government. 2010. *Technical Support Document: Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866*. Available at: www.epa.gov/oms/climate/regulations/scc-tsd.pdf

Keen, Michael. 1990. "Welfare Analysis and Intertemporal Substitution." *Journal of Public Economics* 42(1): 47–66.

Kolb, Jeffrey A., and Joel D. Scheraga. 1990. "Discounting the Benefits and Costs of Environmental Regulations." *Journal of Policy Analysis and Management* 9(3): 381–390.

Laibson, David I. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112(2): 443–477.

Lesser, Jonathan A., and Richard O. Zerbe. 1994. "Discounting Procedures for Environmental (and Other) Projects: A Comment on Kolb and Scheraga." *Journal of Policy Analysis and Management* 13(1): 140–156.

Lind, Robert C. 1982. *Discounting for Time and Risk in Energy Policy.* Washington, D.C.: Resources for the Future.

Lind, Robert C. 1990. "Reassessing the Government's Discount Rate Policy in Light of New Theory and Data in a World Economy with a High Degree of Capital Mobility." *Journal of Environmental Economics and Management* 18(2, part 2): S8–S28.

Loewenstein, George, and Richard H. Thaler. 1989. "Anomalies: Intertemporal Choice." *Journal of Economic Perspectives* 3(4): 181–193.

Marglin, Stephen A. 1963. "The Opportunity Costs of Public Investment." *Quarterly Journal of Economics* 77(2): 274–289.

Newell, Richard G., and William A. Pizer. 2003. "Discounting the Distant Future: How Much Do Uncertain Rates Increase Valuations?" *Journal of Environmental Economics and Management* 46(1): 52–71.

Nordhaus, William N. 2008. *A Question of Balance: Weighing the Options on Global Warming Policies.* New Haven and London: Yale University Press.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89(1): 103–124.

Page, R. Talbot. 1977. *Conservation and Economic Efficiency: An Approach to Materials Policy*. Baltimore: The Johns Hopkins University Press for Resources for the Future.

Portney, Paul R., and John P. Weyant. 1999. *Discounting and Intergenerational Equity*. Washington, D.C.: Resources for the Future.

Ramsey, Frank P. 1928. "A Mathematical Theory of Saving." *Economic Journal* 138(152): 543–559.

Robinson, Lisa A., and James K. Hammitt. 2011. "Behavioral Economics and Regulatory Analysis." *Risk Analysis* 31(9): 1408–1422.

Shefrin, Hersh M., and Richard H. Thaler. 1988. "The Behavioral LifeCycle Hypothesis." *Economic Inquiry* 26(4): 609–643.

Stern, Nicholas H. 2009. *The Global Deal: Climate Change and the Creation of a New Era of Progress and Prosperity*. New York: Public Affairs.

Strotz, Robert H. 1956. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies* 23(3): 165–180.

Trostel, Philip A., and Grant A. Taylor. 2001. "A Theory of Time Preference." *Economic Inquiry* 39(3): 3379–3395.

U.S. Environmental Protection Agency. 2010. *Guidelines for Preparing Economic Analyses*. National Center for Environmental Economics, Office of Policy, EPA. Available at: http://yosemite.epa.gov/ee/epa/eerm.nsf/vwAN/EE-0568-50.pdf/$file/EE-0568-50.pdf.

U.S. Office of Management and Budget. 2003. Circular A-4. Available at: www.whitehouse.gov/sites/default/files/omb/inforeg/regpol/circular-a-4_regulatory-impact-analysis-a-primer.pdf.

Varian, Hal R. 1990. *Intermediate Microeconomics: A Modern Approach.* 2d ed. New York: Norton.

Weitzman, Martin L. 1998. "Why the Far-Distant Future Should Be Discounted at Its Lowest Possible Rate." *Journal of Environmental Economics and Management* 36(3): 201–208.

Weitzman, Martin L. 2001. "Gamma Discounting." *American Economic Review* 91(1): 260–271.

Chapter 7

Valuing Longevity and Health

The support of human life is one of the basic services provided by the environment. Changes in the life support capacity of the environment brought about by, for example, pollution of the air or water can lead to increases in the incidence of disease, impairment of daily activities, and perhaps reduction of life expectancy. Human alteration of the environment can affect health through a number of channels, including:

- organic compounds, which may contaminate aquifers used as sources of drinking water;
- poorly treated sewage or septic tank leachate, which may spread disease-causing bacteria and viruses into drinking water supplies and among shellfish bound for human consumption;
- elevated levels of particulate matter air pollution, which can increase the risk of premature mortality;
- air emissions from manufacturing facilities, auto body repair and painting shops, and the like, which may include carcinogens; and
- climate change induced by greenhouse gas emissions that can cause increased periods of heat and associated heat-related illness and premature mortality.

The purpose of this chapter is to describe and evaluate the currently available methods and techniques for estimating the monetary values of changes in human health that are associated with environmental changes.[1] Two links must be established in estimating these values. The first is the link between the environmental change and the change in health status; and the second is the link between the change in health status and its monetary equivalent, willingness to pay or willingness to accept compensation. There are two alternative strategies for using revealed preference methods to value environmental changes that affect human health. The first strategy is to develop a comprehensive model

1 The authors are indebted to Maureen Cropper for her contributions to this chapter in the first edition of this book, especially the portions dealing with morbidity benefits and the life-cycle models for valuing changes in mortality risks.

of individual behavior and choice in which environmental quality is one of the determining variables. Such models can provide a basis for measuring willingness to pay directly as a function of the environmental change. The second strategy is to deal with the two links separately. Economic values of changes in health status or health risk would be derived first. They would then be combined with independently derived predictions of health changes or risk changes as a function of environmental change. An example of this approach would be to use measures of the value of risk reduction derived from studies of wage rates and occupational risk, combined with epidemiological studies of the relationship between air pollution and mortality rates, to estimate the economic value of an air quality improvement that reduced the risk of premature mortality.

In the first section of this chapter, models and measurement techniques for valuing changes in mortality as measured by the probability of dying are described. In the second section, revealed preference models based on averting and mitigating behavior and household production are reviewed as a means to valuing reduced risk of morbidity and illness. In the third section, some broader issues concerning valuation of changes in health, including measures of health status such as "quality-adjusted life years" (QALYs) and the valuation of changes in the health of children are considered.

Valuing Reduced Mortality Risks

Because some forms of pollution may increase mortality, economists have had to confront the question of the economic value of reducing the risk of premature death. This area of nonmarket valuation has been prone to serious misunderstanding, and consequently, unnecessary controversy, particularly when the values are discussed in the popular press. The root of the problem is that the value of risk reduction is often reported in conjunction with the number of lives that a policy can be expected to save over the long run. This in turn gives the appearance that economists are placing value (prices!) directly on human lives, as if buying and selling human life were acceptable. However, this interpretation of valuing risk reduction is wrong; it is of utmost importance to recognize that what is being valued is the change in the risk of premature death, a fundamentally probabilistic and ex ante concept. Economists are not valuing the life of a specific individual or group of individuals, an ex post concept. When seen from the view of valuing a change in risk, it is straightforward to see that individuals in their day-to-day actions, and governments in their decisions about social policy, do in fact make tradeoffs between changes in the risk of premature death and other goods that have monetary values. These tradeoffs make it possible to infer the implicit prices that people (or their governments) are attaching to changes in the probabilities of their deaths. For example, when a family chooses to save money by driving to Disneyworld, they have implicitly decided to accept a higher risk of accidental death relative to flying. This observed risk–money tradeoff tells us about their willingness to accept a small increase in risk in exchange for money (i.e., other goods and services).

The subject of this section is the economic theory of value as it is applied to the decisions made by individuals and governments concerning the tradeoff between mortality risk and money. To begin, a brief description of the economic basis for assigning monetary values to certain kinds of life-saving activity is described, as are some of the important economic and ethical issues that have been identified in the literature. More detailed discussions of models and methods for measuring willingness to pay for reduced mortality risk are then presented. A review of some of the issues that arise in using willingness-to-pay measures in the evaluation of environmental policy concludes this section.

Willingness to Pay

In keeping with the assumption that individuals' preferences provide a valid basis for making judgments concerning changes in their economic welfare, reductions in the probability of death due to accident or illness should be valued according to what an individual is willing to pay to achieve the reductions or is willing to accept in compensation to forgo the reductions. Schelling (1968) appears to have been the first to propose applying willingness-to-pay concepts in this area. Other early contributors include Mishan (1971) and Jones-Lee (1974, 1976). The use of willingness-to-pay concepts presupposes that individuals treat longevity more or less like any other good. This seems to be a reasonable assumption, at least for small changes in mortality risk. Individuals in a variety of situations act as if their preference functions include life expectancy or the probability of survival as arguments. In their daily lives, they make a variety of choices that involve trading off changes in the risk of death for other economic goods whose values can be measured in monetary terms. As in the trip to Disneyland scenario described above, many examples of people accepting higher risk to save money can be observed in transportation choices. For example, some people travel to work in cars rather than by bus or by walking because of the increased convenience and lower travel time of cars, even though these people increase their risk of dying prematurely. Also, some people accept jobs with known higher risks of accidental death because the jobs pay higher wages (see Chapter 11). In such cases, people must perceive themselves to have been made better off by the alternatives they have chosen; otherwise, they would have chosen some other alternative. When what is being given up (or gained) can be measured in dollars, the individual's willingness to pay (or compensation required) to accept higher risk is revealed by these choices. These choices are the basis of measures of the economic value of reductions in the risk of death.

As the ordinary, everyday nature of these examples should make clear, the economic question being dealt with here is not about how much an individual would be willing to pay to avoid his or her certain death or how much compensation that individual would require to accept that death. Most people would be willing to pay their total wealth to avoid certain death; and there is probably no finite sum of money that could compensate an individual for the sure loss of life. Rather, the economic question is about how much the individual would be willing to pay to achieve a small reduction in the probability of death during a given period or how much compensation that individual would require to accept a small increase in that probability. This is an appropriate question to investigate because most environmental regulatory programs—even those aimed at fairly serious toxins—result in relatively small changes in individuals' mortality risks. Readers interested in critiques of the economic approach to valuation are directed to Broome (1978).

For this kind of situation, the theory of individual choice under uncertainty described in Chapter 5 provides a useful analytical framework. Individuals do not know which of several alternative states of the world will exist at some specified date in the future; but they often must make choices affecting their future utility before the future is revealed. Individuals are assumed to assign probabilities to alternative states of the world and to make their choices so as to maximize their expected utility. One aspect of uncertainty about the future is the date of one's death, or to put it differently, whether one will survive or succumb to some hazard during some time interval. Individuals can affect the probabilities of death during present and future periods by the choices they make. The value of a reduction in risk to an individual is the amount of money that person would be willing to pay to achieve it, other things being equal. As discussed thus far, the willingness-to-pay approach focuses on the individualistic dimensions of human behavior—that is, an individual's willingness to pay to increase his or her own life expectancy. However, there is nothing in the logic of the willingness-to-pay approach to prevent consideration of the effects of kinship and friendship—that is, an individual's altruistic willingness to pay to reduce the probability of death of close relatives and friends. Again, be clear that the question is not the willingness to pay to prevent an imminent or highly probable death of another person (that is, the ransom for the kidnaped child or the search for the lost hiker or boater) but the willingness to pay for a small reduction in the probability of death for the group of which the friend or relative is a part. This would be a form of paternalistic altruism.

A Shorthand Measure: the Value of a Statistical Life

When evaluating policies that reduce the risk of death, economists and policy analysts found it convenient to aggregate the risk reductions experienced by many different individuals into a single measure, referred to as the value of a *statistical life* (VSL) or the value of a *statistical death avoided.* A statistical life can be thought of as the sum of enough risk reductions so that just 1.0 premature death is avoided; and the value of a statistical life is the willingness to pay to achieve that amount of risk reduction. For example, suppose that there were a group of 10,000 people, each of whom has a probability of .0004 of dying during the next year. Suppose a pollution control policy would reduce that probability to .0003, a change of .0001 (1 in 10,000). Furthermore, suppose that each individual in that group expresses a willingness to pay $500 for this policy. Since the policy would affect all of the people equally, it is a form of collective good for the group. The total willingness to pay of the group is $5 million. If the policy is adopted, there will be on average one less death during the year. Thus, the total willingness to pay for the policy resulting in one less death is $5 million. This is the value of statistical life.

The confusion and misunderstanding generated by the terminology "value of a statistical life" (or its even shorter abbreviation "value of life") has led a number of authors to call for changing the nomenclature associated with the concept. Cameron (2010), in her paper entitled "Euthanizing the Value of a Statistic Life" made a compelling case for permanently discarding this term. She noted that nonspecialists all too easily interpret the term to be a monetary value placed ex post on a life (even specialists can occasionally fall into this trap). Further, she argued that routinely aggregating the risk reductions achieved by a policy to the level of one statistical life is often confusing and unnecessary. Cameron suggested that the profession adopt the expression "willingness to swap (WTS) alternative goods and services for a microrisk reduction in the chance of sudden death (or other types or risks to life and health)" (Cameron 2010, 162–163).

The idea of changing terminology has also gained traction within the U.S. EPA. The Agency's preferred alternative is "value of mortality risk" (VMR) which it suggests be reported in "units using standard metric prefixes to indicate the size of the risk change and the associated time scale, for example, \$/μr/person/yr (dollars per micro [10^{-6}] risk per person per year)" (U.S. EPA 2010a, 16). In its review of the EPA white paper suggesting this change, the Science Advisory Board strongly endorsed the value of changing the title, but preferred "value of risk reduction" (U.S. EPA 2010c). While the jury may still be out on the best replacement wording, there is strong consensus that a change is needed. However, since the VSL terminology is so deeply ingrained in the literature, it is not likely that a movement away from it can occur quickly. In this book, the VSL term will be used when referring to past research in order to avoid confusion, but when possible, its use will be avoided.

Human Capital as a Measure of Welfare?

One early concept of the economic value of an individual life is what that individual produces in the way of marketed goods and services in their lifetime and that this productivity is accurately measured by earnings from labor. This argument suggested that earnings before taxes reflect the government's, and therefore society's, interest in each individual's total productivity. With the death of the individual, that output is lost. This approach has a long tradition; in fact, Landefeld and Seskin (1982) trace the idea back almost 300 years. It has been the basis of some widely cited early estimates of the benefits of air pollution control (for example, Lave and Seskin 1970, 1977).

However, the human capital approach is fundamentally at odds with the individualistic perspective of welfare economics and the theory of value. In effect, by asking what the individual is worth to society, the human capital approach ignores the individual's own well-being, preferences, and willingness to pay. Furthermore, it defines the social worth of the individual in a very narrow way; that is, as the individual's market productivity, and thereby ignores the value of that person's health and well-being to loved ones. Although the human capital approach is inappropriate for valuing reductions in the risk of death, it may make economic sense as the starting point for determining compensation for dependents in wrongful death settlements since these are ex post situations.

Although the human capital approach is flawed in principle as an economic measure of welfare change, one could ask if it might be a reasonable approximation to the value of statistical life based on willingness to pay. Unfortunately, both theoretical reasoning and empirical evidence (both presented below) suggest that human capital measures are a poor proxy for the desired willingness-to-pay measure of value for small changes in the risk of death. While it is likely that an individual's income and the consumption it allows are positively related to the utility that person derives from his or her own life, the human capital method does not reflect the probabilistic nature of death and death avoidance and individuals' differing attitudes toward risk. By definition, an individual with no financial wealth could pay no more than the present value of his expected earnings stream to avoid certain death. However, his statistical value of life based on willingness to pay for small probability changes could be several times his discounted earnings stream.

Modeling Individual Choice and Willingness to Pay

In this section, some of the results generated by economic models of individual choice under uncertainty are described. The purposes of analyzing the models here include deriving hypotheses about the determinants of an individual's willingness to pay for a reduction in the risk of mortality, and examining the relationships between these predicted willingness-to-pay measures and other economic variables such as the individual's earnings and purchases of life insurance.

These models are based on the assumption that, as discussed earlier, individuals make choices that affect their risk of death so as to maximize the mathematical expectation of utility. This means that a key assumption of the models is that individuals know the relevant probabilities of dying and know how they are changed by the choices they make. Since individuals must choose alternatives before the uncertainty is resolved, expected utility is an ex ante concept. These models are used to derive the willingness to pay (WTP) for a reduction in the probability of death, defined as the maximum sum of money that can be taken from the individual ex ante without leading to a reduction in that person's expected utility. This is a compensating surplus measure of welfare change. Alternatively, these models could be used to derive the willingness to accept compensation for an increase in the probability of dying, defined as the sum of money that just compensates for the greater risk by increasing consumption sufficiently to equalize the expected utilities of the two alternatives.

Two aspects of the choice problem are of interest in this analysis: uncertainty and time. The uncertainty concerns the unknown timing of the individual's death and the effects of choices of occupation and consumption activity on the probabilities of surviving to any given date. The time aspect concerns the effects of choices made at any given time on the probabilities and utilities associated with future periods. This is of particular relevance for environmental problems, since many of the most important environmental policy issues are characterized by a substantial interval of time between exposure and the perceived effects on health. Most models of intertemporal choice have focused on identifying the optimal consumption stream given an income stream and opportunities for borrowing and lending. The concern here is with choices made in the present that affect the probabilities of survival at future dates. The problem is to identify the marginal willingness to pay now for increases in the probabilities of survival during some future period.

Static Models

To begin, a simple one-period choice model is developed. The marginal willingness to pay for increases in survival probability is identified. Then the model is extended to the intertemporal case to see whether estimates of the value of statistical life based on intra-temporal models can be applied to the intertemporal case.

Assume that an individual derives utility from the consumption of a composite good, z, with a price normalized to one. The initial endowment of z and the probability of surviving to enjoy its consumption, π, are both given to the individual. Let z^0 and π^0 represent this initial endowment. Arbitrarily normalizing the utility function so that the utility of death is zero, expected utility is

$$E[u] = \pi^0 \cdot u\left(z^0\right). \tag{7.1}$$

For an expression for the individual's marginal willingness to pay in units of z (MWTP) for a reduction in π^0 take the total differential of equation (7.1) and set it equal to zero:

$$MWTP = \frac{u\left(z^0\right)}{\pi^0 \cdot \partial u / \partial z^0}. \tag{7.2}$$

As equation (7.2) shows, one of the fundamental results of models of this sort is that if individuals are not free to adjust their survival probabilities in the market, their marginal willingness to pay for enhanced survival will depend on their initial survival situation. Other things being equal, the higher is the risk of death (lower π^0), the higher will be the marginal willingness to pay to reduce that risk. Another result is that diminishing marginal utility of z implies a higher marginal willingness to pay for changes in π with higher initial endowments of z, other things being equal. These propositions can be verified by taking the derivatives of equation (7.2) with respect to π and z.

This model, while providing useful insights, does not suggest a method for estimating WTP other than by stated preference. To use revealed preference methods we must link WTP to voluntary risk-taking behavior—that is, π must be made endogenous. Now suppose that the individual has the opportunity to rearrange his consumption and survival position through exchange, for example by giving up some z in order to improve his chances of surviving to enjoy the remainder. Let p_π represent the price at which consumption can be exchanged for enhanced survival probability. There is assumed to be no opportunity for insurance. Comprehensive models encompassing bequest motivation (utility derived from unconsumed z remaining at death) and insurance behavior can be developed. For examples, see Jones-Lee (1976), Conley, (1976), Thaler and Rosen (1976), and Cropper and Sussman (1988).

The individual chooses z and π so as to maximize expected utility, subject to the budget constraint

$$\text{Max}: E[u] = \pi \cdot u(z) + \lambda\left\{z^0 - z + p_\pi \cdot \left(\pi^0 - \pi\right)\right\}. \tag{7.3}$$

The first-order conditions for a maximum of expected utility can be combined to obtain

$$\frac{u(z)}{\pi \cdot \partial u / \partial z} = p_\pi . \tag{7.4}$$

This expression requires that the individual equate her marginal willingness to pay for enhanced survival (the left-hand side of the expression) with the given price of enhanced survival. This is similar to the expression for marginal willingness to pay for reductions in risk derived in Chapter 5, since here utility and the marginal utility of consumption are zero in the event of death.

The utility-maximizing individual may choose either to forgo consumption in order to enhance her survival probability or to take on increased risk (lower π) in order to enhance consumption opportunities. The actual choices depend on the initial endowments of z^0 and π^0, the price of π, and the individual's preferences. Whatever the final outcome, equation (7.4) allows us to infer the individual's marginal willingness to pay for enhanced survival, since this value will be equated to the observable price of changes in π.

If the risk in question is an environmental risk, it may not be feasible to have a market for reductions in π. However, if there are other risks, what is required is that the individual be able to affect the level of one of these risks and view equal size reductions in any of these risks as equally valuable, or to put it differently, be indifferent as to source of the risk of death. Suppose now that there are three sources of risk of death, one exogenous environmental risk, one job-related risk, and one related to the level of consumption of a private good x_i, an element in the vector of market goods $\boldsymbol{X}$. Assume that in addition to affecting the risk of death, this good conveys utility directly. The corresponding conditional probabilities of death are denoted ρ_e, ρ_j, and $\rho_x(x_i)$. Assuming that these causes of death are independent, the probability of surviving the current period is the product of the probabilities that the individual does not die from each of the three causes (Sussman 1984); that is,

$$\pi = (1-\rho_e)(1-\rho_j)[1-\rho_x(x_i)]. \tag{7.5}$$

In general, good x_i could either increase or decrease the risk of death. Some goods such as skydiving and cigarettes increase risk, while others such as smoke detectors decrease risk. Our interest is in risk-reducing goods, so that $\partial\rho_x/\partial x_i < 0$. Also, suppose that all types of jobs are alike in every respect except for the risk of accidental death, and that riskier jobs have higher wage rates—in other words, the individual receives an annual wage, $M(\rho_j)$, where $\partial M/\partial\rho_j > 0$.

If I is exogenous income, then total income, M^*, is $I + M(\rho_j)$. Expected utility is given by

$$E(u) = (1-\rho_e)(1-\rho_j)[1-\rho_x(x)] \cdot u\left[M^*(\rho_j) - p_x \cdot x_i, X\right]. \tag{7.6}$$

By total differentiation of equation (7.6), willingness to pay for a marginal change in exogenous risk of death, $dI/d\rho_e$, is given by

$$\left(w_{\rho_e} = (1-\rho_j)(1-\rho_x)\frac{u(\cdot)}{\pi\dfrac{\partial u}{\partial M^*}}\right). \tag{7.7}$$

This is the value of the utility lost if the individual dies $u(\cdot)$, converted to dollars by dividing by the expected marginal utility of income ($\pi \cdot \partial u/\partial M^*$), and multiplied by the probability that the individual does not die due to other causes $[(1-\rho_j)(1-\rho_x)]$.

The major question of interest is whether MWTP can be estimated by observing risk-taking behavior in consumption or in the labor market. To answer this, assume that the individual chooses x_i and a job with its associated risk ρ_j, so as to maximize equation (7.6) and derive the first-order conditions for choice of job risk and consumption of x_i. They are

$$\frac{\partial M}{\partial \rho_j} = (1-\rho_e)(1-\rho_x)\frac{u(\cdot)}{\pi\dfrac{\partial u}{\partial M^*}} \tag{7.8}$$

and

$$-\frac{p_x}{\partial\rho_x/\partial x_i} = (1-\rho_e)(1-\rho_j)\frac{u(\cdot)}{\pi\left[\dfrac{\partial u}{\partial M^*} - \dfrac{\partial u/\partial x_i}{p_x}\right]}. \tag{7.9}$$

Equation (7.8) implies that the individual equates the marginal wage income forgone by moving to a safer job with the marginal benefit of a reduction in job risk. The latter is almost identical to the value of an exogenous risk change—equation (7.7)—except that the probability of not dying due to other causes is now $(1-\rho_j)(1-\rho_x)$ instead of $(1-\rho_e)(1-\rho_x)$. If $(1-\rho_e) \approx (1-\rho_j)$, then the marginal price of risk reduction can be used as an estimate of the willingness to pay for a change in exogenous risk. This marginal price, which is an implicit price, can be estimated with a hedonic wage model. That method is described in Chapter 11.

The first-order condition for the choice of x_i differs from the expression for marginal willingness to pay in part because it includes a term for the marginal utility of a dollar spent on x_i. This term is present because of the assumption that x_i conveys utility directly as well as through its effect on risk. Equation (7.9) can be used as an approximation for marginal willingness to pay only if $\partial u/\partial x_i = 0$. Then, if $(1-\rho_e) \approx (1-\rho_x)$, the left-hand side of (7.9), which is the marginal cost of reducing risk through purchasing x_i, can be used to approximate MWTP for an exogenous risk change.

In order to use data on "safety" goods to estimate the value of risk reduction, one must establish that the good's contribution to safety is known. What governs each individual's purchase decision is that person's perception of the risk-reduction capability of the good. In the absence of adequate information, individuals' perceptions might vary substantially and be difficult to observe. This would make the use of safety good purchases for estimating values of risk reduction problematic.

The revealed preference model of consumer choice is based on the assumption that the safety good is divisible; however, this assumption is often invalid. Some of the goods used in actual studies—for example, smoke detectors—are indivisible; that is, their purchase involves a 0–1 decision. The good is purchased if its marginal benefit is equal to or greater than its marginal cost. The equality of willingness to pay and price occurs only for the marginal purchaser of the good. In order to estimate an average value of MWTP, we must have data on the cost of the safety good and on its effect in reducing risk of death for a cross-section of individuals. If marginal cost and marginal risk reduction vary across individuals in the sample, we can estimate the average value of MWTP by using a discrete choice model as described in Chapter 3.

Perhaps the most important conclusion to be drawn from this review of static models of individual choice and willingness to pay is that each person is likely to attach a different value to a small reduction in the probability of dying because of differences in underlying preferences, degree of risk aversion, wealth, the current level of risk exposure, age, the number of dependents, and, perhaps, the quality of life-years expected to be gained from the reduced risk. This conclusion must be kept in mind when interpreting and using the results of empirical estimates of willingness to pay that are based on averages of groups of perhaps quite heterogeneous people.

A second conclusion is that in the case of multiple risks of death where the individual can "purchase" reductions in some component of risk, the observed price or marginal cost of reducing that component of risk can be taken as a close approximation of the individual's willingness to pay for reductions in other components of risk, provided that the safety good or safer job does not also convey utility directly and that the individual values equal reductions in all components of risk equally.

Introducing Time

The static models discussed above are relevant to questions such as transportation and occupational safety in which the individual's actions today affect the probabilities of dying today. However, many of the important environmental and occupational health questions involve actions taken today whose effects on the probability of dying are realized only at some time, perhaps 10 to 20 years, in the future. How much would an individual be willing to pay now to control current pollution when the effects of improved health might be realized only at some future time? The next step in the analysis is to develop a simple multiperiod model that allows us to investigate how willingness to pay now for a reduction in the probability of death is influenced by the time period to which the probability applies. To focus on the intertemporal aspect of the problem, assume only one cause of death and one consumption activity.

As an extension of the model of intertemporal choice developed in Chapter 6, assume that individuals maximize their expected lifetime utility u^*, where u^* is an additively separable function of the consumption stream:

$$E\left[u^*\right]=\sum_{t=1}^{T}\pi_t\cdot D^{t-1}\cdot u\left(z_t\right), \tag{7.10}$$

where $D = 1/(1 + d)$,
$d \equiv$ subjective rate of discount or own time preference, and
$\pi_t \equiv$ probability of surviving from period one through period t. That is, the probability of living for at least t years from now (year one), with

$$\pi_t=\prod_{s=1}^{t}\left(1-\rho_s\right). \tag{7.11}$$

The term ρ_s is the conditional probability of dying during year s; that is, the probability of dying in year s given that the individual has survived to the beginning of year s. Similarly, π_s is the probability of surviving the year given being alive at the beginning of the year. For now, assume that the stream of consumption z_t is given exogenously and cannot be altered by borrowing or lending.

The form of the intertemporal utility function deserves a brief comment. As explained in Chapter 6, the assumption that individuals maximize discounted utilities does not necessarily impose any restrictions on the nature of individuals' preferences regarding present versus future consumption. The subjective rate of discount d could be positive, negative, or zero. A variety of types of preferences and behavior can be encompassed in this model, depending on the value of d. In equilibrium, borrowing and lending behavior depends on market interest rates, present and future income levels, and the rate at which the marginal utility of income is diminishing, as well as time preference.

From equation (7.10) the marginal rate of substitution (MRS) between present consumption and present mortality reduction is

$$MRS_{\pi_1 z_1} = \frac{u(z_1) + \sum_{t=2}^{T} \frac{\pi_t}{\pi_1} D^{t-1} \cdot u(z_t)}{\pi_1 \cdot \partial u / \partial z_1}. \tag{7.12}$$

The second term in the numerator shows the dependence of willingness to pay to reduce present risk on life expectancy and the associated stream of expected future utilities. Higher future consumption (z_t) and longer life expectancy (T and π_t) both increase the willingness to pay now for a higher current survival probability.

Suppose now that the marginal rate of substitution between present mortality reduction and present consumption is known, for example, from a study of the demand for occupational safety. Can this marginal rate of substitution be used to estimate the demand for reductions in future mortality that might be obtained through an environmental or occupational health program? The marginal rate of substitution between a mortality reduction in future period t' and present consumption is

$$MRS_{\pi_{t'} z_1} = \frac{\sum_{t=t'}^{T} \frac{\pi_t}{\pi_{t'}} D^{t-1} \cdot u(z_t)}{\pi_1 \cdot \partial u / \partial z_1}. \tag{7.13}$$

The two marginal rates of substitution might differ for any of three reasons. First, equation (7.12) could be greater than (7.13) because it includes a term for first-period utility, $u(z_1)$. Also, the more distant that period t' is from the present, the smaller the stream of future periods being summed in (7.13). However, equation (7.13) could be greater than (7.12) if $\pi_{t'}$ is sufficiently smaller than π_1. Thus, marginal rates of substitution estimated from one type of probability choice problem—for example, the single-period problem—cannot, in general, be used as predictors of the marginal rates of substitution for other types of probability choice problems with different intertemporal dimensions.

In this model, the individual takes the intertemporal pattern of consumption opportunities as given. There is no borrowing or lending. Cropper and Sussman (1990) have extended this model by incorporating borrowing and lending at a riskless interest rate. This allows individuals to adjust their consumption streams over time. Measures of willingness to pay will be different when this opportunity is available to people. To see this, we turn now to a full life-cycle model of intertemporal choice.

A Life-Cycle Model of Willingness to Pay

Several authors, including Usher (1973), Conley (1976), Cropper and Sussman (1990), and Rosen (1994), have used a life-cycle consumption-saving model with uncertain lifetime to analyze an individual's willingness to pay at age j for a change in the conditional probability of dying at age t, $t \geq j$. In this section, the approach of Cropper and Sussman is followed. This model can be used to examine:

- the relationship between WTP for a change in current probability of death and age;
- the relationship between the present value of expected lifetime earnings (the human capital measure) and WTP;
- the relationship between WTP and the latency of the risk, that is, the interval between the exposure to the risk (a carcinogen, for example) and its manifestation (death due to cancer); and
- the relationship between the willingness to pay in advance to reduce a given risk and the willingness to pay at the time the risk is experienced (a discounting question).

Using this approach shows that: (a) WTP will generally decrease with age; (b) under plausible circumstances, the present value of lifetime earnings will be less than WTP; (c) latency reduces WTP; and (d) the willingness to pay at year one for a reduction in risk in year t' is equal to the willingness to pay in year t' for that probability of change discounted back to year one by a discount factor, which in general will be different from the market interest rate.

In the life-cycle model, an individual of any given age has a probability distribution over the date of his or her death. Let j denote the individual's current age; and let $\rho_{j,t}$ be the probability that the individual dies at the end of the year in which he attains age t; that is, the person lives exactly $t-j$ more years. Since the $\{\rho_{j,t}\}$ constitutes a probability distribution, it must be true that

$$\rho_{j,t} \geq 0, t = j, j+1, \ldots, T, \text{ and that } \sum_{t=j}^{T} \rho_{j,t} = 1, \tag{7.14}$$

where T is the maximum attainable age. The probability that the individual survives to his tth birthday, given that he is alive at age j, is $\pi_{j,t}$, which also is the probability that he dies at $t+1$ or later. Formally,

$$\pi_{j,t} = \sum_{s=t+1}^{T} \rho_{j,s} . \tag{7.15}$$

Let δ_t, be the probability of dying at age t conditional on being alive at the beginning of that year. Thus, the conditional probability of surviving that year is $1-\delta_t$. This term can also be derived from the survival probabilities

$$1-\delta_t = \frac{\pi_{j,t+1}}{\pi_{j,t}}. \tag{7.16}$$

Expected lifetime utility at age j is the sum of the utility of living exactly $t-j$ more years, times the probability of doing so. As in the preceding section, assume that utility is additively separable, and assume that there is no bequest motive. Then we can write expected lifetime utility at age j as

$$V_j = E\left[u^*\right] = \sum_{t=j}^{T} \pi_{j,t} \cdot D^{t-j} \cdot u\left(z_t\right). \tag{7.17}$$

The utility function for each period, $u(z_t)$, is assumed to be increasing in z_t, strictly concave, and bounded from below.

Two points about this formulation of the problem should be emphasized. First, the model is based on the assumption that the utility of living depends only on consumption and not on length of life per se. The concavity of the utility function implies that it is always desirable to spread a given amount of consumption over a longer time span. Thus, lifetime utility is an increasing function of life expectancy. However, this is only because of the effect on consumption, not because of the value of being alive per se. This point is returned to later. Second, this model treats survival probabilities as exogenous to the individual. To keep things simple, no attempt is made to introduce opportunities for the individual to alter risk levels into this model—see Conley (1976) and Viscusi and Moore (1989) for examples of intertemporal models incorporating this additional element of choice.

The individual has to choose a time pattern of consumption, given initial wealth W_j, annual earnings M_t, $t = j, \ldots, T$, and capital market opportunities, so as to maximize expected lifetime utility as given by equation (7.17). Arthur (1981) and Shepard and Zeckhauser (1982, 1984) assume that the individual can save by purchasing actuarially fair annuities and borrowing via life-insured loans. If actuarially fair annuities are available, an individual who invests \$1 at the beginning of his jth year will receive \$$(1 + A_j)$ at the end of the year with probability $1 - \delta_j$ and nothing with probability δ_j. For the annuity to be fair (that is, to have an expected payout of $1 + r$ where r is the riskless rate of interest) there must be an annuity rate of interest a_j that satisfies

$$(1 + a_j)(1 - \delta_j) = 1 + r. \tag{7.18}$$

If the individual borrows, she must cover the possibility that she might die before repaying the loan. Agreeing to pay $1 + a_j$ if she survives is equivalent to paying $1+r$ on survival, plus purchasing a life insurance policy in this amount at actuarially fair rates. Thus, we can call a_j the actuarial rate of interest.

The individual's budget constraint can be expressed as the requirement that the present value of expected consumption equal initial wealth plus the present value of lifetime earnings, where discounting is done at the riskless rate, r:

$$\sum_{t=j}^{T} \pi_{j,t} \cdot (1+r)^{j-t} \cdot z_t = W_j + \sum_{t=j}^{T} \pi_{j,t} \cdot (1+r)^{j-t} \cdot M_t\,. \tag{7.19}$$

Alternatively, making use of equation (7.18) and the fact that

$$\pi_{\mathrm{j,t}} = \prod_{s=j}^{t-1} (1-\delta_s)\,, \tag{7.20}$$

the budget constraint can be expressed in terms of the actuarial rate of interest:

$$\sum_{t=j}^{T} \left[\prod_{s=j}^{t-1} (1+a_s)^{-1} \right] \cdot z_t = W_j + \sum_{t=j}^{T} \left[\prod_{s=j}^{t-1} (1+a_s)^{-1} \right] \cdot M_t\,. \tag{7.21}$$

The pattern of consumption over the life cycle is chosen so as to maximize equation (7.17) subject to (7.19) or (7.21). Formally, it is the solution to the Lagrangian problem:

$$\begin{aligned}\underset{X_t}{\text{Max}}\, V_\lambda = &\sum_{t=j}^{T} \pi_{j,t} \cdot (1+d)^{j-t} \cdot u(z_t) \\ &+\lambda\left[W_j + \sum_{t=j}^{T} \pi_{j,t} \cdot (1+r)^{j-t} \cdot (M_t - z_t)\right].\end{aligned} \tag{7.22}$$

Now consider how a government health and safety regulation that reduces conditional probabilities of death affects lifetime utility. A government regulation can alter the probability that a person dies in any year only if that person is alive at the beginning of the year. Consider a regulation that reduces $\delta_{t'}$, the conditional probability of dying at age t'; that is, the probability that the individual dies between his t'th and $(t' + 1)$st birthdays. Note that when the conditional probability of death is altered at age t', it affects the probabilities of surviving to ages $t' + 1$ and beyond (for example, $\pi_{j,t'+k}$) since, by repeated use of the definition of $\delta_{t'}$,

$$\pi_{j,t'} = (1-\delta_j)(1-\delta_{j+1})\cdots(1-\delta_{t'-1}). \tag{7.23}$$

Formally, let $w_{j,t'}$ be the individual's marginal willingness to pay at age j for a change in $\delta_{t'}$. It is measured by the wealth that must be taken away from that person at age j to keep her expected utility constant, given the reduced risk of death, or

$$w_{j,t'} \equiv \frac{dW_j}{d\delta_{t'}} = -\frac{dV_j / d\delta_{t'}}{dV_j / dW_j}. \tag{7.24}$$

The envelope theorem implies that

$$\frac{dV_j / d\delta_{t'}}{dV_j / dW_j} = \frac{\partial V_\lambda / \partial \delta_{t'}}{\partial V_\lambda / \partial W_j}. \tag{7.25}$$

Thus, using equation (7.17), we can express marginal willingness to pay as

$$w_{j,t'} = (1-\delta_{t'})^{-1} \sum_{t=t'+1}^{T} \pi_{j,t} \cdot \left[\begin{matrix}(1+d)^{j-t} \cdot \dfrac{u(z_t)}{\lambda} + \\ (1+r)^{j-t} \cdot (M_t - z_t)\end{matrix}\right]. \tag{7.26}$$

Willingness to pay at age j for a reduction in the conditional probability of death at age t' equals the gain in expected utility from year t' onward, converted to dollars by dividing by the marginal utility of income in year j, λ, and discounted at the individual's own rate of time preference. Added to this is the effect of the change in $\delta_{t'}$ on the budget constraint. A reduction in $\delta_{t'}$ makes the individual wealthier by increasing the present value of her expected lifetime earnings from age $t' + 1$ onward. However, an increase in survival probabilities also decreases the consumption that the person can afford in each of the years $t' + 1$ through T from a given earnings stream. Thus, her willingness to pay is reduced by the change in the present value of the consumption stream.

It should be noted that $w_{j,t'}$ is the rate at which the individual is willing to trade wealth for a unit change in risk. To compute the dollar value of a small change in risk, equation (7.26) must be multiplied by the magnitude of the risk change. For example, if $w_{j,t'} = \$2 \times 10^6$ but the change in risk is only 10^{-6}, then willingness to pay for the risk change is \$2.

There are a number of points to be made about willingness to pay for reductions in the risk of death on the basis of equation (7.26). First, for a policy that affects the conditional probabilities for dying over a number of years, the total marginal willingness to pay is the sum of the willingness to pay for the changes in each of the $\delta_{t'}$.

Second, a key assumption of models of this type is that expected lifetime utility depends only on expected lifetime consumption, as Linnerooth (1979) has noted. Therefore, what is being calculated is the willingness to pay for the opportunity to continue consumption. As Bergstrom (1982) has pointed out, if the intertemporal objective function is derived from preferences among alternative lotteries, it should include a term that values survival per se. If this term is an increasing function of the $\{\pi_{j,t}\}$, any willingness-to-pay measure derived from equation (7.17) must be regarded as a lower bound to true willingness to pay. This condition was first noted in a static context by Conley (1976) and Cook (1978).

Third, even granting the assumption that utility depends only on consumption, as long as the individual's average utility of consumption exceeds his marginal utility, willingness to pay exceeds human capital, and human capital must be interpreted as a lower bound to willingness to pay. By use of the first-order conditions for utility maximization, the term in brackets in equation (7.26) can be written as

$$(1+r)^{j-t}\left(\frac{u(z_t)}{\partial u / \partial z_t} - z_t + M_t\right). \tag{7.27}$$

This implies that if $u(z_t)/[\partial u/\partial z_t] - z_t > 0$ for all t, then each year's contribution to $w_{j,t'}$ as given by equation (7.27) exceeds that year's contribution to the present value of lifetime earnings, $(1 + r)^{j-t} \cdot M_t$, and willingness to pay must exceed the present discounted value of lifetime earnings. As noted by Conley (1976) and Cook (1978), the condition that $u(z_t)/[\partial u/\partial z_t] - z_t > 0$ implies that the average utility of consumption exceeds its marginal utility, a condition that holds for all increasing, concave utility functions, provided consumption exceeds a subsistence level.

Blomquist (1981), expanding on Linnerooth's analysis, reviewed the results of nine empirical estimates of willingness to pay and showed that the implied value of statistical life was typically much larger than the expected lifetime earnings of the members of the sample population used in each study. His analysis lends strong support to the assertion that values of life based on lifetime earnings are a poor proxy for the willingness to pay for reduced risk of mortality.

Fourth, equation (7.26) implies that as t' (the age at which risk of death changes) increases, $w_{j,t'}$ must decline, at least as long as the individual is above subsistence. This implies that the value of reducing a person's current probability of dying must always be greater than the value of reducing exposure at age j to a carcinogen with a latency period of $t'-j$ years. In the latter case, fewer expected life-years are saved. Furthermore, the longer the latency period, the smaller willingness to pay is, other things being held constant. This means that willingness-to-pay measures based on behavior toward contemporaneous risk (for example, from studies of wages and occupational accident mortality) will not be good proxies for willingness to pay to reduce latent risks.

Cropper and Sussman (1990) have shown that there is an alternative approach to dealing with the latency problem. Suppose we have obtained a measure of $w_{t',t'}$ for a group at age t by examining, for example, the group members' tradeoffs between wages and contemporaneous job risks. Cropper and Sussman have shown that $w_{j,t}{}'(j < t')$ can be calculated from the following expression:

$$w_{j,t'} = \prod_{s=j}^{t'} \left(1+a_s\right)^{-1} \cdot w_{t',t'}. \tag{7.28}$$

That is, the willingness to pay in advance for a reduction in risk in t- is the contemporaneous willingness to pay discounted back by a factor derived from the actuarial interest rates over the interval. Recall that these actuarial interest rates are a combination of the riskless interest rate and conditional probabilities of death. As such, they vary over time for the individual and across individuals, depending on age and other factors.

The fifth point to be made is that it is interesting to see how $w_{j,j}$ varies over the life cycle. Shepard and Zeckhauser (1982, 1984) have used an expression like equation (7.27) to examine this question. If consumption were constant for all t, as would be the case if the riskless rate of interest were equal to the subjective rate of time preference, $w_{j,j}$ would decline monotonically with age. Younger persons would always have a higher WTP to reduce current risk of death than older persons because there would be more years of consumption and utility in the summation. If, however, consumption increases over some portion of the life cycle, $w_{j,j}$ may also increase with age up to some point, and then decline. If, for example, the individual cannot be a net borrower but can lend at the riskless rate of interest, her consumption is likely to be constrained by income at the beginning of her life. This will cause the present value of the utility of consumption and, hence, $w_{j,j}$ to increase up to some point, and then to decline.

Toward Measurement

The most commonly used revealed preference method for estimating the value of reduced risk of death is the hedonic wage-risk tradeoff approach that was introduced above and is described in more detail in Chapter 11. A number of meta-analyses have been undertaken that summarize estimates of the VSL generated by wage-risk tradeoff studies from the literature. In an early study, Viscusi (1993) examined estimates of the VSL for U.S. populations. More recent meta-analyses that provide estimates for the U.S. include Miller (2000), Mrozek and Taylor (2002), Viscusi (2003), Kochi, Hubbell, and Kramer (2006), and the U.S. EPA (2010c). Cropper, Hammitt, and Robinson (2011, 321) summarized the central tendency of the estimates reported by the first four of these studies, indicating that these range from a low of \$2–\$3.3 million up to \$11.1 million (in 2009 USD).

Values for risk reduction might also be revealed by choices regarding purchases of goods that reduce mortality risks. For example, smoke detectors and seat belts are goods whose primary purpose is to produce safety; that is, to reduce the risk of death for those who purchase them. Data on the purchase and utilization of these goods have been used to estimate the values of reducing risk of death. Some results are reviewed in Viscusi (1993). Safety could also be one of the characteristics of a differentiated product like an automobile. Different automobile models have measurable differences in accident rates (probability) and the severity of injury. If these differences are systematically related to the prices of different models of automobiles, then the hedonic price model described in Chapters 4 and 10 can be applied to estimate individuals' willingness to pay for reductions in the risk of accident or death. For an example of this approach based on U.S. data, see Atkinson and Halvorsen (1990).

Stated preference methods have also been increasingly used to estimate the value of risk reduction and VSLs. These methods are attractive in that they permit researchers to investigate the relationships between WTP and such variables as age at risk, income, health status, cause of death, the level of baseline risk, whether the risk is voluntarily encountered or not, and the size of the risk reduction. Stated preference studies also allow the analyst to provide the context to respondents that best fits the policy situation in which the risk change will occur. For example, if the value of risk reduction related to lowering mortality from a regulation that will lower the average exposure to a toxic pollutant is sought, a stated preference survey can target the affected population (e.g., elderly, urban residents, etc.) and can describe the salient aspects of the proposed risk change (e.g., latency, size of the risk reduction, etc.).

Cropper, Hammitt, and Robinson (2011) identified three recently published meta analyses of stated preference VSL estimates (Kochi, Hubble, and Kramer 2006; Dekker et al. 2011; Lindhjem, Navrad, and Braathen 2010). Their estimates of the central tendency of VSL range from \$2.7–\$8.5 million (also in 2009 USD). Note that this range of estimates is below the range reported from the revealed preferences studies summarized above. The fact that the estimates from stated preference methods generally lay below those from revealed preference methods remains one of the puzzles of this literature (Cropper, Hammitt, and Robinson 2011).

Using the Value of Risk Reduction and VSL in Policy Evaluation

The benefit of a policy that reduces the risk of premature death for a specified group of people is the sum of the individual WTPs for the reduction in risk of all of the members of the group. These WTPs could vary across the group for a variety of reasons including differences in their age, income, health status, cause of death, and the level and type of baseline risk. However, historically the practice of environmental and safety policy has been to ignore these differences and to use a VSL based on the WTP of the average individual in the group that provided the data for the WTP estimate. For example, if the source of data were a hedonic wage regression, the sample mean values for all of the independent variables would be used to compute the WTP of the sample mean individual. This in turn would provide the basis for computing the VSL for the sample.

In the United States, this practice has been sanctioned for all federal agencies by the Office of Management and Budget (U.S. Office Management and Budget 2000) and the Environmental Protection Agency (EPA) (U.S. Environmental Protection Agency 2010b). EPA uses a VSL of $7.4 million in 2006 dollars. This figure is based on an analysis of 5 stated preference and 21 wage-risk studies in which a Weibull distribution was fitted to the VSLs and the mean of the Weibull distribution was calculated. As acknowledged in the *Guidelines for Preparing Economic Analyses* (U.S. EPA 2010b, Appendix B) these values are in need of updating (all of the studies from which the data are drawn were performed prior to 1991, indicating that they are based on relatively old data and methods).

In the past decade, the practice of using a single value expressed in terms of VSL for all regulations that alter risk has come under increasing criticism for failing to reflect adequately the variety of factors that have been seen to influence individuals' WTP for risk reduction (Sunstein 2004). EPA is working to update their guidance and considering adjustments related to the demographics of the affected populations (age and health status) and the characteristics of the risk. In the latter category, they note that eight different risk dimensions have been considered and shown to have effects on the valuation of risk reduction including whether the risk is involuntary, man-made, controllable, or continuous, among others (U.S. EPA 2010a, 2010b). The fact that EPA and other federal agencies continue to rely on a single value for VSL in the presence of evidence suggesting that these factors can significantly affect the magnitude of the value is problematic (Kniesner, Viscusi, and Ziliak 2010; National Research Council 2008). Recent studies continue to support the importance of differential values across many subpopulations (examples include work by Evans and Smith 2010, Hammitt and Haninger 2010, and Cameron and DeShazo 2013).

Discounting Statistical Lives?

The practice of discounting statistical lives is common in the economic appraisal and evaluation of environmental and safety public health initiatives. The practice allows policymakers to compare policies that affect probabilities of premature death with different time patterns. However, the practice is also controversial from ethical perspectives. Discounting is justified on the basis of two factors. The first is the observation that people tend to prefer present consumption over future consumption (time preference). The second factor is the productivity of capital investments meaning that $1 of resources invested today rather than consumed will make it possible to consume more than $1 at some future time.

The practice of discounting lives saved might be attractive to policymakers because it gives the appearance of making possible the comparison of policies that affect probabilities of premature death with different time patterns and different costs, but without engaging in the controversial practice of assigning monetary values to statistical lives saved. However, at a fundamental level, it is not possible to separate the thing being discounted (lives) from its economic value, especially in the realm of policymaking and choices among alternative policies with different costs. Thus, if economic values should be discounted, then discounting should also apply to statistical lives saved. Discounting of statistical lives saved is a controversial practice. Here, the objection to the practice offered by moral philosophers will be outlined first. Then, two complementary arguments in favor of discounting from economics will be offered.

Speaking explicitly of discounting lives saved, the philosopher Douglas MacLean (1990) has argued that mere differences in the timing of events can have no moral significance. Similar people should not be treated differently solely on account of differences in the timing of their deaths. For example, the premature death of any 20-year-old person today can be considered as no worse than the premature death of a similar 20-year-old 10 years from now. Similarly, if it is better to save 100 lives today than to save 99 lives today, it must also be better save 100 lives 10 years from now than to save 99 lives today.

This is just a special case of a more general point that Broome (a philosopher and economist) attributes to the 19th century philosopher/economist Henry Sidgwick. Broome said:

> [*F*]*rom a universal point of view* the time at which a man lives cannot affect the value of his happiness. A universal point of view must be impartial about time, and impartiality about time means that no time can count differently from any other. In overall good, judged from a universal point of view, good at one time cannot count differently from good at another. Nor can the good of a person born at one time count differently from the good of a person born at another.
>
> (Broome 1992, 92, italics in original)

Thus, whether considering lives or well-being, the amount realized must be given the same moral weight independent of the time of its realization. This is equivalent to saying that in moral terms the discount rate is zero.

Looking at deaths alone, the ethical perspective has validity. However, it abstracts from two important considerations. The first consideration is individual's preferences concerning the timing of the benefits and costs that they experience and that contribute to their well-being. The second is the opportunity cost of resources committed to life-saving policies. Each consideration will be discussed in turn.

As shown in Chapter 6, if individuals wish to maximize their lifetime utility, they must equate their marginal rate of substitution between present and future consumption with the market rate of interest. This is true whether they have a positive, zero, or negative time preference. Furthermore, as long as market goods and nonmarket goods such as health and risk reduction are substitutes in preferences, the marginal utilities of future nonmarket goods should also be discounted. It follows that if policy choices should be based on (or at least reflect) individuals' preferences regarding the benefits and costs they receive, then the benefits and costs of policies including policies that affect the risk of premature death should be discounted.

For example, consider the question of choosing between policy A that prevents x immediate deaths in a group now (e.g., reduces the risks of fatal accidents) and policy B that reduces the same group's exposure to a carcinogen thereby preventing x deaths in, say, 30 years. The group would undoubtedly prefer policy A because of the larger number of life years saved. This would be reflected in their higher willingness to pay now for policy A than for policy B. One way for policymakers to reflect this preference would be to discount the avoided deaths of policy B.

Turning to the second consideration, discounting concerns choices among policies that involve commitments of resources that have opportunity costs; and these opportunity costs can include forgoing other opportunities to reduce mortality at other times. The practice of discounting is a way of bringing these opportunity costs into the decision process. Consider a policy involving costs now that avoids x deaths per year in perpetuity. The undiscounted sum of lives saved is infinite. Without discounting, this would justify an infinite commitment of resources. Surely, however, there are limits on the cost that the present generation is obliged to incur to save future lives. The discounted present value of this stream of lives saved converges on x/r, and the present generation can (must) then decide what cost it is willing to incur to save a finite sum of discounted lives.

What about policy A that prevents x deaths in a group of 40-year-old people now vs. policy B that prevents y ($> x$) deaths in a group of otherwise similar 40-year-old people in 30 years? The ethical perspective claims that policy B is preferred because of the larger number of deaths avoided. However, if the resources used in this policy were instead invested to earn r% for 1 year and then used in a policy similar to B (call it policy C), then y' ($> y$) deaths would be avoided. Thus, C should be preferred to B, and so forth for policies D, E, etc. involving further postponement of the lifesaving policy. None of the policies would be undertaken since an additional postponement can always increase the undiscounted number of deaths avoided. Discounting avoids this counterintuitive result.

These two lines of argument (time preference and opportunity cost) come together in the following way. In order to identify policies that are Potential Pareto Improvements (PPIs), both the benefits and costs of life-saving policies must be discounted at a rate that reflects the interaction of time preference and opportunity cost—the market interest rate in simple models.

Valuing Reduced Morbidity

Morbidity is a general term that refers to cases of disease or being in less than good health. Morbidity can be classified in a variety of ways, among them duration of the condition (chronic or acute), degree of impairment of activity, or type of symptom. An episode of acute morbidity would last only a matter of days and would have a well-defined beginning and end. Chronic morbidity refers to cases of a longer-term illness of indefinite duration. The degree of impairment could be defined in terms of, for example, "restricted activity days" on which a person is able to undertake some, but not all, normal activities; "bed disability days" on which a person is confined to bed, either at home or in an institution, for all or most of a day; or "work loss days" on which a person is unable to engage in ordinary gainful employment. However, these measures of morbidity reflect responses to ill health rather than the health condition itself. Whether a given clinical manifestation of ill health results in any restriction on activity, bed disability, or work loss depends upon a number of socioeconomic variables, such as employment and labor-force status, non-labor sources of income, the presence of other income-earners in the household, and so forth. Morbidity can also be measured by "symptom days," that is, by the occurrence of specific symptoms such as an asthma attack, a headache, a cough, throat irritation, or diarrhea.

The choice of symptoms for defining and measuring morbidity has implications for the economic valuation of health effects. The economic perspective on health focuses attention on effects that people are aware of and wish to avoid (that is, effects that would reduce their utility). Yet some biomedical clinical research focuses on effects of questionable significance to individuals and measures health effects that are difficult to relate to individual perceptions and behavior. The question of how to define morbidity is tied to a legal and policy issue that vexes the U.S. Environmental Protection Agency: What constitutes an adverse health effect? The Clean Air Act of 1970 calls for setting air quality standards so as to protect individuals from *adverse* health effects. The question of whether or not an effect is adverse arises (for example) when clinical studies reveal that exposure to an air pollutant under controlled conditions leads to detectable changes in organ structure or function without necessarily causing pain, impeding people's activities, or reducing life expectancy. Are these changes adverse? From an economic perspective based on the willingness-to-pay definition of value, the answer depends on whether the changes matter to the individual and whether the individual reveals or expresses a willingness to pay to avoid the effect, or requires compensation to experience it.

Measures of morbidity must also take into account the fact that, unlike mortality, morbidity is not a discrete event but a process involving time. Cases observed during a period of time may fall into one of four categories: (a) onset of morbidity occurs before the period and morbidity terminates by either recovery or death during the period; (b) onset of morbidity occurs before the period and morbidity terminates after the period; (c) onset of morbidity occurs during the period and morbidity terminates during the period; and (d) onset of morbidity occurs during the period and morbidity terminates after the period. The prevalence rate encompasses all four categories. It is defined as the total number of cases in the period as a percentage of the average number of persons in the population during the period. The incidence rate covers only the third and fourth categories. It is defined as the number of new cases during the period as a percentage of the average number of persons in the population. Incidence rate data would be more appropriate for investigating causal relationships between changes in exposure and changes in health status. Because willingness to pay to reduce morbidity is likely to depend on the length as well as the number of cases, the prevalence rate and measures incorporating data on duration would be more appropriate for analyzing the social costs of morbidity.

Broadly speaking, monetized estimates of reduced morbidity take one of two forms, those based on individual preferences (willingness to pay or required compensation), and those based on the resource and opportunity costs associated with illness. The latter form is typically referred to as a cost-of-illness or sometimes a damage cost measure. These are examples of the damage function method described in Chapter 2. They seek to identify the real costs of illness in the form of lost productivity and output and the increase in resources devoted to medical care. Costs per case of illness or per day are multiplied by the number of cases or days sick to determine an aggregate value. However, as shown formally in this section, the damage cost approach to valuation will yield an incorrect measure of welfare change. Thus, while costs of illness avoided may be relevant for some policy decisions, they are not a reasonable substitute for WTP values.

To see this in a simple way, consider the case of an individual who experiences one less day of asthma attacks because of an improvement in air quality. The benefit to that person might include avoiding the lost wages associated with being unable to work one day and the reduction in costs for medicine and treatment; but in addition, the individual avoids the discomfort associated with the attack itself. The first two components are captured by the cost-of-illness approach. Only a comprehensive willingness-to-pay measure would capture the discomfort component as well.

Now consider an individual who experiences no asthma attacks at present levels of air pollution because he spends money operating an air purifier and stays at home indoors on high-pollution days in order to prevent the attacks that would be associated with exposure to the outdoor air. If air quality is improved, this individual benefits from being able to reduce the monetary expenditures and the lost wages and opportunities for leisure activities that are associated with these defensive activities. This individual will benefit from reduced air pollution even though there is no observed reduction in the actual incidence of asthma attacks related to air pollution or in the associated cost of illness.

As these two examples show, air pollution that affects human health can reduce people's well-being through four channels: the medical expenses associated with treating disease induced by air pollution (including the opportunity cost of time spent in obtaining such treatment); the lost wages resulting from the inability to work; the defensive or averting expenditures and activities associated with attempts to prevent disease induced by air pollution (including the opportunity cost of time); and the disutility associated with the symptoms and lost opportunities for leisure activities caused by the illness. Improving environmental quality can yield benefits to individuals by reducing some or all of these adverse effects. Portions of the first three of these effects have readily identifiable monetary counterparts, but effects of the fourth kind may not. A truly comprehensive benefit measure should be capable of capturing all of these relevant effects. Measures based solely on decreases in medical costs or lost wages are not comprehensive because they omit major categories of beneficial effects.

Although individual willingness to pay is the correct starting point for analyzing health-related values, there is one important respect in which society's valuation of changes in health might diverge from that of the affected individual. Society has developed several mechanisms for shifting some of the costs of illness away from the individual who is ill and onto society at large. These mechanisms include medical insurance, which spreads the costs of treatment among all policyholders, and sick leave policies, which shift at least part of the cost of lost work days onto the employer and ultimately onto the consumers of the employer's products. An individual's expressed willingness to pay to avoid illness would not reflect those components of the costs of her illness that are borne by or shifted to others. However, the value to society of avoiding her illness includes these components. Empirical measures of the value of reducing illness must take account of these mechanisms for shifting costs. This will be discussed in more detail below.

In this section, a general model of individual choice is developed that captures the principal ways by which individuals can affect their health status. This model is used to derive measures of the value of decreases in pollution that affect health and of improvements in health per se. Several extensions of the basic model are described. The problems that arise when this approach to valuation is applied to reductions in the incidence of chronic disease are considered.

A Basic Model of Health Production and Choice

Most of the formal models used for deriving the value of reduced morbidity use some variant of the health production function first developed by Grossman

(1972). Cropper (1981) introduced a pollution variable into the health production function. Harrington and Portney (1987) extended the model to examine explicitly the relationships among willingness to pay for a reduction in pollution, reductions in the cost of illness, and changes in defensive expenditures. The model developed here is an expanded version of the Harrington and Portney model. Dickie (2003) presented an excellent exposition of these models.

The health production function relates exogenous variables (including environmental variables such as air pollution) and choice variables (such as preventive medicine and treatment costs) to some measure of health status. It is assumed that individuals know their health production function, choose the output level optimally, and choose inputs so as to minimize the cost of production of any level of health. Of course, these are strong assumptions.

As originally formulated by Grossman (1972), the health production function was dynamic, allowing for "investments" in "health capital" that yielded benefits in the form of reduced illness over several time periods. The simple model presented here will abstract from this intertemporal dimension of the problem. Intertemporal models are briefly discussed in a later section.

Let health in any time period be measured by the number of days sick, represented by s. This is a simplification in that it makes no distinction between one episode of illness of two days' duration and two separate illnesses of one day each, and in that differences in the types of symptoms and the severity of illness are ignored. Some of the implications of richer specifications of the health variable are discussed in a later section.

Assume that one of the determinants of health status is the level of exposure to, or dose of, some environmental contaminant. Dose is represented by the scalar variable d, which depends on the concentration of pollution, c, and the amount of an averting activity, a, undertaken to avoid or reduce exposure to pollution. Examples of averting activities include filtering tap water before drinking and staying indoors on days of high air pollutant levels. If the contaminant is an air pollutant, c could be interpreted as the number of days during which some measure of air pollution exceeds a stated standard, the mean value of the pollutant averaged over the relevant time period, or the highest value recorded for the pollutant during that period. Since a change in pollutant emissions is likely to change all of these measures in a predictable way, the choice of a measure for d should be based on whatever is the best predictor of changes in health status. Similar questions arise for other forms of environmental contamination, such as chemicals in drinking water and pesticide residues in food.

Assume also that there is a set of mitigating activities and treatments that can be undertaken to reduce the health impact of any given exposure to pollution, represented by b. Examples of mitigating activities include taking antihistamines and visiting a doctor to obtain relief from a sinus headache. In this model, assume that the level of mitigating activities can be chosen by the individual so as to maximize utility.

The health production function for an individual can be written as

$$s = s(d, b) \tag{7.29}$$

$$d = d(c, a) \tag{7.30}$$

and by substitution

$$s = s(c, a, b) \tag{7.31}$$

with

$$\partial s/\partial c > 0 \tag{7.32}$$

$$\partial s/\partial a, \partial s/\partial b < 0\,. \tag{7.33}$$

The health production function can be estimated from cross-section data on illness, pollution, and averting and mitigating activities. It would also be necessary to control for other determinants of health status such as physical and socioeconomic characteristics of individuals for example, age, sex, use of tobacco, income, and education.

The individual derives utility from the consumption of a numeraire good, z, normalized with a price of one, and leisure, f. Illness causes disutility; thus,

$$u = u(z, f, s)\,, \tag{7.34}$$

with

$$\partial u/\partial z, \partial u/\partial f > 0 \tag{7.35}$$

$$\partial u/\partial s < 0\;. \tag{7.36}$$

The individual chooses z, f, a, and b so as to maximize utility subject to the budget constraint of

$$I + p_w(T - f - s) = z + p_a a + p_b b\;, \tag{7.37}$$

where
I = nonlabor income,
p_w = the wage rate,
T = total time available,
p_a = the price of averting activities, and
p_b = the price of mitigating activities.

Where there are time costs associated with the averting or mitigating activities, they should be incorporated into the full income budget constraint. For an example, see Gerking and Stanley (1986).

The first-order conditions for a maximum include

$$\partial u/\partial z = \lambda \tag{7.38}$$

$$\partial u/\partial f = \lambda \cdot p_w \tag{7.39}$$

and

$$\lambda \cdot \frac{p_b}{\partial s / \partial b} = \frac{\partial u}{\partial s} - \lambda \cdot p_w = \lambda \cdot \frac{p_a}{\partial s / \partial a}, \tag{7.40}$$

where λ is the Lagrangian multiplier and can be interpreted as the marginal utility of income.

The Marginal Value of Reduced Pollution

This model of choice can be used to derive an observable measure of the individual's marginal willingness to pay to reduce pollution. An individual's willingness to pay for a small reduction in ambient pollution is the largest amount of money that can be taken away from that person without reducing her utility. If pollution enters the utility function directly, for example, because of aesthetic disamenities associated with pollution, then there are additional benefits that are not associated with health. This point is discussed further in the section on complex models.

In the health production model, in which pollution affects utility only through health, willingness to pay is the reduction in the cost of achieving the optimal level of health made possible by the decrease in pollution. For example, if a reduction in ozone levels from 0.16 to 0.11 parts per million (ppm) reduces the number of days of respiratory symptoms from 6 to 4, and if an expenditure of \$20 on averting activities or on medicine has the same effect, then (all else being equal) the individual's maximum willingness to pay is \$20 for the ozone reduction.

Formally, marginal willingness to pay for a reduction in pollution (w_c) is given by the reduction in sick time associated with the reduction in pollution times the marginal cost of reducing sick time. The latter is given by the cost of an additional mitigating input divided by the reduction in sick time that input produces, or, alternatively, by the cost of averting behavior divided by the reduction in sick time that averting behavior produces. To see this, first totally differentiate the indirect utility function, $v(I, p_w, p_a, p_b, c)$, and solve for willingness-to-pay (w_c) defined as dI / dc to obtain:

$$\frac{dI}{dc} = -\frac{\partial v / \partial c}{\partial v / \partial I} = -\frac{\partial v / \partial c}{\lambda}. \tag{7.41}$$

Next note that the effect of c on utility consists of two components: the direct loss of utility due to illness and the opportunity cost of time spent sick valued at the wage rate, or

$$\frac{\partial v}{\partial c} = \frac{\partial u}{\partial s} \cdot \frac{\partial s}{\partial c} - \lambda \cdot p_w \cdot \frac{\partial s}{\partial c} = \left(\frac{\partial u}{\partial s} - \lambda \cdot p_w \right) \frac{\partial s}{\partial c}. \tag{7.42}$$

Then, substitute the first-order condition equation (7.40) for the term in parentheses in equation (7.42):

$$\frac{\partial v}{\partial c} = \left(\lambda \cdot \frac{p_b}{\partial s / \partial b}\right) \frac{\partial s}{\partial c}. \tag{7.43}$$

Finally, substitute equation (7.43) into equation (7.41) to obtain

$$w_c = -p_a \frac{\partial s / \partial c}{\partial s / \partial a} = p_a \frac{\partial a}{\partial c}. \tag{7.44}$$

A similar procedure leads to

$$w_c = -p_b \frac{\partial s / \partial c}{\partial s / \partial b} = p_b \frac{\partial b}{\partial c}. \tag{7.45}$$

The right-hand terms in equation (7.44) and equation (7.45) follow from application of the implicit function rule.

Most of these results can also be derived from the expenditure function, as follows:

$$e \equiv \min \left[z + p_a \cdot a + p_b \cdot b + p_w \cdot s\right] + \mu \left[u^0 - u(z, f, s)\right], \tag{7.46}$$

where $s = s(c, a, b)$. The first-order conditions include

$$\frac{p_a}{\partial s / \partial a} = \mu \cdot \frac{\partial u}{\partial s} - p_w. \tag{7.47}$$

So,

$$w_c = -\frac{\partial e}{\partial c} = \left(\mu \cdot \frac{\partial u}{\partial s} - p_w\right) \frac{\partial s}{\partial c} = p_a \cdot \frac{\partial s / \partial c}{\partial s / \partial a}, \tag{7.48}$$

which is the same as equation (7.44). Equation (7.45) can be found by similar manipulations.

There are several things to note about these expressions for marginal willingness to pay. First, the ratios

$$\frac{\partial \mathrm{s} / \partial \mathrm{c}}{\partial \mathrm{s} / \partial \mathrm{b}} \text{ and } \frac{\partial \mathrm{s} / \partial \mathrm{c}}{\partial \mathrm{s} / \partial \mathrm{a}} \tag{7.49}$$

can be interpreted as marginal rates of technical substitution (MRTS) between pollution and the other variable in producing a constant level of sickness. Marginal willingness to pay can be expressed in terms of any of the MRTSs between pollution and another input in the production of health, since to minimize the cost of producing health, the values of marginal products of all inputs must be equal at the margin. Second, all of the measures are functions of observable variables that can be calculated given knowledge of the health production function. Third, as shown by the right-hand terms, marginal willingness to pay can be calculated from the reductions in expenditures on either mitigating or averting behavior that are required to attain the original health status, holding all else constant. As noted in Chapter 4, this will not be, in general, equal to the observed reduction in mitigating or averting behavior associated with the reduction in pollution.

In order to compute equations (7.44), (7.45), or (7.48), it is necessary to estimate a production function for the health outcome of interest and evaluate the numerator and denominator of the equation at current levels of all inputs. In practice, this has proven to be a difficult task. However, for one effort to implement this basic model, see Gerking and Stanley (1986). In addition to having data on the relevant health outcome and on ambient pollution levels, one must identify averting and mitigating behaviors and measure their costs. In practice the most effective method of reducing exposure, given ambient pollution levels, is to spend time indoors. Although the amount of time spent indoors could be measured, determining its cost would be difficult. Devices that reduce indoor pollution concentrations (air conditioners, air filters) have costs that can be measured; but they produce other services, such as reducing indoor temperature, so that it is inappropriate to allocate all of these costs to pollution avoidance. The implications of this possibility are discussed in the section on complex models.

Because of the difficulties of implementing these measures, it is useful to consider an alternative expression that shows the relationship between the observable cost of illness and marginal willingness to pay. The first step in deriving this expression is to obtain the demand functions for a and b: $a^{*}\left(I, p_{w}, p_{a}, p_{b}, c\right)$ and $b^{*}\left(I, p_{w}, p_{a}, p_{b}, c\right)$. These functions give the optimal quantities of a and b as functions of income, prices, and pollution. The second step is to take the total derivative of the health production function:

$$\frac{ds}{dc} = \frac{\partial s}{\partial a} \cdot \frac{\partial a^{*}}{\partial c} + \frac{\partial s}{\partial b} \cdot \frac{\partial b^{*}}{\partial c} + \frac{\partial s}{\partial c} \cdot \tag{7.50}$$

This gives the impact of a change in pollution on illness after taking account of the optimal adjustments of a and b to the pollution change. This expression can be rearranged as follows:

$$\frac{\partial s}{\partial c} = \frac{ds}{dc} - \frac{\partial s}{\partial a} \cdot \frac{\partial a^{*}}{\partial c} - \frac{\partial s}{\partial b} \cdot \frac{\partial b^{*}}{\partial c}, \tag{7.51}$$

and multiplied by the first-order conditions of equation (7.40)

$$-\frac{p_{a}}{\partial s / \partial a} = p_{w} - \frac{\partial u / \partial s}{\lambda} \tag{7.52}$$

to obtain

$$\begin{aligned} -p_{a} \frac{\partial s / \partial c}{\partial s / \partial a} = & \left(p_{w} - \frac{\partial u / \partial s}{\lambda}\right) \cdot \frac{ds}{dc} - \left(p_{w} - \frac{\partial u / \partial s}{\lambda}\right) \cdot \frac{\partial s}{\partial a} \cdot \frac{\partial a^{*}}{\partial c} \\ & - \left(p_{w} - \frac{\partial u / \partial s}{\lambda}\right) \cdot \frac{\partial s}{\partial b} \cdot \frac{\partial b^{*}}{\partial c} \end{aligned} \tag{7.53}$$

or, after rearranging:

$$w_c = p_w \cdot \frac{ds}{dc} + p_b \cdot \frac{\partial b^*}{\partial c} + p_a \cdot \frac{\partial a^*}{\partial c} - \frac{\partial u / \partial s}{\lambda} \cdot \frac{ds}{dc}. \tag{7.54}$$

This expression says that MWTP is the sum of the observable reductions in the cost of illness and averting activities and the monetary equivalent of the disutility of illness. The change in the cost of illness consists of the economic value of reductions in sick time and mitigating expenditures. The term $p_w\left(ds/dc\right)$ includes both actual lost wages and lost leisure time valued at the wage rate.

To compute ds/dc we do not need to estimate a health production function, but can instead estimate a dose-response function—a reduced-form relationship between illness and ambient pollution controlling for other variables that affect health status. In the health production framework a dose-response function is obtained by substituting the demand functions for b and a into the health production function. Full implementation of equation (7.54) as a measure of value therefore requires estimation of these demand functions.

As a practical matter, the first three terms in equation (7.54) can be approximated after the fact by using the observed changes in illness and averting and mitigating expenditures. In this way, equation (7.54) can be used to derive a lower bound to individual WTP. Since the last term in the equation is negative, $\partial u/\partial s < 0$, the first three terms—the value of lost time plus the change in averting and mitigating expenditures—give a lower bound to WTP. In the health literature, the term "cost of illness" typically refers only to the social cost of lost earnings plus the medical expenditures associated with illness. This term therefore ignores two components of the social cost of illness—the social value of averting expenditures and the cost of lost leisure time that result from illness.

The Marginal Value of Reduced Illness

The marginal willingness to pay for an exogenous reduction in illness falls out of the above model as a special case. Suppose that averting behavior is not possible and that mitigation (b) reduces realized s from its exogenous level s^* according to

$$s = f\left(s^*, b\right) = s^* - s\left(b\right). \tag{7.55}$$

The analogs of equations (7.41)–(7.44) are

$$\frac{dI}{ds^*} = -\frac{\partial v / \partial s^*}{\partial v / \partial I} = -\frac{\partial v / \partial s^*}{\lambda} \tag{7.41$'$}$$

$$\frac{\partial v}{\partial s^*} = \frac{\partial u}{\partial s^*} - \lambda p_w \tag{7.42$'$}$$

$$\frac{\partial v}{\partial s^*} = \lambda \cdot \frac{p_b}{\partial s / \partial b} \tag{7.43$'$}$$

$$w_{s^*} = -p_b \cdot \frac{\partial b}{\partial s^*} \,. \tag{7.44'}$$

Similarly, the analog to equation (7.53) is

$$w_{s^*} = p_w + p_b \frac{\partial b^*}{\partial s^*} - \frac{\partial u / \partial s^*}{\lambda} \,, \tag{7.54'}$$

or, marginal willingness to pay is the sum of the cost of illness (lost wages and mitigation costs) and the monetary equivalent of the lost utility (pain and suffering).

Valuing Nonmarginal Changes in Pollution

In the preceding subsection, a measure of value for a small change in health status is defined. This value is likely to be a function of health status itself. Specifically, for reasons analogous to the standard assumption of diminishing marginal utility, the marginal willingness to pay for further decreases in morbidity is likely to decrease as health status increases. In principle, this functional dependence of marginal willingness to pay on health status should be taken into account whenever large (that is, nonmarginal) changes in morbidity are being valued. The proper way to do this is straightforward in principle, but may be hard to implement in practice.

If the function relating marginal willingness to pay to health status is known, the value of the nonmarginal change is simply the integral of this function over the relevant range. Typically, however, what empirical methods produce is a point estimate of marginal willingness to pay. If it can be assumed that marginal willingness to pay is (approximately) constant over the relevant range, the value of the nonmarginal change can be calculated by multiplying the marginal willingness to pay times the change in health status. The larger the change in morbidity and the more rapidly marginal willingness to pay decreases with increases in health status, the larger the error in using this simple approach. Alternatively, one could compute a lower bound to MWTP. As shown in Chapter 4, the change in averting and mitigating expenditures, holding illness constant, constitutes a lower bound to willingness to pay for pollution improvement.

Extensions of the Health Production Model

The basic model of choice and value described in the preceding section involves a considerable amount of simplification in order to obtain some insights about the relationship between observed behavior and economic value. In this section, the model is extended in several directions toward greater realism to see to what extent the conclusions about values are preserved.

Multiple Symptoms

The model can be generalized to the case of many symptoms and various forms of averting and mitigating behavior. Consider the special case in which each mitigating activity enters the separate health production function for one symptom, $s_i(c, a, b)$. Following procedures similar to those of the basic model, it can be shown that

$$w_c = \sum p_i \cdot \frac{\partial s_i / \partial c}{\partial s_i / \partial b_i}, \tag{7.56}$$

where p_i is the price of b_i. Estimation of w_c in this case involves estimating production functions for all symptoms. Similarly, it can be shown that

$$w_c = p_a \cdot \frac{\partial s_i / \partial c}{\partial s_i / \partial a} \tag{7.57}$$

for any symptom.

Unfortunately, simple expressions such as these cannot be obtained when mitigating activities affect more than one symptom. Suppose that there are two symptoms, s_1 and s_2, neither of which restricts activities, and two mitigating activities, b_1 and b_2, with health production functions:

$$s_1 = s_1\left(c, a, b_1, b_2\right) \tag{7.58}$$

and

$$s_2 = s_2\left(c, a, b_1, b_2\right). \tag{7.59}$$

For example, b_1 could be a multi-symptom cold remedy. The first-order conditions are:

$$\frac{\partial u / \partial s_1}{\lambda} \cdot \frac{\partial s_1}{\partial b_1} + \frac{\partial u / \partial s_2}{\lambda} \cdot \frac{\partial s_2}{\partial b_1} - p_1 = 0 \tag{7.60}$$

$$\frac{\partial u / \partial s_1}{\lambda} \cdot \frac{\partial s_1}{\partial b_2} + \frac{\partial u / \partial s_2}{\lambda} \cdot \frac{\partial s_2}{\partial b_2} - p_2 = 0 \tag{7.61}$$

$$\frac{\partial u / \partial s_1}{\lambda} \cdot \frac{\partial s_1}{\partial a} + \frac{\partial u / \partial s_2}{\lambda} \cdot \frac{\partial s_2}{\partial a} - p_a = 0\,. \tag{7.62}$$

As Dickie and Gerking (1991) showed, as long as the derivatives of the health production function are known and the number of mitigating and averting activities (three here) equals or exceeds the number of symptoms (two), this system of equations can be used to solve for the unobserved utility terms, for example,

$$\frac{\partial u / \partial s_1}{\lambda}. \tag{7.63}$$

The solution can be used to substitute into the following expression to obtain a measure of willingness to pay:

$$w_c = -\frac{\partial v / \partial c}{\lambda} = -\left[\frac{\partial u / \partial s_1}{\lambda} \cdot \frac{\partial s_1}{\partial c} + \frac{\partial u / \partial s_2}{\lambda} \cdot \frac{\partial s_2}{\partial c}\right]. \tag{7.64}$$

However, if the number of symptoms exceeds the number of mitigating and averting activities, then the expression for MWTP will necessarily include some unobservable marginal utility terms.

Shifting the Cost of Illness

The measures derived above give each individual's willingness to pay to avoid pollution-induced illness. However, where social mechanisms have been developed for spreading some of the costs of illness, the social value of reducing pollution may exceed the aggregate of the individuals' willingness to pay. For example, where employers grant a certain number of days of paid sick leave, the individual does not incur any financial cost for the loss of a workday. Thus, estimates of WTP from either revealed or stated preference will be conditioned on part of these costs being shifted and will not include, in general, the lost wages component of the first term on the right of equation (7.54). Nevertheless, society loses that person's output for that day.

The burden of this lost output is borne initially by the employer, but ultimately shows up as some combination of higher prices for outputs and lower equilibrium wages. A nonmarginal reduction in pollution that resulted in fewer lost workdays would reduce the costs of paid sick leave plans. In principle, this would result in changes in general equilibrium wages and prices.

Similarly, medical insurance has the effect of shifting the marginal cost of treatment away from the sick individual. A 100 percent coverage plan reduces the price of medical treatment to zero. Thus, marginal willingness-to-pay measures based on the health production function would have to focus on changes in averting expenditures. Again, changes in the cost of treatment covered by insurance should be added to the aggregate of individuals' willingness to pay to obtain the social benefit of reduced pollution-induced illness.

As Harrington and Portney (1987) pointed out, the introduction of sick leave and medical insurance plans changes incentives by changing the prices of these activities and thus could alter individuals' choices of mitigating and averting activities and days of illness. If the effects of these plans on prices have been taken into account in estimating the models of individual behavior, then the measures of social values described here will be correct.

Health Capital and the Health Production Function

The basic model outlined above ignores any dynamic aspects of health and illness. In that model, averting or mitigating behavior and pollution in any one period of time affect only illness in that time period. It would be more in the spirit of Grossman's (1972) original treatment of the health production function to formulate the choice problem as one of optimizing investment in health capital over time. Health capital can be thought of as a measure of health status that depreciates over time unless it is maintained or augmented with some kind of investment. To formalize this, let h_t be the stock of health capital at t. Then h_t can be augmented by investment of time, y_t, in activities such as exercising and by the purchase of "health" goods, x_t. Health capital is also subject to depreciation at the rate of g percent per year. The health stock changes over time according to

$$\frac{dh_t}{dt} = I_t\left(x_t, y_t\right) - g \cdot h_t . \tag{7.65}$$

The rate of depreciation itself can be made a function of age and other characteristics.

One approach to deriving willingness to pay is to assume that health capital is an argument in the symptom production functions. Then, measures of willingness to pay similar to those of the basic model can be derived. Cropper (1977, 1981) used a different approach. She made the rate of depreciation a function of pollution, among other things. She also assumed that illness was not a direct argument in the utility function. She derived a willingness-to-pay measure that involves two terms. The first term is the value of the reduction in sick time. The second is the change in the rate of investment in health capital caused by the change in pollution. The sign of this term is ambiguous, since reducing pollution both lowers the price of an investment in health capital and increases the desired quantity of health capital. This second term is analogous to the expressions involving averting and mitigating activities. Once again, the conclusion is that true willingness to pay is different from changes in the cost of illness because of the induced changes in other activities and expenditures that affect health.

Complex Models

The models outlined so far have been simple in the sense that pollution affects only sick days and averting behavior serves only to reduce exposure to pollution. However, if either pollution or averting activities also affect utility directly, then the model becomes more complex in the sense that it is no longer possible to derive expressions for marginal willingness to pay that are functions only of observable data. Rather, the expressions for MWTP become contaminated with unobservable marginal utility and/or marginal disutility terms.

Suppose that in addition to causing illness, pollution causes disutility directly—for example, by impairing the view from one's house. Then the expression for willingness to pay in this modified version of the basic model is

$$w_c = \frac{\partial v / \partial c}{\lambda} = -\frac{(\partial u / \partial s - \lambda \cdot p_w)\partial s / \partial c + \partial u / \partial c}{\lambda}. \tag{7.66}$$

The first-order condition can be used to eliminate $(\partial u/\partial s)/\lambda$; but the unobservable $(\partial u/\partial c)/\lambda$ remains. Observations on averting and mitigating behavior are not sufficient to measure MWTP. However, if pollution varies across housing sites, then the hedonic housing price model provides a way around this problem (see Chapter 10). Similarly, suppose a generates utility in addition to reducing exposure. For example, a could be air conditioning or water filters that reduce unpleasant tastes in tap water. Then the first-order condition for a becomes

$$\left(\frac{\partial u}{\partial s} - \lambda p_w\right) = \frac{\left(\lambda p_a - \frac{\partial u}{\partial a}\right)}{\partial s/\partial a}. \tag{7.67}$$

When this is substituted into the indirect utility function, the unobservable $(\partial u/\partial a)/\lambda$ remains.

Chronic Morbidity

Now we turn to some of the problems that arise when these models of choice and value are applied to the case of chronic morbidity. One problem is how to quantify and measure chronic morbidity. Chronic morbidity is more like a state of being than like an illness where incidence and duration are the primary characteristics of concern. It may be more useful to model chronic morbidity in terms of state-dependent preferences. For example,

$$u = u(\boldsymbol{X}, s), \tag{7.68}$$

with $s = 0$ for health and $s = 1$ for chronic disease.

A second problem concerns the choice between an ex ante and an ex post perspective. We have already discussed some of these issues in the context of mortality risks. Suppose that air pollution increases the risk of chronic lung disease. The ex ante perspective asks for a healthy individual's willingness to pay to reduce the risk of incurring chronic lung disease at some time in the future. The ex post perspective would be applicable to individuals already suffering from chronic lung disease who would be asked their willingness to pay to be restored to a healthy state. As discussed in Chapter 5, the ex ante and ex post perspectives might give quite different values to the same chronic health effect, even leaving aside the likelihood that an individual who actually experiences a chronic health effect might learn something about it that would alter his willingness to pay. Which perspective is appropriate would depend on the nature of the policy question being asked. If a potential cure for the chronic condition is available, the ex post perspective yields the correct answer to the valuation question. If the policy in question is one that would reduce exposure to, say, air pollution leading to reduced risk of chronic disease, then the ex ante perspective is appropriate.

Toward Measurement

In order to implement measures based on the health production function, such as equations (7.44) and (7.54), we would require the following data for a cross section of individuals over some time period:

1 Frequency, duration, and severity of pollution-related symptoms;
2 Ambient pollution levels to which the individual is exposed;
3 Actions the individual takes to avoid or mitigate the effects of air pollution;
4 Costs of avoidance and mitigating activities;
5 Other variables affecting health outcomes (age, general health status, presence of chronic conditions, and so forth).

These data would be used to estimate health production and input demand functions, which, in turn, would be used to calculate the expression for marginal willingness to pay.

An alternative approach to the revealed preference method based on the health production function is to employ stated preference methods. For example, people could be asked directly what they would be willing to pay to reduce pollution. If willingness to pay for morbidity benefits varies with health status, age, and income, pollution control policies may have very different benefits depending on the characteristics of the target population. To value the morbidity benefits of different policies we must therefore know the distribution of key variables in the population affected by the policy, and how WTP (or the cost of illness) varies with these characteristics.

In principle, each of the valuation techniques presented above is capable of describing how the value of reduced morbidity varies with the characteristics of the respondent. The averting behavior models suggest that WTP depends on any variables that affect the marginal product of pollution, mitigating activities, or avoidance activities. In practice these variables would include health status (whether or not the respondent has a chronic respiratory condition), age, and perhaps education. WTP will also vary with factors such as earned income or education that affect the cost of averting activities. The effect of these variables on WTP could be calculated from the health production function or from equations describing the unit cost of averting activities. When WTP is estimated by a stated preference method, data are typically gathered on variables that would enter the health production function or affect the level of averting (mitigating) behavior undertaken. WTP responses can then be regressed on these variables.

Significant progress has been made in the last decade in developing and implementing stated preference approaches for valuing morbidity, either alone or in conjunction with increased mortality. Cameron and DeShazo (2013) reported on an extensive and thorough study where reduced risk of both morbidity and mortality are valued. The authors considered four possible future health states ranging from "current health" to "sickness," "recovered/remissions years," and "lost life years." Using a stated choice experiment (a form of stated preference elicitation that will be covered in Chapter 12), they elicited values associated with reducing the risk of an undesirable future health state. In addition to providing value estimates for the type of good most appropriate for many environmental and health assessment (the ex ante reduced risk of illness/death), their careful approach allows estimates of WTP that vary by age, risk beliefs, health status and other attributes. In another study that considers both mortality and morbidity reductions simultaneously, Adamowicz et al. (2011) developed a choice experiment that was implemented across a sample of Canadians. They too developed estimates of WTP that vary across sub-populations including age, gender, health status, among others and, importantly, they estimated separate values for reducing mortality and morbidity from microbial disease versus cancer risk.

A range of studies valuing reduced morbidity using stated preference methods have also appeared recently. Using a nationwide survey with a sample of over 3,500 people, Viscusi, Huber, and Bell (2012) estimated values for reducing gastrointestinal illness from drinking water. Their results confirm that these values can vary considerably across the population. Specifically, households who perceive themselves to be at higher risk, who consume large amounts of tap water, and/or who identify themselves as environmentalists, report the highest values. These findings substantiate the need for benefit-cost analysis to carefully consider the population that will be affected by a policy or regulation. Other examples of using stated preference methods to value reduced morbidity include values for foodborne illness (Teisl and Roe 2010), chronic pain (Chuck et al. 2009), skin cancer (Bateman and Brouwer 2006), and asthma (Blomquist, Dickie, and O'Conor 2011).

What are the prospects for measuring WTP using the averting behavior approach? It is usually difficult to measure the cost of activities that reduce exposure to pollution or that prevent symptoms from occurring altogether. The mere presence of an air conditioner in a home or in a car is an imperfect measure of reduced exposure to air pollution, and the cost of averting behavior is inherently difficult to measure. For example, to determine the cost of spending leisure time indoors rather than outdoors, we cannot rely on observed prices but must question individuals directly. Also measuring the cost of averting behavior is complicated by the fact that many avoidance activities produce joint products.

Two recent papers overcame these challenges using carefully collected data and survey design. Richardson, Champ, and Loomis (2012) used revealed preference information to construct a health production function for exposure to smoke from forest fires in Los Angeles County. The use of a home air cleaner is found to be an effective averting action that reduces expected symptom days. Combining the effectiveness of air cleaners with their cost yields an estimate of about $84/day to avoid a day of symptoms resulting from exposure to wildfire smoke. In contrast, Mansfield, Johnson, and Van Houtven (2006) combined revealed preference diary data on outdoor activity and a stated preference conjoint analysis of parents' WTP to avoid restrictions on their children's outdoor activities to provide an estimate of the cost of averting behavior for parents who keep their children indoors to reduce exposure to high ozone level.

Special Topics

Subtle Health Effects

In addition to the mortality and morbidity effects discussed in this chapter, there is now substantial evidence that chemical contaminants can produce subtle physical and mental changes in people. For example, relatively low levels of lead in blood have been implicated in decrements to IQ in children, in low birth weight, and in lower body growth rates and, presumably, height at maturity (U.S. Environmental Protection Agency 1986). To the extent that subtle effects such as low birth weight are associated with other adverse health effects and illness, they pose no special methodological problems from an economic perspective. The models and techniques discussed in this chapter can be used to estimate the monetary values of avoiding these adverse health effects. However, problems arise in estimating monetary values for avoiding more subtle effects such as decrements in IQ and reduced growth rates. Some studies have estimated the costs of special education and lost productivity of individuals with impaired intellectual development due to elevated levels of lead in blood (for example, U.S. EPA 1997). However, from a conceptual perspective, these cost-of-illness and lost-wage measures are only partial and incomplete measures of the total willingness to pay to avoid adverse health effects. Surely, there is a loss of utility or well-being associated with intellectual impairment; but basing estimates of the value of this loss on an individual's willingness to pay is problematic.

Valuing Health Effects on Children

The economic valuation of policies that protect or improve the health of children raises interesting ethical, as well as empirical questions, especially what normative perspective to adopt in valuation. For a discussion of some of these issues, see U.S. EPA (1999). The standard theory of welfare economics is based on the assumptions that each individual is the best judge of how well off he/she is in a given situation, and that individuals have well-defined preferences over alternative states and choose rationally among alternatives subject to the usual constraints. These assumptions define what might be called the consumer sovereignty normative perspective of welfare economics. When we turn to the welfare economics of children, there appear to be three alternative ethical or normative perspectives. The first is a natural extension of consumer sovereignty to children or "children's sovereignty." The second perspective can be termed "parental sovereignty," and would use the parents' values for changes in the health of their children. According to the third or "adult as child" perspective, values would be based on what the adult would have chosen for him/herself in childhood. None of these ethical perspectives is entirely satisfactory. Each is discussed in turn.

Children's Sovereignty

This perspective is consistent with the individualistic basis for welfare economics, however, it is not ethically attractive. Children are immature and lack the cognitive ability to make choices about health and safety. They may not have well-defined preferences over the full range of alternatives necessary to make reasoned choices, and they do not control the financial resources that are required to make tradeoffs between money and health or safety.

Parental Sovereignty

How is this perspective justified on ethical grounds? One possibility is to assert guardianship or stewardship. Another possibility is parental altruism of some form. Some authors simply presume altruism without much discussion. Others note that parents do not always seem to be the best judges of what is good for their children and sometimes engage in activities such as smoking and drinking that actually harm their children.

Parental sovereignty also has some ethically unattractive implications at a more fundamental level. The economic analysis of fertility choice emphasizes the utility that children convey to parents and the potential economic benefits they bring through providing labor for household production, and in the long term, economic security for their parents. If the marginal utility of a child (or its marginal productivity) is decreasing in the number of children, then the value to the parent of reducing the risk of death to the child or preventing disease depends on the number of children in the family and, perhaps, on its birth order.

Child as Adult

This is the perspective most consistent with the basic welfare economics principles, but one that is very difficult to implement. We cannot observe the relevant choices, so stated preference methods would have to be used, and they would impose difficult cognitive tasks on respondents.

In conclusion, the choice of an ethical or normative perspective involves some difficult questions with no easy answers. However, each perspective has implications for how we go about measuring values for children. Perhaps it is more accurate to say that each of the methods discussed in this chapter maps back to one of these perspectives. The household production function and averting behavior methods are consistent with Parental Sovereignty, since it is the choices of parents that we observe. When we attempt to use benefits transfer where the values come from adults and are adjusted on some basis, this implies the Child as Adult perspective. Finally, cost of illness measures may still have a role as lower bound estimates of value from either of the perspectives.

Quality-Adjusted Life Years

The models described above take as measures of the outcome of policy the change in the probability of dying and the reduction in days spent sick. However, as noted above, these outcome measures are simplifications of a more complex reality. Another outcome measure that has been used extensively in the health economics literature is the quality-adjusted life year (QALY). The QALY is a measure of the performance of medical treatments and interventions. It captures in a single metric two important dimensions of medical outcomes: the degree of improvement in health, and the time interval over which the improvement occurs, including any increase in the duration of life itself. Most of what is said here about QALYs applies also, with appropriate modification, to disability-adjusted life years, health-adjusted life years, and related concepts. Duration is measured in years of life; and quality is indexed by a number between 0, representing death, and 1, representing perfect health. Therefore, a treatment that is expected to increase the duration of life by 1 year of perfect health is said to produce 1 QALY. A treatment that improves health status from an index number of 0.25 to 0.75 for two years also produces 1 QALY, and a treatment that extends a life by 5 years at a quality level of 0.4 produces 2 QALYs. In the evaluation of alternative health policies or treatment programs, the numbers of QALYs produced for each patient or recipient of treatment are simply added up to obtain an aggregate measure of program effectiveness. As a summary measure with simple intuitive appeal, can the QALY be used in benefit-cost analysis to describe the outcomes of environmental policies? There are no simple answers to this question.

Many of the earliest papers on this topic relied on the judgments of medical professionals to provide the quality weights for different health states (Torrance 1986). However, the preferred approach has been to obtain the weights by some form of questioning of a sample of individuals representative of the population of interest. There are three principal ways of asking questions in order to elicit values or weights. They all start with a description of a health state including symptoms, degree or level of pain, degree of impairment of activity or function, and so forth (Gold et al. 1996; Fabian 1994).

In the simplest form of questioning, respondents are simply asked to assign a weight or numerical value between one and zero that reflects the utility they assign to the health state relative to states of perfect health and of death, respectively. Often, respondents are provided with a visual aid such as a horizontal line with a scale between zero and one marked on it.

The second approach is known as the time tradeoff approach. Respondents are asked to choose between two options: living in a given state of less than perfect health for a fixed period of time (T), for example, 5 years, and living a shorter period of time (N) in perfect health. The number of years of perfect health is varied until the individual expresses indifference between the two options. The value or weight attached to the impaired health state is simply N/T.

The third approach to questioning is the standard gamble question derived from the method first outlined by von Neumann and Morgenstern (1944) for eliciting cardinal utilities. Respondents are asked to choose between two options, where option A is living with the impaired health state with certainty for the rest of one's normal life span, and option B is a gamble in which one outcome is living for the same period of time in perfect health (with a probability of p) and the other outcome is immediate death (with a probability of $1 - p$). The probability is then varied until the individual expresses indifference between the gamble and the given health state with certainty. This means that the expected utilities of the two choices are equal. The quality weight for the given health state is simply the probability (p) that makes the individual indifferent between the two choices.

Advocates of the use of QALYs in policy evaluation cite as one advantage of the concept that QALYs are based on individuals' preferences. To the extent that QALYs are derived from the responses of representative individuals, there is something to this claim. However, as several authors have shown, given the way that QALYs are used in policy evaluation, they are consistent with utility theory and the economic theory of individual preferences only if individuals' utility functions and preference structures satisfy some quite restrictive conditions. Some examples will illustrate the nature of some of these restrictions (this discussion is based on Freeman, Hammitt, and De Civita 2002). For further discussion and illustration, readers should consult Broome (1993), Johansson (1995), Fabian (1994), Garber et al. (1996), Johnson and Lievense (2000), Hammitt (2002), Pliskin, Shepard, and Weinstein (1980), and Bleichrodt, Wakker, and Johannesson (1997). The first condition is risk neutrality over longevity, which means that an individual is indifferent to patterns of mortality risks that have the same life expectancy. For example, an individual must be indifferent between living 25 more years for certain, and a gamble offering a 50 percent chance of living 50 more years and a 50 percent chance of dying immediately.

The second condition is constant proportional tradeoff (of longevity for health), which implies that the fraction of remaining longevity an individual would trade to improve his health from one state to another (for the rest of his life) does not depend on his longevity. For example, if he is willing to give up 10 of 50 remaining years to improve his health from fair to excellent, he would also be willing to give up one of 5 remaining years for the same health improvement. Alternatively, if future QALYs are discounted, as is recommended practice (Gold et al. 1996), then it is the fraction of discounted longevity an individual is willing to give up that must remain constant (Johannesson, Pliskin, and Weinstein 1994). Johnson and Lievense (2000) cited evidence suggesting the assumption of constant proportional tradeoff is unduly restrictive.

An additional condition is that an individual's preferences for health and longevity are utility independent of his wealth and future income, which means that his preferences for risks that affect health or longevity do not depend on income. This assumption implies that the effect of income on utility is positively related to the quality weight for each health state. There is little empirical information available on this point, although the notion that the marginal utility of income is smaller in impaired health than in full health is consistent with one study (Sloan et al. 1998).

Because QALYs impose restrictive assumptions on preferences, the ranking of health interventions using QALYs may differ systematically from the ranking using WTP. For example, given these restrictions, the QALY value of reducing mortality risk within the current year to different people should be proportional to life expectancy. This implies that the value of reducing risk to a 20-year-old is about three times larger than the value of reducing risk to a 65-year-old. In contrast, individual WTP to reduce mortality risk does not fall as sharply with decreasing life expectancy, and may even increase as life expectancy declines over some range of ages. Using WTP, it is not necessarily more valuable to reduce mortality risk to a younger person, and in any case, the differential value assigned is likely to be smaller than under the QALY approach. WTP to reduce mortality risk does not fall in proportion to life expectancy because the opportunity cost of spending on risk reduction also falls with decreasing life expectancy, as the individual has less to save for (Hammitt 2000).

In summary, QALYs are not, in general, an appropriate measure for use in benefit-cost analysis and welfare assessment. Where WTP estimates of health values are available, they are a superior reflection of individuals' preferences. When WTP estimates are not available, QALYs may provide useful information but their limitations must be carefully considered.

Summary

The economic framework for valuing reductions in illness and risk of death has been described in this chapter. The aggregate benefit of a reduction in the risk of death and incidence of illness is the sum of what each of the affected people is willing to pay to reduce his or her own risks of death and illness.

For an individual to have a positive WTP for a reduction in risk of death or illness implies that the individual can perceive and is aware of changes in his health status. It does not require that the individual know that the reduction is attributable to a specific environmental change. If the value of reduced risk or symptom days is known, policy analysts can calculate benefits if they can predict the magnitude of the reduction caused by pollution control. Assumptions about individuals' knowledge do play important roles in the empirical estimation of WTP in some circumstances, however. For example, if WTP for risk reductions is to be inferred from wage-risk premia, then it must be assumed that individuals know the relative risk levels of different jobs. Also, if WTP for reduced illness is to be inferred from an individual's averting or mitigating behavior in response to changes in pollution, then it must be assumed that the individual knows the relationship between pollution and her illness experience.

In valuing risks to life, two forms of revealed preference methods have been widely used. One is based on compensating wage differentials received by workers in risky occupations; the other is based on the cost of goods that increase safety, such as seat belts and smoke detectors. Safety goods are typically 0–1 activities whose benefits exceed their costs for most persons. Thus, unless the data permit estimation of the parameters of a discrete choice model, basing risk valuations on these activities is likely to understate the value of a risk reduction. On the other hand, if the safety good also conveys utility directly, then the value of risk reduction alone will be overestimated. The compensating wage approach may also understate the value of a risk reduction because persons who are willing to be paid to accept increased risk (such as structural ironworkers) may have lower values for risk reduction than the average person.

A serious shortcoming of studies of job-risk premia and use of safety equipment is that they value only voluntarily assumed risks of accidental death (for example, risk of death on the job, or risk of dying in an auto accident). By contrast, environmental risks are largely involuntary, and may lead to a painful illness (such as cancer) before death occurs. These considerations suggest that individuals might be willing to pay more to reduce environmental risks than to reduce risk of death in an auto accident. On the other hand, to the extent that environmental risks may not manifest themselves until after a long latency period, fewer years of life will be lost than are risked from death in an auto accident during the current year. This suggests that willingness to pay estimates for a reduction in current risk of death may overstate willingness to pay to reduce environmental risks.

One advantage of stated preference methods is that the risks being valued can be tailored to circumstances relevant to environmental health policy and the most relevant populations can be sampled. For example, people can be asked their willingness to pay to reduce exposure to a pollutant that increases their risk of dying of cancer in 20 years. However, as the risk being valued becomes smaller in magnitude and more distant in time, individuals may have trouble understanding what it is they are valuing.

References

Adamowicz, Wiktor, Diane Dupont, Alan Krupnick, and Jing Zhang. 2011. "Valuation of Cancer and Microbial Disease Risk Reductions in Municipal Drinking Water: An Analysis of Risk Context Using Multiple Valuation Methods." *Journal of Environmental Economics and Management* 61(2): 213–226.

Arthur, W.B. 1981. "The Economics of Risks to Life." *American Economic Review* 71(1): 54–64.

Atkinson, Scott E., and Robert Halvorsen. 1990. "Evaluation of Risks to Life: Evidence from the Market for Automobiles." *Review of Economics and Statistics* 72(1): 133–136.

Bateman, Ian J., and Roy Brouwer. 2006. "Consistency and Construction in Stated WTP for Health Risk Reductions: A Novel Scope-Sensitivity Test." *Resource and Energy Economics* 28(3): 199–214.

Bergstrom, Theodore C. 1982. "When Is a Man's Life Worth More Than His Human Capital?" In *The Value of Life and Safety*, edited by Michael W. Jones-Lee. Amsterdam: North-Holland.

Bleichrodt, Han, Peter Wakker, and Magnus Johannesson. 1997. "Characterizing QALYs by Risk Neutrality." *Journal of Risk and Uncertainty* 15(2): 107–114.

Blomquist, Glenn. 1981. "The Value of Human Life: An Empirical Perspective." *Economic Inquiry* 19(1): 157–164.

Blomquist, Glen C., Mark Dickie, and Richard M. O'Conor. (2011). "Willingness to Pay for Improving Fatality Risks and Asthma Symptoms: Values for Children and Adults of All Ages." *Resource and Energy Economics* 33(2): 410–425.

Broome, John. 1978. "Trying to Value a Life." *Journal of Public Economics* 9(1): 91–100.

Broome, John. 1992. *Counting the Cost of Global Warming*. Cambridge: White Horse Press.

Broome, John. 1993. "Qalys". *Journal of Public Economics* 50(2): 149–167.

Cameron, Trudy A. 2010. "Euthanizing the Value of a Statistical Life." *Review of Environmental Economics and Policy* 4(2): 161–178.

Cameron, Trudy A., and J.R. DeShazo. 2013. "Demand for Health Risk Reductions." *Journal of Environmental Economics and Management* 65(1): 87–109.

Chuck, Anderson, Wiktor Adamowicz, Philip Jacobs, Arto Ohinmaa, Bruce Dick, and Saifudin Rashiq. 2009. "The Willingness to Pay for Reducing Pain and Pain-Related Disability." *Value in Health* 12(4): 498–506.

Conley, Bryan C. 1976. "The Value of Human Life in the Demand for Safety." *American Economic Review* 66(1): 45–55.

Cook, Philip J. 1978. "The Value of Human Life in the Demand for Safety: Comment." *American Economic Review* 68(4): 710–711.

Cropper, Maureen L. 1977. "Health, Investment in Health, and Occupational Choice." *Journal of Political Economy* 85(6): 1273–1294.

Cropper, Maureen L. 1981. "Measuring the Benefits from Reduced Morbidity." *American Economic Review* 71(2): 235–240.

Cropper, Maureen L., and Frances G. Sussman. 1988. "Families and the Economics of Risks to Life." *American Economic Review* 78(1): 255–260.

Cropper, Maureen L., and Frances G. Sussman. 1990. "Valuing Future Risks to Life." *Journal of Environmental Economics and Management* 19(2): 160–174.

Cropper, Maureen L., James. K. Hammitt, and Lisa A. Robinson. 2011. "Valuing Mortality Risk Reductions: Progress and Challenges." *Annual Review of Resource Economics* 3: 313–336.

Dekker, Thijs, Roy Brouwer, Marjan Hofkes, and Klaus Moeltner. 2011. "The Effect of Risk Context on the Value of a Statistical Life: A Bayesian Meta-model." *Environmental and Resource Economics* 49(4): 597–624.

Dickie, Mark. 2003. "Defensive Behavior and Damage Cost Methods." In *A Primer on Non-market Valuation*, edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Boston: Kluwer Academic.

Dickie, Mark, and Shelby Gerking. 1991. "Valuing Reduced Morbidity: A Household Production Approach." *Southern Economic Journal* 57(3): 690–702.

Evans, Mary F. and V. Kerry Smith (2010). "Measuring How Risk Tradeoffs Adjust with Income." *Journal of Risk and Uncertainty* 40(1): 33–55.

Fabian, Robert. 1994. "The Qualy Approach." In *Valuing Health for Policy: An Economic Approach*, edited by George Tolley, Donald Kenkel, and Robert Fabian. Chicago: University of Chicago Press.

Freeman, A. Myrick III, James K. Hammitt, and Paul De Civita. 2002. "On Quality Adjusted Life Years (QALYs) and Environmental/Consumer Safety Valuation." *AERE Newsletter* 22(1): 7–12.

Garber, Allen M., Milton C. Weinstein, George W. Torrance, and Mark S. Kamlet. 1996. "Theoretical Foundations of Cost Effectiveness Analysis." In *Cost-Effectiveness in Health and Medicine: Report to U.S. Public Health Service by the Panel on Cost-Effectiveness in Health and Medicine*, edited by Marthe R. Gold et al. Washington, D.C.: U.S. Department of Health and Human Services.

Gerking, Shelby, and Linda R. Stanley. 1986. "An Economic Analysis of Air Pollution and Health: The Case of St. Louis." *Review of Economics and Statistics* 68(1): 115–121.

Gold, Marthe R., D.L. Patrick, G.W. Torrance, D.G. Fryback et al. 1996. "Identifying and Valuing Outcomes." In *Cost-Effectiveness in Health and Medicine: Report to U.S. Public Health Service by the Panel on Cost-Effectiveness in Health and Medicine*, edited by Marthe R. Gold, Joanna E. Siegel, Louise B. Russell, and Milton C. Weinstein. Washington, D.C.: U.S. Department of Health and Human Services

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80(2): 223–255.

Hammitt, James K. 2000. "Valuing Mortality Risk: Theory and Practice." *Environmental Science and Technology* 34(8): 1396–1400.

Hammitt, James K. 2002. "QALYs vs. WTP." *Risk Analysis* 22(5): 985–1001.

Hammitt, James K., and Kevin Haninger. 2010. "Valuing Fatal Risks to Children and Adults: Effects of Disease, Latency, and Risk Aversion." *Journal of Risk and Uncertainty* 40(1): 57–83.

Harrington, Winston, and Paul R. Portney. 1987. "Valuing the Benefits of Health and Safety Regulations." *Journal of Urban Economics* 22(1): 101–112.

Johannesson, Magnus, Joseph S. Pliskin, and Milton C. Weinstein. 1994. "A Note on QALYs, Time Tradeoff, and Discounting." *Medical Decision Making* 14(2): 188–193.

Johansson, Per-Olov. 1995. *Evaluating Health Risks: an Economic Approach*. Cambridge: Cambridge University Press.

Johnson, F. Reed, and Katharine Lievense. 2000. *Stated-Preference Indirect Utility and Quality-Adjusted Life Years*. Prepared for Health Canada. Durham: Triangle Economic Research.

Jones-Lee, Michael W. 1974. "The Value of Changes in the Probability of Death or Injury." *Journal of Political Economy* 99(4): 835–849.

Jones-Lee, Michael W. 1976. *The Value of Life: An Economic Analysis*. Chicago: University of Chicago Press.

Kniesner, Thomas J., W. Kip Viscusi, and James P. Ziliak. 2010. "Policy Relevant Heterogeneity in the Value of Statistical Life: New Evidence from Panel Data Quantile Regressions." *Journal of Risk and Uncertainty* 40(1): 15–31.

Kochi, Ikohu, Bryan Hubbell, and Randall Kramer. 2006. "An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis." *Environmental Resource Economics* 34(3): 385–406.

Landefeld, J. Steven, and Eugene P. Seskin. 1982. "The Economic Value of Life: Linking Theory to Practice." *American Journal of Public Health* 72(6): 555–566.

Lave, Lester B., and Eugene P. Seskin. 1970. "Air Pollution and Human Health." *Science* 169(3947): 723–733.

Lave, Lester B., and Eugene P. Seskin. 1977. *Air Pollution and Human Health*. Baltimore: Johns Hopkins University Press for Resources for the Future.

Lindhjem, Henrick, Stale Navrud, and Nils Axel Braathen. 2010. "Valuing Lives Saved from Environmental, Transport, and Health Policies: A Meta-analysis of Stated Preference Studies, Organization of Economic Co-operation and Development."

Available at: http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=env/epoc/wpnep%282008%2910/final.

Linnerooth, Joanne. 1979. "The Value of Human Life: A Review of the Models." *Economic Inquiry* 17(1): 52–74.

MacLean, Douglas. 1990. "Comparing Values in Environmental Policies: Moral Issues and Moral Arguments." In *Valuing Health Risks, Costs, and Benefits for Environmental Decision Making*, edited by P. Brett Hammond, and Rob Coppock. Washington, D.C.: National Academies Press.

Mansfield, Carol, F. Reed Johnson, and George Van Houtven. 2006. "The Missing Piece: Valuing Averting Behavior for Children's Ozone Exposures." *Resource and Energy Economics* 28(3): 215–228.

Miller, Ted R. 2000. "Variations Between Countries in Values of Statistical Life." *Journal of Transportation Economics and Safety* 34(2): 169–188.

Mishan, Ezra J. 1971. "Evaluation of Life and Limb: A Theoretical Approach." *Journal of Political Economy* 79(4): 687–705.

Mrozek, Janusz R., and Laura O. Taylor. 2002. "What Determines the Value of Life? A Meta-Analysis." *Journal of Policy Analysis and Management* 21(2): 269–286.

National Research Council. 2008. *Estimating Mortality Risk Reduction and Economic Benefits from Decreasing Tropospheric Ozone Air Pollution*. Washington, D.C.: National Academies Press.

Pliskin, Joseph S., Donald S. Shepard, and Milton C. Weinstein. 1980. "Utility Functions for Life Years and Health Status." *Operations Research* 28(1): 206–224.

Richardson, Leslie A., Patricia A. Champ, and John B. Loomis. 2012. "The Hidden Cost of Wildfires: Economic Valuation of Health Effects of Wildfire Smoke Exposure in Southern California." *Journal of Forest Economics* 18(1): 14–35.

Rosen, Sherwin. 1994. "The Quantity and Quality of Life: A Conceptual Framework." In *Valuing Health for Policy: An Economic Approach*, edited by George Tolley, Donald Kenkel, and Robert Fabian. Chicago: University of Chicago Press.

Schelling, Thomas C. 1968. "The Life You Save May Be Your Own." In *Problems in Public Expenditure Analysis*, edited by Samuel B. Chase, Jr. Washington, D.C.: Brookings Institution.

Shepard, Donald S., and Richard J. Zeckhauser. 1982. "Life Cycle Consumption and Willingness to Pay for Increased Survival." In *The Value of Life and Safety*, edited by Michael W. Jones-Lee. Amsterdam: North-Holland.

Shepard, Donald S., and Richard J. Zeckhauser. 1984. "Survival versus. Consumption." *Management Science* 30(4): 423–439.

Sloan, Frank A., W. Kip Viscusi, Harrell W. Chesson, Christopher J. Conover, and Kathryn Whetten-Goldstein. 1998. "Alternative Approaches to Valuing Intangible Health Losses: The Evidence for Multiple Sclerosis." *Journal of Health Economics* 17(4): 475–497.

Sunstein, Cass R. 2004. "Valuing Life: A Plea for Disaggregation." *Duke Law Journal* 54(2): 385–445.

Sussman, Frances G. 1984. "A Note on the Willingness to Pay Approach to the Valuation of Longevity." *Journal of Environmental Economics and Management* 11(1): 84–89.

Teisl, Mario F., and Brian E. Roe. 2010. "Consumer Willingness-to-Pay to Reduce the Probability of Retail Foodborne Pathogen Contamination." *Food Policy* 35(6): 521–530.

Thaler, Richard G., and Sherwin Rosen. 1976. "The Value of Life Savings." In *Household Production and Consumption*, edited by Nester E. Terleckyj. New York: Columbia University Press.

Torrance, George W. 1986. "Measurement of Health State Utilities for Economic Appraisal: A Review." *Journal of Health Economics* 5(1): 1–30.

U.S. Environmental Protection Agency. 1986. *Air Quality Criteria for Lead*. Research Triangle Park: U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency. 1997. *The Benefits and Costs of the Clean Air Act: 1970 to 1990*. Prepared for the U.S. Congress, Washington, D.C.: U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency. 1999. *Valuing Health for Environmental Policy with Special Emphasis on Children's Health Protection*. Proceedings of the Second Workshop in the Environmental Policy and Economics Workshop Series. Available at: http://yosemite1.epa.gov/ee/epa/eerm.nsf/vwSER/D811C47399E896C48525679D0074ED76?OpenDocument (accessed March 11, 2002).

U.S. Environmental Protection Agency. 2010a. Valuing Mortality Risk Reductions for Environmental Policy: A White Paper. Draft.

U.S. Environmental Protection Agency. 2010b. *Guidelines for Preparing Economic Analyses*. Washington, D.C.: U.S. Environmental Protection Agency. Available at: http://yosemite.epa.gov/ee/epa/eed.nsf/pages/guidelines.html

U.S. Environmental Protection Agency. 2010c. *Review of 'Valuing Mortality Risk Reductions for Environmental Policy: A White Paper'*. Available at: http://yosemite.epa.gov/sab/sabproduct.nsf/d21b76bff879fa0a8525735a00766807/298E1F50F844BC23852578DC0059A616/$File/EPA-SAB-11-011-unsigned.pdf.

U.S. Office of Management and Budget. 2000. *Guidelines to Standardize measures of Costs and Benefits and the Format of Accounting Statements*. Memorandum from Jacob J. Lew, March 22, 2000, Washington, D.C.: U.S. Office of Management and Budget. Available at: www.whitehouse.gov/omb/memoranda/m0008.pdf (accessed February 19, 2002).

Usher, Dan. 1973. "An Imputation to the Measure of Economic Growth for Changes in Life Expectancy." In *The Measurement of Economic and Social Performance* vol. 38 of *Studies in Income and Wealth*, edited by Milton Moss. New York: National Bureau of Economic Research.

Viscusi, W. Kip. 1993. "The Value of Risks to Life and Health." *Journal of Economic Literature* 31(4): 1912–1946.

Viscusi, W. Kip. 2003. "Racial Differences in the Labor Market Values of a Statistical Life." *Journal of Risk and Uncertainty* 27(3): 239–256.

Viscusi, W. Kip, and Michael J. Moore. 1989. "Rates of Time Preference and Valuation of the Duration of Life." *Journal of Public Economics* 38(3): 297–317.

Viscusi, W. Kip, Joel Huber, and Jason Bell. 2012. "Heterogeneity in Values of Morbidity Risks from Drinking Water." *Environmental & Resource Economics* 52(1): 23–48.

von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Chapter 8

Environmental Quality as a Factor Input

In addition to providing services directly to individuals as consumers, environmental and resource systems can affect the costs and output levels in the production sector of an economy. Costs and output levels can be affected, for example, by changes in the flow of minerals and petroleum from the ground, the negative effects of air pollution on the flow of food and fiber from agriculture, changes in precipitation and/or temperature associated with the accumulation of greenhouse gases, and the impact of pollution on the costs of manufactured goods through requirements for more frequent cleaning, repair, and replacement of materials. The effects of these changes will be transmitted to individuals through the price system in the form of changes in the costs and prices of final goods and services and changes in factor prices and incomes. All of these examples involve a common economic mechanism. Improvements in the resource base or environmental quality lower costs and prices, and increase the quantities of marketed goods, thus leading to increases in consumers', and perhaps producers', surpluses. Similarly, increases in pollution can cause economic harm to producers and consumers by decreasing their surpluses.

Most of the early studies of the effects of air and water pollution on producers were based on the damage function approach (Freeman 1982, chs. 5, 9). This approach involves (a) estimating a dose/damage function that relates some measure of pollution to a physical measure of damage; (b) applying this function to estimates of the inventory of materials exposed or at risk; and (c) multiplying the result by some unit value.

For example, in the case of damages to materials and structures, the physical damage might be corrosion, soiling, or loss of thickness of paint. This measure would have to be translated into an estimate of the increase in the frequency of some repair or replacement activity. The unit cost of this activity would provide the basis for estimating a monetary damage. The problem with this approach is that paint thickness and corrosion rates are not the economically relevant impacts. Rather, what is required is some understanding of the responses of producers in terms of increased frequency of repair, degree of degradation of performance, and so forth, all of which translate into increases in the cost of production. Similarly, in the case of agriculture the damage function approach focuses on reduction in harvestable yield and multiplies this by a market price. However, this approach ignores adaptive behavior on the part of farmers and impacts on consumers resulting from possible changes in market price and changes in the cost of production of a crop.

Properly specified economic models of the effects of pollution on producers make use of cost functions or production functions to link the physical effects of changes in environmental quality to changes in market prices and quantities, and ultimately to changes in consumers' and producers' surpluses. Either directly, in the case of the production function approach, or implicitly, in the case of the cost function approach, these models incorporate the whole range of possible producers' responses to changes in pollution levels (through, for example, material substitution, increased protection activities, and changes in maintenance and repair schedules).

In the first section of this chapter, the theory is laid out in more detail for the case of single-product firms in a competitive industry. In the second section, the analysis is developed for the more likely case of multiproduct firms. In the third section, the method through which welfare effects on factor owners and consumers are passed through vertically linked markets for inputs and intermediate products when there are several intermediate stages of production are examined. In the fourth section, the effects that market distortions, such as monopoly power, have on welfare measures are considered. In the fifth section, simple models describing how the methods presented in this chapter can be used to value changes in the productivity of natural resource systems such as commercially exploited forests and fisheries are presented. For other treatments of many of these topics, see Bockstael and McConnell (2007), Just, Hueth, and Schmitz (2004), and McConnell and Bockstael (2005). The basic theory for estimation of productivity benefits is outlined in Chapter 4.

Basic Theory[1]

The purpose of this section is to derive welfare measures for changes in some parameter, q, which enters directly into the production functions of single-product firms. This parameter can be interpreted as a measure of environmental quality or the quality of a resource input into production, or alternatively, it could be a measure of innovation or technological change. Whatever the interpretation, an increase in this variable is assumed to increase the output attainable with any given set of inputs. Following on the earlier contributions of Anderson (1976), Schmalensee (1976), and Just and Hueth (1979), Just, Hueth, and Schmitz (2004) provided a rigorous analysis of how to measure changes in welfare due to price distortions in factor and product markets. These models provide a basis for analyzing the effects of productivity-induced changes in product and factor prices. As we will see, in the case of single-product firms the results that they derive for

1 This section is adapted, with permission, from Freeman and Harrington (1990).

price distortions carry over in a straightforward manner to the case of parametric shifts of cost and supply functions caused by environmental changes.

In this section, two alternative measures of welfare change for both marginal and nonmarginal changes in q for the case of a single-product firm are considered. One measure is the value of the marginal product of q. The other measure is based on the aggregate cost function for the industry and can be interpreted as the area between the old and new supply curves. The analysis in this section is essentially short-run, focusing on changes in quasi-rents to firms and on consumers' surpluses. In the long run, quasi-rents are competed away, except for those accruing to specialized factors owned by the firms. These rents can be viewed as increases in the prices of such factors.

Consider a competitive industry with N firms producing a single good, x. The ith firm is assumed to have a production function $x_i = x_i\left(\boldsymbol{V}_i, k_i, q\right)$, $\left(i = 1,\ldots,N\right)$, where $\boldsymbol{V}_i \equiv \left(v_{i1},\ldots,v_{iJ}\right)$ is a vector of variable factor inputs to firm i, k_i is a fixed factor, and q measures environmental or resource quality. For example, q could be a measure of air or water quality or natural soil fertility. Alternatively, q could be some measure of technical efficiency such as the rate at which an electrical generating station converts thermal energy to electricity. In this case, the model developed here would measure the benefits of investments in increasing technical efficiency. To simplify the notation, the k_i terms will be suppressed in all that follows. Both $\boldsymbol{V}_i$ and q have positive marginal products, but q is given exogenously to the industry.

Assume that the industry faces perfectly elastic supplies for all factor inputs at prices f_j, or in vector notation, $\boldsymbol{F}$. If factor supplies are less than perfectly elastic, the changes in rents to factor owners as factor prices change must also be included in the social welfare measure. Also, as Just, Hueth, and Schmitz (2004, ch. 9 and its appendix) show in their analysis of multimarket changes, general equilibrium supply and demand curves must be used, as discussed below.

Let aggregate industry output be denoted by $y = \sum_{i=1}^{N} x_i$. The industry faces an inverse demand function for its product, $p = p\left(y\right)$, where p is the market price, and income and all other prices are assumed constant. Define an aggregate production function

$$y = y\left(v_{11},\ldots,v_{NJ},q\right) \tag{8.1}$$

to be the sum of the N firm production functions.

The Production Function Approach

Now assume that demand functions are compensated so that consumer welfare changes can be measured by the appropriate areas. Then the social welfare W associated with producing y is the area under the demand curve for y, less the cost of the inputs:

$$W\left(v_{11}, ..., v_{NJ}, q\right) = \int_0^y p(u)\,du - \sum_{i=1}^{N} \boldsymbol{F} \cdot \boldsymbol{V}_i \ . \tag{8.2}$$

The first-order conditions for choosing the v_{ij} so as to maximize social welfare are

$$\frac{\partial W}{\partial v_{ij}} = \frac{\partial y}{\partial v_{ij}} \cdot p(y) - f_j = 0 \quad \text{for all } i, j \ . \tag{8.3}$$

Because each firm is a price taker, this welfare optimum is also the competitive equilibrium. These first-order conditions define input demand functions $v_{ij}^*\left(f_j, q\right)$, and, in turn, an output function

$$y^*(q) = y\left[v_{11}^*(f_1, q), \ldots, v_{NJ}^*\left(f_J, q\right), q\right] \tag{8.4}$$

and a social welfare function

$$W(q) = y\left[v_{11}^*(f_1, q), \ldots, v_{NJ}^*\left(f_J, q\right), q\right]. \tag{8.5}$$

The asterisks indicate optimally chosen quantities. In what follows, the f_j will not appear as function arguments, except where factor prices are assumed to be variable.

Using the envelope theorem, we have

$$\frac{\partial W}{\partial q} = p\left(y^*\right) \frac{\partial y^*\left[v^*(q), q\right]}{\partial q}. \tag{8.6}$$

The net welfare gain from an increase in q is the value of the marginal product of q in the production function. Note that $\partial y^* / \partial q$ is not the observed increase in y^*, but rather, it is the increase in y that would occur holding all other inputs constant. Knowledge of the production function is required to implement this measure.

The Cost Function Approach

The variable cost functions for firms can be added to obtain an aggregate variable cost function for the industry:

$$C = C\left[y^*(q), q\right] \ . \tag{8.7}$$

Assuming profit maximization with price equal to marginal cost, this also defines a market supply curve. The total contribution to social welfare made by the production and consumption of this good is the sum of compensating surpluses and quasi-rents to firms:

$$W = \int_0^y p(u)\,du - C\left(y^*, q\right). \tag{8.8}$$

This, of course, is maximized when output is set where price equals marginal cost,

$$p(y) = \frac{\partial C\left(y^*, q\right)}{\partial y} \ , \tag{8.9}$$

and, as before, the optimal output is a function of the parameter q.

Again applying the envelope theorem, we have the following expression for welfare change associated with marginal changes in q:

$$\frac{\partial W}{\partial q} = -\frac{\partial C\left(y^*, q\right)}{\partial q}. \tag{8.10}$$

This marginal value can be calculated if the cost function is known. The duality of the cost function and the production function ensures that the two measures of marginal welfare change given by equations (8.6) and (8.10) are equivalent.

The change in total welfare is positive if increasing q reduces costs. However, this fact does not necessarily imply that both producers' surplus and compensating surplus must increase. Compensating surplus (CS) is given by the area under the compensated demand curve less the actual expenditure, or

$$CS(q) = \int_0^{y^*} p(u)\,du - \left[p\left(y^*\right) \cdot y^*(q)\right]. \tag{8.11}$$

The marginal effect of a change in q is the derivative of this expression, or

$$\frac{\partial CS}{\partial q} = -\frac{\partial p}{\partial y^*} \cdot \frac{\partial y^*}{\partial q} \cdot y^*. \tag{8.12}$$

Of the three terms in equation (8.12), $\partial p / \partial y^* \leq 0$ by the law of demand, and $y^* > 0$. Thus, CS will decrease if, and only if, equilibrium output y^* decreases with an increase in q. This possibility cannot be ruled out.[2] It is possible for the marginal cost curve to increase over the relevant range, leading to an increase in price, even though total cost is reduced. Therefore, it is possible for consumers to be made worse off by an increase in q.

The marginal change in producers' quasi-rent, R, is the marginal welfare change given by equation (8.10) less the change in consumers' surplus given by equation (8.12); that is,

$$\frac{\partial R}{\partial q} = -\frac{\partial C}{\partial q} + \frac{\partial p}{\partial y^*} \cdot \frac{\partial y^*}{\partial q} \cdot y^*. \tag{8.13}$$

2 To determine the sign of $\partial y^* / \partial q$, differentiate the first-order condition (8.9):

$$\frac{\partial p}{\partial y^*} \cdot \frac{\partial y^*}{\partial q} = \left(\frac{\partial^2 C}{\partial y^{*2}} \cdot \frac{\partial y^*}{\partial q}\right) + \frac{\partial^2 C}{\partial y^* \partial q},$$

and rearrange to obtain

$$\frac{\partial y^*}{\partial q} = \frac{\left(\partial^2 C / \partial y^* \partial q\right)}{\left[\left(\partial p / \partial y^*\right) - \left(\partial^2 C / \partial y^{*2}\right)\right]}.$$

Since the denominator of this expression is always negative, $y^*(q)$ is decreasing whenever $\partial^2 C / \partial y^* \partial q$ is positive.

The first term is positive, while the second term is usually negative. Thus, quasi-rent could decrease only if the second term were larger in absolute value. The second term can also be expressed in terms of the price elasticity of demand:

$$\varepsilon = \frac{\partial y^*}{\partial p} \cdot \frac{p}{y^*} < 0 \cdot \tag{8.14}$$

When this expression is substituted into equation (8.13), we see that the magnitude of the second term varies inversely with the elasticity of demand. For sufficiently inelastic demands, the fall in price brought about by the increase in quantity could actually harm producers by reducing their quasi-rents.

Nonmarginal Changes in q

For nonmarginal changes in q, say a change from q^0 to q^1, the aggregate benefit can be found by integrating either equation (8.6) or equation (8.10):

$$\begin{aligned} W_q &= \int_{q^0}^{q^1} p\left(y^*\right) \cdot \frac{\partial y\left[V^*(q), q\right]}{\partial q} dq \\ &= -\int_{q^0}^{q^1} \left\{ \frac{\partial C\left[y^*(q), q\right]}{\partial q} \right\} dq. \end{aligned} \tag{8.15}$$

As equation (8.15) indicates, both of these terms require integration along a path. To implement them, not only are we required to know either the production function or the cost function, but we also must know how equilibrium output changes with q, or how the levels of inputs change with q, or both.

However, W_q may be calculated directly from the demand curve and the cost function if the initial and final output levels y^{*0} and y^{*1} are known. This calculation is the change in the area bounded by the demand curve and cost curve:

$$W_q = \int_0^{y^{*1}} p(y)dy - C\left(y^{*1}, q^1\right) - \int_0^{y^{*0}} p(y)dy - C\left(y^{*0}, q^0\right) \cdot \tag{8.16}$$

This is shown in Figure 8.1, where $p_y(y)$ is the inverse demand curve and $MC_y\left(q^0\right)$ and $MC_y\left(q^1\right)$ indicate the industry marginal cost curves as a function of y^* and q. Graphically, equation (8.16) is equivalent to

$$\begin{aligned} W_q &= (a+b+c+d+e) - (c+e) - (a+b+c) + (b+c) \\ &= (b+d). \end{aligned} \tag{8.17}$$

The welfare change can be represented diagrammatically by the area between the old and new cost curves bounded by the demand curve. The division of the gain between producers and consumers is also shown graphically in Figure 8.2. Consumers gain the areas $u + v + w$, and producers lose area w but gain $s + t + u$.

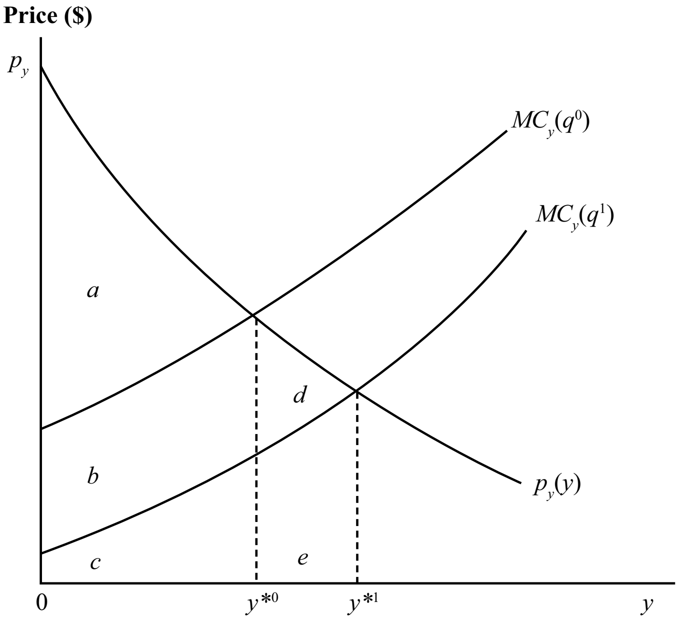

Figure 8.1 The welfare change for single-product firms

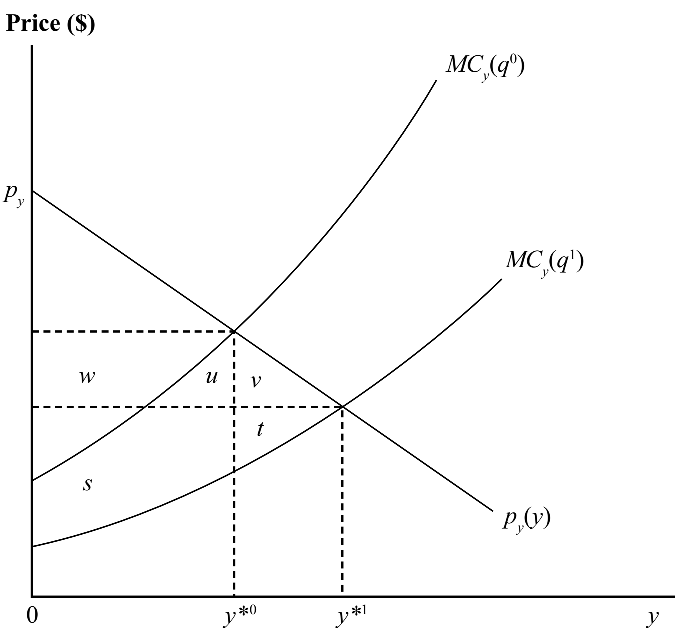

Figure 8.2 The division of the welfare gain between consumers and producers

The net social gain is $s + t + u + v$. If the cost and the demand functions are known, these areas can be calculated by taking the appropriate integrals.

Conclusions

In this section, two alternative measures are identified for the welfare change associated with a change in a nonmarket input such as q. One is based on shifts in the cost or supply curves (the cost function approach); the other is based on the direct impact of the change in q on the production function (the production function approach). Of course, because of duality, they are both based on essentially the same information and must give the same results. As a practical matter, the choice of approach depends on the availability of data.

In some cases, it may be possible to estimate the production function from data on physical inputs and outputs and environmental quality. For example, the availability of experimentally derived crop loss functions for ozone (see Heck et al. 1983) makes it feasible to use the production function approach to measure the benefits to agriculture of reducing ozone air pollution. Using time series data, Schlenker and Roberts (2009) estimated the relationship between corn, soybean, and cotton yields and temperature to assess the degree to which global warming scenarios will impact agriculture. Similarly, Auffhammer, Ramanathan, and Vincent (2006) estimated the effect that atmospheric black carbon (so-called brown clouds) and greenhouse gases have on rice harvests through their effects on solar radiation levels. Alternatively, one could estimate the cost function for an industry (Mathtech 1982) or the rent or profit function (Mjelde et al. 1984).

The production function and cost function methods are based on observations of the optimizing behavior of producers. An alternative to observing this behavior is to simulate it through formal optimization models. One example of this approach is the model used by Adams, Hamilton, and McCarl (1984) and Adams and McCarl (1985) to estimate the effects on the agricultural sector of controlling ozone air pollution. They combined a mathematical programming model of representative farms in the five Corn Belt states with an experimentally derived damage function to simulate producers' responses to changes in ozone levels. In another study of the economics of controlling ozone air pollution and agriculture, Kopp et al. (1985) constructed supply functions from region-specific farm budget and farm practice data. This model lacks the formal optimization characteristics of the Adams and McCarl model, but it allows for more detailed treatment of differences in farm economics across regions.

Multiproduct Firms

In the case of multiproduct firms with joint production technologies, things are not so simple. For marginal changes, the production function and cost function approaches can still be applied with appropriate modification. However, for nonmarginal changes, it is not correct simply to add up areas between observed shifts of marginal cost or supply curves. When production is characterized by jointness, the model used for measurement must take account of the interconnectedness of the marginal cost or supply functions for the various outputs. The gain in welfare for an increase in q is measured by the sum of the changes in consumers' surpluses in the markets for all of the affected products plus the aggregate change in quasi-rents to the affected firms. However, in measuring the changes in quasi-rents, it is necessary to take account of how changes in the output of one good affect the marginal cost curves of the other goods being jointly produced.

For notational simplicity, assume there are only two products. The other notation is unchanged. The production function for the ith firm is given by

$$x_i\left(x_{1i}, x_{2i}, v_{i1}, \ldots, v_{iJ}; q\right) \geq 0. \tag{8.18}$$

That is, $x_i \geq 0$ is feasible production and $x_i = 0$ is efficient production. Also, assume $\partial x_i/\partial q \geq 0$, which assures that increasing q will increase the welfare associated with the production of x_{1i} and x_{2i}.

Both the aggregate production function and cost function approaches still provide a straightforward basis for measuring the marginal welfare change. As before, assume that each firm is a price taker in all of its product and factor markets and that the industry faces infinitely elastic factor supply curves. In a fashion similar to the analysis leading to equations (8.6) and (8.10) define the aggregate production function as the sum of the N firm production functions:

$$y\left(y_1, y_2, \boldsymbol{V}, \mathrm{q}\right) = 0 \ , \tag{8.19}$$

where $y_1 = \sum_{i=1}^{N} x_{1i}$, $y_2 = \sum_{i=1}^{N} x_{2i}$, and $\boldsymbol{V} = \left(v_{11}, \ldots, v_{NJ}\right)$ is the vector of all inputs for all firms. As before, the social welfare function is

$$W(\boldsymbol{V}, q) = \int_0^{y_1} p_1\left(u_1\right) du_1 + \int_0^{y_2} p_2\left(u_2\right) du_2 - \sum_{i=1}^{N} \boldsymbol{F} \cdot \boldsymbol{V}_i \ , \tag{8.20}$$

subject to $y\left(y_1, y_2, \boldsymbol{V}, q\right) = 0$.

The first-order conditions for a maximum of social welfare are

$$W_1 = p_1\left(y_1\right) - \lambda \partial y/\partial y_1 = 0 \, , \tag{8.21}$$

$$W_2 = p_2\left(y_2\right) - \lambda \partial y/\partial y_2 = 0 \, , \tag{8.22}$$

$$W_{vj} = -f_j - \lambda \partial y/\partial v_j = 0 \, , \tag{8.23}$$

where λ is the Lagrangian multiplier on the production constraint. These conditions define optimal $y_1^*(q)$, $y_2^*(q)$, and $\boldsymbol{V}^*(q)$ in the usual way. By the envelope theorem,

$$\partial W/\partial q = -\lambda \cdot \partial y/\partial q. \tag{8.24}$$

If we now differentiate the production function with respect to q, we have

$$\left(\frac{\partial y}{\partial y_1}\cdot\frac{\partial y_1^*}{\partial q}\right)+\left(\frac{\partial y}{\partial y_2}\cdot\frac{\partial y_2^*}{\partial q}\right)+\frac{\partial y}{\partial q}+\left(\sum_j\frac{\partial y}{\partial v_j}\cdot\frac{\partial v_j^*}{\partial q}\right)=0\,. \tag{8.25}$$

Combining this with equation (8.24) yields

$$\begin{aligned}\frac{\partial W}{\partial q}&=-\lambda\cdot\left(\frac{\partial y}{\partial q}\right)\\ &=\left(\lambda\cdot\frac{\partial y}{\partial y_1}\cdot\frac{\partial y_1^*}{\partial q}\right)+\left(\lambda\cdot\frac{\partial y}{\partial y_2}\cdot\frac{\partial y_2^*}{\partial q}\right)\\ &\quad+\frac{\partial y}{\partial q}+\left(\sum_i\lambda\cdot\frac{\partial y}{\partial v_j}\cdot\frac{\partial v_j^*}{\partial q}\right)=0.\end{aligned} \tag{8.26}$$

Substituting terms from the first-order conditions yields

$$\frac{\partial W}{\partial q}=\left[p_1\left(y_1^*\right)\cdot\frac{\partial y_1^*}{\partial q}\right]+\left[p_2\left(y_2^*\right)\cdot\frac{\partial y_2^*}{\partial q}\right]-\sum_j f_j\frac{\partial v_j^*}{\partial q}\,. \tag{8.27}$$

Alternatively, in terms of the aggregate cost function, $C\left(y_1^*, y_2^*, q\right)$, the benefits are

$$\frac{\partial W}{\partial q}=-\frac{\partial C\left(y_1^*, y_2^*, q\right)}{\partial q}\,. \tag{8.28}$$

In the case of nonmarginal changes in q, it is tempting to say that all that needs to be done is to add the areas between the new and old supply curves for each of the products of the multiproduct firms, but that would be wrong. Because of the interdependencies of the marginal cost curves under joint production, adding up these areas will not give a correct measure of the change in quasi-rents. The benefit of a nonmarginal change is the increase in the social values of the outputs, net of any changes in the joint cost of production. Again, for a change from q^0 to q^1 we have

$$W_q=\int_{y_1^0}^{y_1^1}p_1(y_1^*)dy_1+\int_{y_2^0}^{y_2^1}p_2(y_2^*)dy_2+\left[C\left(y_1^0, y_2^0, q^0\right)-C\left(y_1^1, y_2^1, q^1\right)\right]. \tag{8.29}$$

The change in joint costs, ΔC, can be decomposed into three steps:

$$\begin{aligned}\Delta C=&\left[C\left(y_1^0, y_2^0, q^0\right)-C\left(y_1^0, y_2^0, q^1\right)\right]\\ &+\left[C\left(y_1^0, y_2^0, q^1\right)-C\left(y_1^1, y_2^0, q^1\right)\right]\\ &+\left[C\left(y_1^1, y_2^0, q^1\right)-C\left(y_1^1, y_2^1, q^1\right)\right].\end{aligned} \tag{8.30}$$

Substituting this into equation (8.29) gives

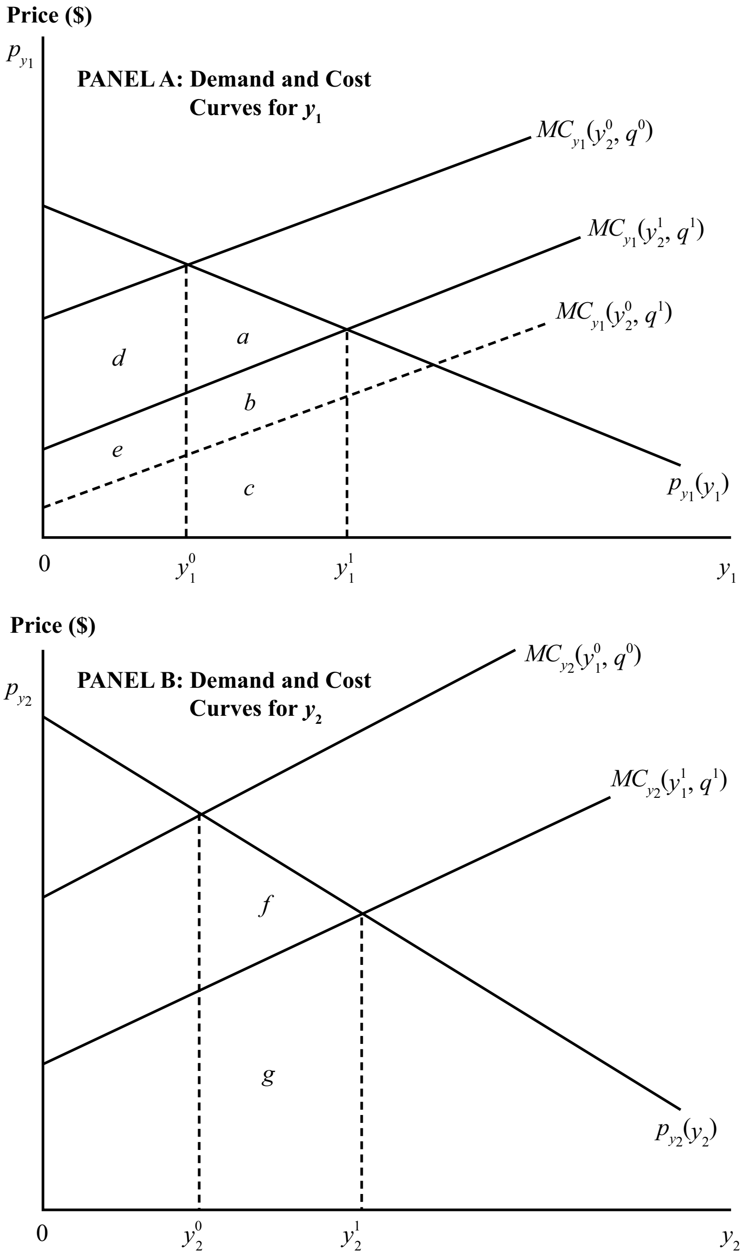

Figure 8.3 The welfare measure for multiproduct firms when outputs are substitutes in production

$$W_q = \int_{y_1^0}^{y_1^1} p_1(y_1^*)dy_1 + \int_{y_2^0}^{y_2^1} p_2(y_2^*)dy_2 - \int_{q^0}^{q^1} \left[\frac{\partial C\left(y_1^0, y_2^0, q\right)}{\partial q}\right] dq$$
$$- \int_{y_1^0}^{y_1^1} \left[\frac{\partial C\left(y_1, y_2^0, q^1\right)}{\partial q}\right] dy_1 - \int_{y_2^0}^{y_2^1} \left[\frac{\partial C\left(y_1^1, y_2, q^1\right)}{\partial q}\right] dy_2. \tag{8.31}$$

Note that in each of the markets one of the two marginal cost curves used in defining the welfare measure presented here is not actually observed. Implementation of equation (8.31) requires full knowledge of the joint cost function. Basing measures on comparisons of observed marginal cost curves before and after the change will result in error, the sign and magnitude of which will depend on the specific characteristics of the joint technology.

A graphical interpretation of equation (8.31) may prove to be helpful. First consider the case where y_1 and y_2 are substitutes in production in the sense that $\partial^2 C/\partial y_1 \partial y_2 = \partial^2 C/\partial y_2 \partial y_1 > 0$. This case is shown in Figure 8.3. The solid cost curves represent the pre- and post-change observed marginal cost curves. The dashed cost curve in panel A corresponds to the fourth integral in (8.31), where q has changed holding y_2 at y_2^0.

The geometric areas in panels A and B corresponding to the five integrals in equation (8.29) are given by the following:

$$W_q = (a + b + c) + (f + g) + (d + e) - c - g$$
$$= a + b + d + e + f. \tag{8.32}$$

The welfare change is measured by the area between the two cost curves for y_1 holding y_2 at y_2^0 plus the observed welfare triangle f in the market for y_2 (panel B). An empirical measure based on areas between observed cost curves would yield an underestimate of the true welfare change.

In contrast, Figure 8.4 shows the case where y_1 and y_2 are complements in production—that is $\partial^2 C/\partial y_1 \partial y_2 = \partial^2 C/\partial y_2 \partial y_1 < 0$ geometric equivalent to equation (8.31) is

$$W_q = (a + b + c) + (f + g) + d - (b + c + e) - g$$
$$= a + d - e + f. \tag{8.33}$$

Again, measuring the first term requires knowledge of the unobserved cost curve for y_1 after the change in q but holding y_2 at y_2^0.

It is of interest to know the magnitude of the error that would result from failure to recognize the existence of joint costs. To get an approximation of the error, form a second-order Taylor polynomial expanded around $\left(y_1^0, y_2^0, q^0\right)$. The approximate change in cost is

$$\Delta C = C\left(y_1^1\ y_2^1, q^1\right) - C\left(y_1^0\ y_2^0, q^0\right) = \Delta C^1 + \Delta C^2 + \Delta C^J, \tag{8.34}$$

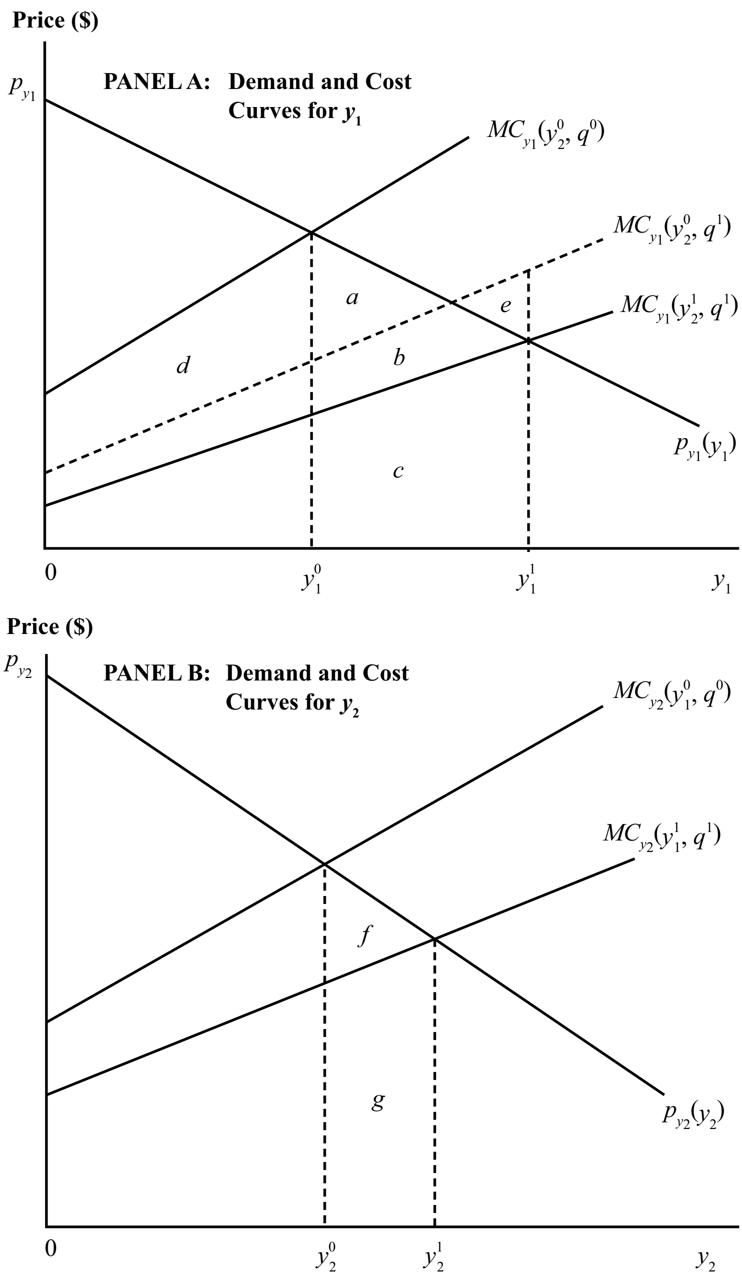

Figure 8.4 The welfare measure for multiproduct firms when outputs are complements in production

where

$$\Delta C^1 = \left[\left(y_1^1 - y_1^0\right)\cdot\frac{\partial C}{\partial y_1}\right] + \left[\frac{1}{2}\left(y_1^1 - y_1^0\right)^2\cdot\frac{\partial^2 C}{\partial y_1{}^2}\right] + \left[\left(y_1^1 - y_1^0\right)\left(q^1 - q^0\right)\frac{\partial^2 C}{\partial y_1 \partial q}\right] \tag{8.35}$$

$$\Delta C^2 = \left[\left(y_2^1 - y_2^0\right)\cdot\frac{\partial C}{\partial y_2}\right] + \left[\frac{1}{2}\left(y_2^1 - y_2^0\right)^2\cdot\frac{\partial^2 C}{\partial y_2}\right] + \left[\left(y_2^1 - y_2^0\right)\left(q^1 - q^0\right)\frac{\partial^2 C}{\partial y_2 \partial q}\right] \tag{8.36}$$

$$\Delta C^J = \left(y_1^1 - y_1^0\right)\left(y_2^1 - y_2^0\right)\frac{\partial^2 C}{\partial y_1 \partial y_2}\cdot \tag{8.37}$$

The third term, ΔC^J, is the contribution to the change in cost by the jointness in y_1 and y_2.

There is another way of looking at the difficulties involved here. Recall that one component of the welfare change is the change in quasi-rents to firms. With joint production, it is generally not possible to measure the quasi-rent of a multiproduct firm from data from just one market. However, Just, Hueth, and Schmitz (2004) have shown that in the case of price changes, there are circumstances in which it is possible to measure the change in quasi-rents in either a single-factor market or a single-product market. Their analysis can be extended to the case of parametric shifts in the production technology. What is required is that there is either a necessary input or a necessary output. A necessary output is one for which there is some positive minimum price at which the firm will choose to stop producing not only that output but all other products as well—in other words, production will shut down completely. All of the quasi-rent to a firm can be attributed to the necessary output and can be measured by integrating above the supply curve for that output from the shutdown price to the current market price. Alternatively, the quasi-rent can be measured by the area under the demand curve for any necessary factor input. An input is deemed to be necessary if there is some price for that input at which its derived demand falls to zero, and if all of the firm's outputs fall to zero when that input is set at zero.

The measurement of changes in quasi-rents can be shown graphically for the case of two necessary outputs, y_1 and y_2, and one necessary input, v_1. Suppose that, as shown in Figure 8.5, an increase in q shifts the supply functions for both outputs and the demand function for the necessary input to the right. The consumer surplus component of the welfare change is the areas $w_1 + v_1 + w_2 + v_2$ in panels B and C. The increase in the quasi-rents to all firms can be measured alternatively by $u_1 + t_1 - w_1$ in panel B, or $u_2 + t_2 - w_2$ in panel C, or b in panel A.

To see this, first suppose that y_1 is the necessary output. At q^0 if p_1 is at p_1^0, other things being equal, firms shut down and quasi-rents are zero. If p_1 increases to p_1^1,

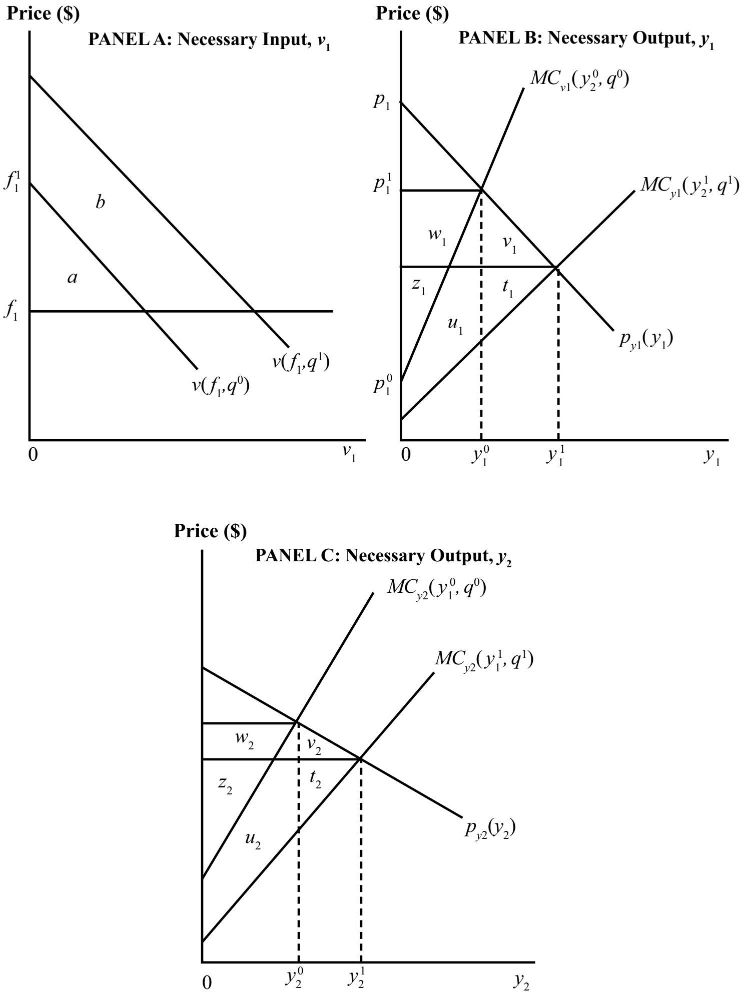

Figure 8.5 The welfare gain for multiproduct firms with necessary outputs

firms produce both goods in positive quantities; but all of the quasi-rents are attributable to the increase in p_1 and are measured by the area $w_1 + z_1$. Comparing quasi-rents at q^1 and q^0 gives $\Delta R = R^1 - R^0 = u_1 + t_1 + z_1 - w_1 - z_1 = u_1 + t_1 - w_1$. For the necessary input v_1, if its price is f_1^1, firms shut down and quasi-rents are zero. If f_1 falls, firms produce both goods. All of the quasi-rents are attributable to the increased use of v_1 and are measured by the area a.

The three alternative measures of the total welfare gain are:

$$y_1 \text{ is necessary:} \qquad W_{1q} = w_1 + v_1 + t_1 + w_2 + v_2 + u_1 - w_1 \qquad (8.38)$$
$$= u_1 + v_1 + t_1 + w_2 + v_2$$

$$y_2 \text{ is necessary:} \qquad \begin{aligned} W_{2q} &= w_1 + v_1 + w_2 + v_2 + u_2 + t_2 - w_2 \\ &= w_1 + v_1 + u_2 + v_2 + t_2 \end{aligned} \tag{8.39}$$

$$v_1 \text{ is necessary:} \qquad W_{vq} = W_1 + v_1 + w_2 + v_2 + b. \tag{8.40}$$

Where empirical measures of producers' benefits have been reported in the literature, production has typically been modeled either implicitly or explicitly as being nonjoint, even in those cases where multiproduct firms are typical. For example, Adams, Hamilton, and McCarl (1984), Adams and McCarl (1985), Kopp et al. (1985), and Adams, Crocker, and Katz (1984) have all provided estimates of the benefits to agriculture of controlling ozone air pollution, using variations of the cost function approach. The first two studies used supply functions derived from models of farm behavior rather than from econometric estimates. Independence of the marginal cost or supply curves by crop was a characteristic of these models. Adams, Crocker, and Katz (1984) used econometric estimates of supply functions for four crops; but the estimating equations imposed nonjointness on the production technology even though many farms produced at least two of the four crops modeled (corn and soybeans). Other researchers, such as Mjelde et al. (1984), Garcia et al. (1986), and Mathtech (1982), aggregated across products to obtain industry profit or cost functions. However, consistent aggregation of this sort is possible only if marginal cost functions are independent, a condition that is not satisfied when there is joint production. The analysis presented here suggests that in the future it will be important to confront the question of joint production directly. If in fact the industries being studied are characterized by joint production, the empirical models of production must reflect this; and the measures of welfare change must take this into account in the way just described.

Vertically Linked Markets

The next case to consider is that of vertically linked markets in which the output of one set of firms is purchased as an input by another set of firms. Assume that q affects costs and prices in one industry that is part of a set of vertically linked industries. The model used to derive welfare measures in this case must take account of the fact that every price change affects both buyers and sellers, but in different directions; that is, for a price increase, the buyer loses while the seller gains, and vice versa for price decreases. For simplicity, in what follows assume that all firms are single-product firms. Some firms are producers that purchase primary factor inputs and sell an intermediate product y_w. Other firms buy the intermediate product and factor inputs and sell y_r at retail to consumers. Assume that all factor supply curves but one are infinitely elastic; the exception is the supply curve for labor used in the wholesale industry, v_w. Assume also that the compensated supply curve for v_w is upward sloping and that y_w and y_r are produced and sold in perfectly competitive markets. Figure 8.6 shows the case.

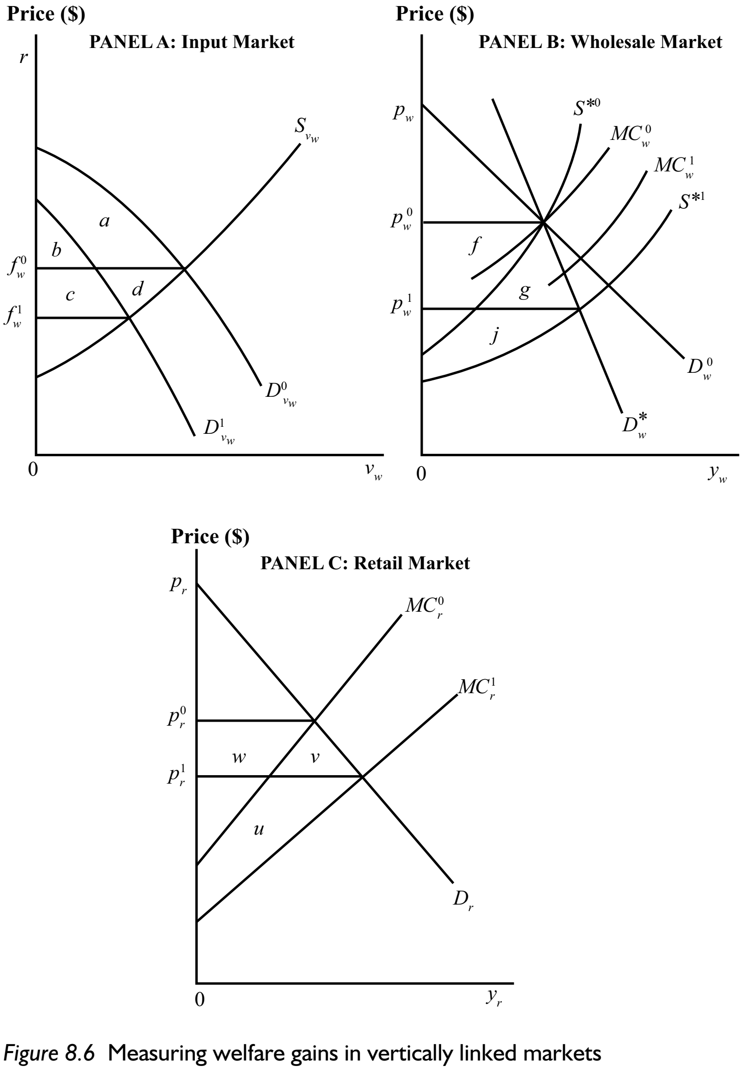

Figure 8.6 Measuring welfare gains in vertically linked markets

Assume now that q affects the producers of y_w. Suppose that at the initial level of q the supply function for y_w is MC_w^0 and the demand curve is D_w^0 as shown in panel B of Figure 8.6. The first effect of an increase in q is to shift MC_w down to MC_w^1. This results in a similar downward shift of MC_r and a decrease in the retail price, as shown in panel C of Figure 8.6. In the wholesale market, the demand curve for y_w is derived on the assumption that the retail price is constant, but when the retail price falls, the derived demand for y_w shifts to the left. The wholesale producers respond to the changes in demand along a path that reflects the adjustments in price in the retail market. The path of this adjustment can be described by what Just, Hueth, and Schmitz (2004) call a general equilibrium demand curve. This demand curve shows the maximum willingness to pay by the y_r industry given its costs and the retail price at which it must sell its product, and assuming that all other input prices are constant. This is shown as D_w^* in panel B of Figure 8.6.

At the same time, the demand curve for v_w in panel A of Figure 8.6 is shifting for two reasons. First, the decrease in p_w pushes the demand curve for v_w to the left. Second, the change in q also affects the productivity of, and demand for, v_w. The direction of this effect depends on the specific characteristics of the production function and on the elasticity of demand for y_w. As the price of v_w changes, this shifts MC_w in the same direction. There are also two general equilibrium supply functions, S^{*0} and S^{*1}, in panel B of Figure 8.6, one for each level of q. They describe the supply adjustments to changing factor prices, and they show the minimum supply price for y_w for each level of q, taking account of changes in factor prices as the derived demand for v_w shifts. The net result is an equilibrium in this market at the intersection of D_w^* and S^{*1}, as shown in panel B of Figure 8.6.

Three groups of economic agents are affected by the change in q, so there are three components to the measure of welfare gain. The first group, the consumers of y_r, receives an increase in consumer surplus, shown as $w + v$ in panel C of Figure 8.6. The second group, the owners of v_w, experiences a change in their rents because of the change in its price. They may either gain or lose depending upon how the demand for v_w changes; in Figure 8.6, panel A, they lose the area $c + d$. The third group consists of the firms in the wholesale and retail industries that experience changes in their quasi-rents. The change in quasi-rents in the retail industry can be measured directly in the retail market by the area $u - w$. As discussed above, if v_w is a necessary input, then the change in quasi-rents to the wholesale industry can be measured in the factor market by the area $c - a$. Considering all of the changes, and netting out transfers between buyers and sellers resulting from price changes, we have

$$\begin{aligned} W_q &= \Delta PS_v + \Delta R_w + \Delta R_r + \Delta CS_r \\ &= -(c+d) + (c-a) + (u-w) + (w+v) \\ &= -(d+a) + (u+v). \end{aligned} \tag{8.41}$$

Following Just, Hueth, and Schmitz (2004), it can be shown that the welfare gain can also be measured in the directly affected wholesale market, provided that the measurements are based on the general equilibrium supply and demand curves. These curves reflect the adjustments of prices in linked markets that transfer gains and losses among buyers and sellers. Specifically,

$$W_q = \Delta CS_w^* + \Delta R_w^*, \tag{8.42}$$

where $C_w^* \left(= \Delta R_r + \Delta CS_r\right)$ and $R_w^* \left(= \Delta R_w + \Delta PS_v\right)$ are the areas behind the general equilibrium demand and supply curves respectively. So from equation (8.42),

$$W_q = f + g + j - f = g + j. \tag{8.43}$$

Thus, if general equilibrium demand and supply functions for the directly impacted market can be estimated, it may be more convenient to measure welfare gains using this expression than to attempt to measure separately the changes in quasi-rents and surpluses in all of the remaining vertically linked markets.

Now consider the case of multiproduct firms that are part of a chain of vertically linked markets. As just shown, the welfare gain can be measured in the directly affected market. However, the welfare measure must be modified to take account of the multiproduct characteristic of the producing industry. Specifically, if there is any jointness in production, it is not correct simply to add the areas $g + j$ across all products. Rather, the correct measure is the sum of the consumer surplus changes as measured by areas like $f + g$, and the change in producing firms' quasi-rents as measured by $j - f$, in the market for one essential output.

Monopoly Markets

The welfare measures derived in the preceding sections come from models in which all markets are perfectly competitive. In this section the effects on welfare measurement when there is monopoly power in the output market is considered.

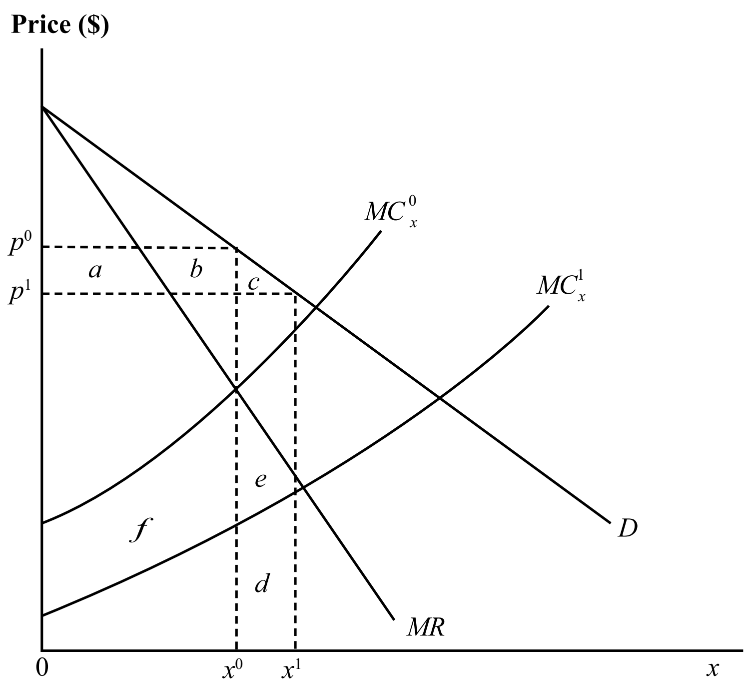

Figure 8.7 The welfare change in a monopoly market

The correct welfare measure in the case of monopoly can be derived graphically in a straightforward manner, but measurement is another question. Figure 8.7 shows the demand, marginal revenue, and marginal cost curves for a monopolist. If an increase in q shifts the marginal cost curve outward, the welfare measure is the sum of the increases in consumer surplus and monopoly quasi-rents:

$$W_q = \Delta CS + \Delta R, \tag{8.44}$$

where

$$\Delta CS = a + b + c \tag{8.45}$$

$$\Delta R = \Delta TR - \Delta C = d + e - (d - f) = e + f. \tag{8.46}$$

Thus,

$$W_q = a + b + c + e + f. \tag{8.47}$$

As in the case of perfect competition, consumers always gain because of the price reduction, but in this case the monopolist also always gains from an increase in q. The monopolist's profit is

$$I = p(x) \cdot x - C(x, q) \tag{8.48}$$

$$\frac{\partial I}{\partial q} = \left(p + x\frac{dp}{dx} - \frac{\partial C}{\partial x} \right) \frac{dx}{dq} - \frac{\partial C}{\partial q}. \tag{8.49}$$

The first-order condition for profit maximization requires that the term in parentheses be equal to zero. So,

$$\frac{\partial I}{\partial q} = -\frac{\partial C}{\partial q} > 0. \tag{8.50}$$

The problem posed for measurement is that the marginal cost curve of the monopolist cannot be observed from market data on optimally chosen prices and quantities. One possible approach to deriving marginal cost functions is to construct models of the firm based on engineering or technological data (see Russell and Vaughan 1976 for an example).

Valuing Changes in the Productivity of Natural Resource Systems

The productivity of commercially exploited natural biological systems can depend on such things as the flow of nutrients into the system, the population of a predator species, climatic variables such as precipitation and temperature, or the level of a pollutant. Any of these factors might be subject to human manipulation, so it would be useful to have an economic framework for evaluating the welfare consequences of policies to change these things. Since the outputs of such resource systems are traded in markets, the framework presented in this chapter is applicable. The welfare consequences come in the form of changes in producers' and consumers' surpluses. However, there are some special features of natural resource systems that sometimes need to be taken into account.

This section provides a brief description of how an environmental quality variable can be introduced into the standard economic models of two types of natural resource systems—the commercial forest and the commercial fishery. In the case of the forest, an investment in environmental quality may alter the optimum time of harvest, so the intertemporal features of the forest optimization model must be examined. Moreover, an environmental quality variable might affect the economic value of a nonmarket output such as recreation. Thus, the interaction between the nonmarket and market outputs must be analyzed in a multiple-use framework.

In the case of the fishery, there are two commonly used bio-economic models in the literature, the Schaefer–Gordon growth model, and the Beverton–Holt stock-recruitment model. Environmental parameters can be easily introduced into both models, but the economic implications of a change in environmental quality are more transparent in the case of the Schaefer–Gordon model, as will be shown. In addition, the economic value of an environmental quality change will depend upon the institutions for ownership or management of the fishery. This is demonstrated by a comparison of the economic value of an environmental change under the alternatives of optimum management and open-access exploitation.

Commercial Forests and the Role of Time

The models described in the preceding sections of this chapter were timeless in the sense that environmental changes and changes in prices and quantities were contemporaneous. There was no need to model explicitly any intertemporal effects. Where the linkages between environmental changes and market changes are not contemporaneous, time must be built into the model explicitly. An example is the case of the effects of changes in air quality on commercial forest productivity.

One of the things that makes the case of the effects of environmental change on commercial forests interesting from an economic perspective is that a change affecting a stand of young trees today will not have an effect on marketed outputs for perhaps 40 years, when the trees are harvested and sold. If there were an active market in forestland, any increase in the growth rate of young trees now would have an immediate impact on the market price of land where there are standing, growing trees. Even so, any consumer surplus benefit from lower prices of forest products would not be realized until the increased harvests actually took place. In what follows, the product price effects on consumers are ignored and only rents accruing to forests and owners are considered.

Suppose that the trees on a plot of land grow in net value over time according to

$$G_t = G(t,q), \tag{8.51}$$

where G_t is the stumpage value—that is, the market price of the harvested volume at age t, net of harvest costs and transportation to the mill. This is a very general formulation that allows for growth in harvestable volume as well as for changes in the price per unit of volume because of changes in quality. We also assume that replanting and management costs are zero.

Assuming no economically relevant alternative uses for a unit of land, the landowner's economic problem is to choose a sequence of harvesting dates to maximize the present value (V) of the stream of net receipts at each harvest—that is,

$$\begin{aligned}\underset{t}{\text{Max}} : V &= \sum_{h=1}^{\infty} e^{-h \cdot r \cdot t} G(t, q) \\ &= \frac{G(t, q) e^{-r \cdot t}}{1 - e^{-r \cdot t}}\end{aligned} \tag{8.52}$$

where t is the age of the stand at the time of harvest, r is the interest rate, and h indexes the generation of the stand. For simplicity, assume that there are no costs for planting, thinning, and other management activities during the rotation. For a more complete treatment of the forestry optimization problem, see Samuelson (1976), Hyde (1980), and Bowes and Krutilla (1989).

The stand should be harvested at the age that satisfies

$$\frac{\partial V}{\partial t} = \frac{\dfrac{\partial G}{\partial t} - \left[r \cdot G(\cdot)\right] - \dfrac{r \cdot G(\cdot) e^{-r \cdot t}}{1 - e^{-r \cdot t}}}{1 - e^{-r \cdot t}} = 0 \,, \tag{8.53}$$

or

$$\frac{\partial G}{\partial t} = \left[r \cdot G(\cdot)\right] + \frac{r \cdot G(\cdot) e^{-r \cdot t}}{1 - e^{-r \cdot t}} \,. \tag{8.54}$$

The stand should be allowed to grow as long as the marginal gain in value through growth $\partial G/\partial t$ exceeds the interest forgone by not realizing G through harvest ($r \cdot G$) plus the opportunity cost of postponing the stream of returns from future rotations. Harvest should occur when the marginal gain from waiting just equals the marginal opportunity cost.

Two questions of interest are the effects of changes in q on V and on the optimum rotation length. From equation (8.52):

$$\frac{\partial V}{\partial q} = \frac{\partial G}{\partial q} \cdot \frac{e^{-r \cdot t}}{1 - e^{-r \cdot t}} \,. \tag{8.55}$$

The increase in q increases G at the time of the next harvest, and the second term in this expression gives the present value of the stream of these increases over the infinite future. So, if the growth function were known, including the effect of q, then calculating the effect of changes in the steady-state level of q on the stream of rents would be straightforward. However, q often varies over time, because, for example, of trends in pollution associated with economic growth: therefore, such calculations would require more detailed knowledge of the growth function than is currently available. Answering the second question appears to be more difficult. If t^* is the solution to equation (8.54), it appears to be impossible to determine unambiguously the sign of $\partial t^*/\partial q$.

Environmental Quality and Multiple-Use Management of Forests

Suppose that in addition to the periodic harvest of marketable products, the forest unit provides a flow of nonmarket services such as recreation that depends on the age of the forest and on q. Models for measuring the value of recreational resources are discussed in Chapter 9.

Let the value of the nonmarket service be given by $R(t,q)$. Over one harvest cycle, the present value of the nonmarket service flow is

$$V_n = \int_0^{t^*} R(t,q)dt. \tag{8.56}$$

The objective function becomes:

$$\underset{t}{\text{Max}} : V = \sum_{h=1}^{\infty} e^{-h \cdot r \cdot t}\left[G(t,\ q) + \int_0^t R(t,\ q)\,dt\right]. \tag{8.57}$$

As Hartman (1976) first showed, including the value from the standing forest in the objective function will either increase the optimum age at harvest (t^*) or make harvesting uneconomic entirely. This is because the standing forest adds a second term, the marginal benefit of delaying the harvest, to the left-hand side of equation (8.54). The optimum harvest (if it exists) occurs when

$$\frac{\partial G}{\partial t} + R(\cdot) = r \cdot G(\cdot) + \frac{r \cdot G(\cdot)e^{-r \cdot t}}{1 - e^{-r \cdot t}}. \tag{8.58}$$

An increase in q can be valuable both because of its increase in the present value of the flow of harvestable product and because of its effect on the value of the flow of services from the standing forest. For further discussion of the Hartman result and multiple-use forest management in general, see Bowes and Krutilla (1989), Swallow and Wear (1993), and Swallow, Talukdar, and Wear (1997).

One benefit of modeling exercises of this sort is to provide guidance for the natural science research that is required to support future economic analyses. As these models suggest, it is important to know how changes in q affect the whole time pattern of the growth of trees and the flow of nonmarket services over the life of the forest.

Environmental Quality and Commercial Fisheries

The economic analysis of fisheries rests on a foundation of a biological model of the growth and mortality of a species. Two alternative biological models have dominated the literature on the economics of fisheries. The first is the so-called Schaefer–Gordon model, which makes the growth rate of the aggregate biomass of the species at any point in time a function of the current level of biomass. The alternative Beverton–Holt model explicitly describes both the number of fish in each age cohort and their weight or size. The Beverton–Holt model may be more realistic from a biological perspective; but it is also more complex from an economic perspective.

The Schaefer–Gordon Model

In this model the relationship between the growth of the aggregate stock, g, and the aggregate size of the stock, z, takes the following form:

$$g = b \cdot z\left(1 - \frac{z}{k}\right) = b \cdot z - \frac{b}{k} \cdot z^2, \tag{8.59}$$

where b is the intrinsic growth rate and k is the carrying capacity of the environment. As this shows, when the stock has grown to the carrying capacity $(z = k)$, the growth rate is zero. This quadratic growth function results in a logistic time path for the stock in the absence of harvest, with z approaching the carrying capacity asymptotically.

Environmental parameters have been incorporated into the Schaefer–Gordon growth model in several empirical studies of fisheries. Either the intrinsic growth rate or the carrying capacity (or both) can be made a function of an environmental quality parameter, q. For example, in his study of the North American lobster fishery, Bell (1972) included seawater temperature as one variable helping to explain the annual harvest. Lynne, Conroy, and Prochaska (1981) examined the effects of changes in the acreage of marine wetlands on annual harvests of blue crabs in the Florida Gulf Coast fishery. For other examples, see Bell (1989), Swallow (1994), and Barbier and Strand (1998).

The Schaefer–Gordon model can be incorporated into an economic model to determine the optimum levels of harvest and stock. This model can also be used to trace the effects of changes in environmental parameters on harvest, stock, and economic welfare. If an environmental quality parameter can be increased by public policy at some cost, optimum levels of investment in environmental quality can also be determined.

The standard approach is to specify a production function that makes the fishery industry's annual harvest, h, a function of economic inputs and the stock of fish to be caught, z. For simplicity, the economic inputs are aggregated into a measure of effort, e. A unit of effort can be called a "boat" and can be interpreted as an optimal combination of labor and capital. Assume that each unit of effort has a cost of p_e. For any chosen level of h, the production function can be solved for the required level of effort, given the size of the stock. Since the size of the stock at any point and time is the net result of past growth (which depends in part on q) and past harvest, the cost function can be written as

$$C = p_e \cdot e = C\left(h, z, q\right). \tag{8.60}$$

Managing the fishery for economic objectives is a dynamic problem, since current harvests and costs depend on past harvest decisions and how they have affected the size of the stock. The economic objective is to maximize the present value of the net economic return from the fishery over time, subject to the biological constraint imposed by the growth function. If $p_h\left(h\right)$ is the inverse demand function for fish, the objective function is

$$\underset{h_t}{\text{Max}}: \int_0^\infty \left[\int_0^h p_h\left(h_t\right) dh - C\left(h_t, z_t, q\right)\right] e^{-r \cdot t} \cdot dt, \tag{8.61}$$

subject to

$$\frac{dz}{dt} = g\left(z_t, q\right) - h_t. \tag{8.62}$$

For any given initial conditions, this problem can be solved for the sequence of harvests that maximizes the objective function. The economic dimensions of the problem are easiest to see if we focus attention on the conditions for the long-run, steady-state equilibrium where growth and harvest are equal and the stock is of constant size. The first-order conditions for this optimum are

$$p_h = \frac{\partial C}{\partial h} + \rho \tag{8.63}$$

and

$$r = \frac{\partial g}{\partial z} - \frac{1}{\rho} \cdot \frac{\partial C}{\partial z}, \tag{8.64}$$

where r is the interest rate and ρ is the shadow value of the stock, reflecting its effect on the rate of growth and the cost of harvest. See, for example, Fisher (1981) or Clark (1976, 1985). Equation (8.63) says that the price of fish must equal the full marginal cost of their harvest, which includes the opportunity cost of decreasing the stock, ρ. Equation (8.64) defines the optimum intertemporal tradeoff. The interest rate is the opportunity cost of forgoing \$1 worth of harvest now, and must equal the benefit of forgoing harvest, which has two components. The first component is the contribution the additional stock makes to future growth. The second reflects the contribution that the extra stock makes to lowering harvest costs.

Valuation of Changes in q in the Schaefer–Gordon Model

An increase in q shifts the cost function in equation (8.61) and changes the growth function of (8.62). This leads to new solution values for h and z in every period. Assuming optimal management and the satisfaction of (8.63) and (8.64), the welfare value in each period is the increase in the net value term in brackets in equation (8.61)—that is, the area between the marginal cost curves bounded by the demand curve. The stream of changes in that value must be discounted to its present value to obtain the total welfare value of the change in q.

Taking wetland acreage as a measure of q, Ellis and Fisher (1987) drew on earlier work by Lynne, Conroy, and Prochaska (1981) to calculate the net welfare value of increases in wetlands in the blue crab fishery of the Gulf Coast of Florida. They abstracted from the dynamic, intertemporal dimension of the fishery management problem by focusing only on a single-period optimum and by assuming that the cost of harvest was independent of this stock. This is equivalent to asserting that $\rho = 0$, so that an optimum is defined by the price of harvest equaling its marginal cost.

However, the economic value of a change in environmental quality depends not only on the economic and biological parameters of the model, but also on the institutional arrangements for ownership or public management of the resource. The preceding discussion (and the Ellis–Fisher analysis) is based implicitly on the assumption of either private ownership with perfect competition or public regulation to achieve the economic optimum. However, most fisheries resources are characterized by absence of private ownership and more or less open access to the resource, and where there is public regulation, it is seldom designed to achieve an economically efficient outcome. Consequently, the net economic value of the resource will be lower; and the economic value of changes in q will also be affected by the ownership and management arrangements.

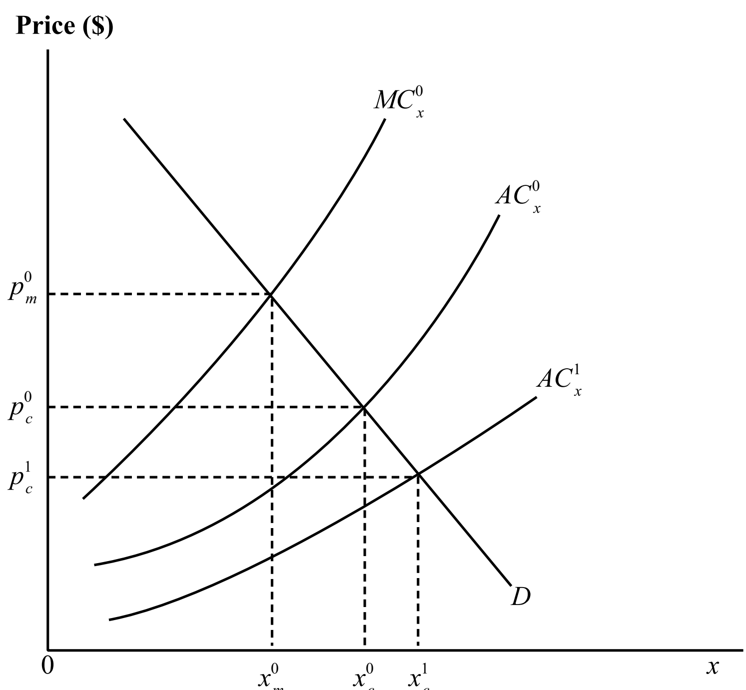

Figure 8.8 Welfare measurement for open-access resources: the case of a fishery

A simple static model can be used to show the effect of open access on the welfare value of a change in the productivity of the fisheries resource such as might result from an improvement in water quality (see also McConnell and Strand 1989). It is well known that under open access without regulation, competition drives rent to zero (Gordon 1954; Scott 1955). Let MC_x^0 in Figure 8.8 be the marginal cost of harvesting fish at the original level of q. If the fishery is privately owned, or if it is optimally managed, the output and price will be given by the intersection of the marginal cost of harvest function and the demand curve at p_m^0 and x_m^0. An improvement in q would shift the marginal cost of harvest function outward to the right, resulting in a higher quantity and a lower equilibrium price. The new marginal cost curve is not shown in Figure 8.8. The welfare measure would be the same as that presented above.

If there is open access to the fishery, the condition for equilibrium is each fisherman earning zero profits. This requires that price be equal to the average cost of harvesting fish. Compared to the efficient price and output under private ownership, more fishermen enter in pursuit of profit. The increase in fish caught decreases the stock, raising costs or decreasing the price of fish, or both. Entry continues until these forces eliminate the incentive for entry. At the initial level of q, the average cost of harvesting fish is AC_x^0, as shown in Figure 8.8. The open-access equilibrium is at p_c^0 and x_c^0.

An improvement in q also shifts the average cost curve outward, resulting in a lower price and a higher quantity; however, since price equals average cost both before and after the change in environmental quality, there is no change in producer surplus. The benefit consists entirely of the increase in consumer surplus associated with the price decrease. The more elastic is the demand curve for fish, the smaller is the welfare gain associated with the environmental improvement. Freeman (1991) presented some illustrative calculations of the magnitude of these effects based on the Ellis–Fisher data. In the limit, if this fishery is small relative to the market and the demand for fish is perfectly elastic, there is no welfare gain under open access. The physical improvement in productivity brought about by the higher water quality is entirely dissipated by the uneconomic competition of fishermen for the potential increase in rents. However, with inelastic demand the welfare gain from an increase in q is slightly higher under open access than under optimal regulation.

Empirical Examples

Problems associated with nutrient over-enrichment in many freshwater and coastal systems have been documented around the world and are increasingly receiving policy attention (Rabotyagov et al. 2014). A myriad of ecosystem services are affected by low oxygen levels (hypoxic conditions) and "dead zones" that result from excess nutrient flowing into these systems. One potentially important welfare consequences of this water quality problem is damage to commercial fisheries and the methods described here are directly relevant.

Several papers have studied the welfare losses associated with eutrophication from excess nutrients flowing into the Pamlico Sound estuary, which is located in North Carolina. Huang, Smith, and Craig (2010) employed a spatial bioeconomic model to relate hypoxic conditions to brown shrimp stock and found that hypoxic conditions accounted for about a 13 percent decline in harvest in the early 2000s. Using the bioeconomic model developed by Huang, Smith, and Craig, Huang et al. (2012) studied the welfare effects of hypoxic conditions considering both supply and demand effects. Because the North Carolina shrimp industry is small in the world market, and therefore demand is elastic, the authors found that there is almost no loss in consumer surplus from hypoxic conditions in the region. While there are losses in producer surplus associated with hypoxia, they are relatively small when supply elasticity is accounted for.

Another fishery where welfare effects from low oxygen levels have been addressed is the blue crab fishery, also in North Carolina. Smith (2007) developed a spatial bioeconomic model to allow for two locations (patches) where hypoxic conditions occur in one but not the other. The model was used to estimate changes in producer and consumer surplus from a 30 percent decrease in nitrogen entering the fishery (which is the primary driver of hypoxic conditions) under both open access and an effort constrained scenario. Smith found that the benefits of addressing the water quality problem depend on the management regime with total benefits ranging from $1 million to $7 million annually.

The Beverton–Holt Model

The Schaefer–Gordon model has been criticized as being too aggregated, and not being biologically realistic for many species. Also, one of its key predictions is not supported by observation for some species. The prediction is that if effort increases, eventually the stock must decrease, leading to a decline in growth and harvest and eventually to the biological and economic collapse of the fishery. Yet for several fisheries, plots of effort and harvest over time show increasing effort associated with almost no change in harvest. Townsend (1986) showed this to be the case for the North American lobster fishery originally analyzed by Bell (1972), and McClelland (1991) has also shown this to be the case for several of the Florida Gulf Coast fisheries analyzed by Bell (1989). This observed pattern of harvest and effort is entirely consistent with the Beverton–Holt biological model of the fishery.

The Beverton–Holt model combines a model of the growth of individual members of a species and the number of individuals in a single expression. The first element of the model is the recruitment of k individuals of the same age into a cohort at time $t = 0$. In the basic model, k is considered to be exogenous; but clearly, it could be a function of environmental factors. Consider, for example, the oyster. If recruitment is defined as the attachment of larva to solid surfaces where shell development begins, the limiting factor in this stage of development may be the quantity of suitable surface available for attachment, not the number of eggs released and fertilized in the spawning stage.

The second element of the model is an expression for the population of the cohort at any point in time. Over the lifetime of the species, the total population changes because of natural mortality, *m*, and the harvest of fish, *f*. The latter is sometimes referred to as fishing mortality. Both of these forms of mortality can vary over time. Thus, the population of the cohort at time *t* is given by

$$n_t = k \cdot e^{-(f_t + m_t)^t}. \tag{8.65}$$

The third element of the model is an expression giving the weight of a typical member of the cohort at any point in time. Each fish grows over time, so its weight is given by

$$w_t = w_\infty \left[1 - e^{-\alpha(t)}\right]^\beta. \tag{8.66}$$

The total biomass of the cohort at time *t* is the product of equations (8.65) and (8.66). The last element of the model shows the impact of fishing effort on the cohort. The total weight of fish harvested at any point in time is

$$h_t = w_t \cdot n_t \cdot f_t, \tag{8.67}$$

where f_t depends, at least in part, on the level of effort.

Cost and revenue functions can be specified and the optimum time path of the exploitation of a cohort can be derived. For examples see Clark (1976, 1985). However, since the number of recruits in one cohort is largely independent of the fishing mortality experienced by earlier cohorts, the intertemporal properties of the Beverton–Holt model are not nearly as interesting as those of the Schaefer–Gordon model.

What is of interest for our purposes is the number of places at which environmental quality parameters can enter into the biological model. For example, environmental variables can affect the number of fish available for harvest at a point in time, their individual weights, or both. Recruitment and natural mortality may both depend upon environmental quality. The individual growth rate, α, and the upper limit on size are also likely to depend on environmental quality. An interesting and potentially fruitful area for future research is the role of environmental variables in the Beverton–Holt model and the development of expressions for the economic value of these changes.

Summary

When an environmental quality variable affects the production costs of firms, the welfare value of the change is measured by changes in the surpluses of producers and consumers. Measuring these values requires a model of the market for the output. In the simplest case of single-product firms in a competitive industry, developing such a model and deriving estimates is straightforward, given knowledge of how *q* affects the production or cost functions, or both. In this chapter, the simple model has been extended to deal with several kinds of complications, including the cases of multiproduct firms, and vertically linked markets. Monopoly power models for optimum management of biological resources—such as forests and fisheries—have also been adapted for the purposes of welfare measurement by considering how environmental quality parameters can be incorporated into these models.

References

Adams, Richard M., and Bruce A. McCarl. 1985. "Assessing the Benefits of Alternative Ozone Standards on Agriculture: The Role of Response Information." *Journal of Environmental Economics and Management* 12(3): 264–276.

Adams, Richard M., Scott A. Hamilton, and Bruce A. McCarl. 1984. *The Economic Effects of Ozone on Agriculture*. Corvallis: U.S. Environmental Protection Agency.

Adams, Richard M., Thomas D. Crocker, and Richard W. Katz. 1984. "Assessing the Adequacy of Natural Science Information: A Bayesian Approach." *Review of Economics and Statistics* 66(4): 568–575.

Anderson, James E. 1976. "The Social Cost of Input Distortions: A Comment and a Generalization." *American Economic Review* 66(1): 235–238.

Auffhammer, Maximillian, V. Ramanathan, and Jeffrey R. Vincent. 2006. "Integrated Model Shows That Atmospheric Brown Clouds and Greenhouse Gases Have Reduced Rice Harvests in India." *Proceedings of the National Academy of Sciences* 103(52): 19668–19672.

Barbier, Edward B., and Ivar Strand. 1998. "Valuing Mangrove-Fishery Linkages: A Case Study of Campeche, Mexico." *Environmental and Resource Economics* 12(2): 1151–1166.

Bell, Frederick W. 1972. "Technological Externalities and Common Property Resources: An Empirical Study of the U.S. Northern Lobster Fishery." *Journal of Political Economy* 80(1): 148–158.

Bell, Frederick W. 1989. *Application of Wetland Evaluation Theory to Florida Fisheries*. Tallahassee: Florida Sea Grant College.

Bockstael, Nancy E., and Kenneth E. McConnell. 2007. *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models*. Dordrecht: Springer.

Bowes, Michael D., and John V. Krutilla. 1989. *Multiple-Use Management: The Economics of Public Forestlands*. Washington, D.C.: Resources for the Future.

Clark, Colin W. 1976. *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*. New York: John Wiley.

Clark, Colin W. 1985. *Bioeconomic Modeling and Fisheries Management*. New York: John Wiley.

Ellis, Gregory M., and Anthony C. Fisher. 1987. "Valuing the Environment as Input." *Journal of Environmental Management* 25(2): 149–156.

Fisher, Anthony C. 1981. *Resource and Environmental Economics*. New York: Cambridge University Press.

Freeman, A. Myrick, III. 1982. *Air and Water Pollution Control: A Benefit-Cost Assessment*. New York: John Wiley.

Freeman, A. Myrick, III. 1991. "Valuing Environmental Resources Under Alternative Management Regimes." *Ecological Economics* 3(3): 247–256.

Freeman, A. Myrick, III, and Winston Harrington. 1990. "Measuring Welfare Values of Productivity Changes." *Southern Economics Journal* 56(4): 892–904.

Garcia, Phillip, Bruce L. Dixon, James W. Mjelde, and Richard M. Adams. 1986. "Measuring the Benefits of Environmental Change Using a Duality Approach: The

Case of Ozone and Illinois Cash Grain Farms." *Journal of Environmental Economics and Management* 13(1): 69–80.

Gordon, H. Scott. 1954. "The Economic Theory of a Common Property Resource: Fisheries." *Journal of Political Economy* 62(2): 124–142.

Hartman, Richard C. 1976. "The Harvesting Decision When a Standing Forest Has Value." *Economic Inquiry* 14(1): 52–58.

Heck, Walter W., Richard M. Adams, William W. Cure, Allen S. Heagle, Howard E. Heggestad, Robert J. Kohut, Lance W. Kress, John O. Rawlings, and O. Clifton Taylor. 1983. "A Reassessment of Crop Loss from Ozone." *Environmental Science and Technology* 17(12): 572A–581A.

Huang, Ling, Martin D. Smith, and J. Kevin Craig. 2010. "Quantifying the Economic Effects of Hypoxia on a Fishery for Brown Shrimp *Farfantepenaeus aztecus*." *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* 2: 232–248.

Huang, Ling, Lauren A.B. Nichols, J. Kevin Craig, and Martin D. Smith. 2012. "Measuring Welfare Losses from Hypoxia: The Case of North Carolina Brown Shrimp." *Marine Resource Economics* 27(1): 3–23.

Hyde, William F. 1980. *Timber Supply, Laced Allocation, and Economic Efficiency*. Washington, D.C.: Resources for the Future.

Just, Richard E., and Darrell L. Hueth. 1979. "Multimarket Welfare Measurement." *American Economic Review* 69(5): 947–954.

Just, Richard E., Darrell L. Hueth, and Andrew Schmitz. 2004. *The Welfare Economics of Public Policy: A Practical Approach to Project and Policy Evaluation.* Northampton: Edward Elgar.

Kopp, Raymond J., William J. Vaughan, Michael Hazilla, and Richard T. Carson. 1985. "Implications of Environmental Policy for U.S. Agriculture: The Case of Ambient Ozone Standards." *Journal of Environmental Management* 20(4): 321–331.

Lynne, Gary D., Patricia Conroy, and Frederick J. Prochaska. 1981. "Economic Valuation of Marsh Areas for Marine Production Processes." *Journal of Environmental Economics and Management* 8(2): 175–186.

Mathtech. 1982. *Benefits Analysis of Alternative Secondary National Ambient Air Quality Standards for Sulfur Dioxide and Total Suspended Particulates, Final Analysis*. Princeton: Mathtech.

McClelland, John D. 1991. "On Selecting Models for Fishery Management." An honors paper for the Department of Economics, Bowdoin College, Brunswick, ME.

McConnell, Kenneth E., and Nancy E. Bockstael. 2005. "Valuing the Environment as a Factor of Production." In *Handbook of Environmental Economics*, edited by Karl-Göran Mäler and Jeffery R. Vincent. Amsterdam: Elsevier.

McConnell, Kenneth E., and Ivar E. Strand. 1989. "Benefits for Commercial Fisheries When Demand and Supply Depend on Water Quality." *Journal of Environmental Economics and Management* 17(3): 284–292.

Mjelde, James W., Richard M. Adams, Bruce L. Dixon, and Phillip Garcia. 1984. "Using Farmers' Actions to Measure Crop Loss Due to Air Pollution." *Journal of the Air Pollution Control Association* 34(4): 360–364.

Rabotyagov, Sergey S., Catherine L. Kling, Philip W. Gassman, Nancy N. Rabalais, and Eugene E. Turner. 2014. "Causes, Impacts, Policy Challenges, and a Model of the Gulf of Mexico Hypoxic Zone." *Review of Environmental Economics and Policy* 8(1): 58–79.

Russell, Clifford S., and William J. Vaughan. 1976. *Steel Production: Processes, Products, and Residuals*. Baltimore: Johns Hopkins University Press for Resources for the Future.

Samuelson, Paul A. 1976. "Economics of Forestry in an Evolving Society." *Economic Inquiry* 14(4): 466–492.

Schmalensee, Richard. 1976. "Another Look at the Social Valuation of Input Price Changes." *American Economic Review* 66(1): 239–243.

Schlenker, Wolfram, and Michael J. Roberts. 2009. "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change." *Proceedings of the National Academy of Sciences* 106(37): 15594–15598.

Scott, Anthony. 1955. "The Fishery: The Objectives of Sole Ownership." *Journal of Political Economy* 63(2): 116–124.

Smith, Martin D. 2007. "Generating Value in Habitat-Dependent Fisheries: The Importance of Fishery Management Institutions." *Land Economics* 83(1): 59–73.

Swallow, Stephen K. 1994. "Renewable and Nonrenewable Resource Theory Applied to Coastal Agriculture, Forest, Wetland, and Fishery Linkages." *Marine Resource Economics* 9(4): 291–310.

Swallow, Stephen K., and David N. Wear. 1993. "Spatial Interactions in Multiple-Use Forestry and Substitution and Wealth Effects for the Single Stand." *Journal of Environmental Economics and Management* 25(2): 103–120.

Swallow, Stephen K., Piyali Talukdar, and David N. Wear. 1997. "Spatial and Temporal Specialization in Forest Ecosystem Management under Sole Ownership." *American Journal of Agricultural Economics* 79(2): 311–326.

Townsend, Ralph E. 1986. "A Critique of Models of the American Lobster Fishery." *Journal of Environmental Economics and Management* 13(3): 277–291.

Chapter 9

Recreation Demand

Many natural resource systems such as lakes, rivers and streams, estuaries, and forests are used extensively for various kinds of recreation activities, including fishing, hunting, boating, hiking, and camping. As places to conduct such activities, natural resource systems provide valuable services. From an economic perspective, these services have two important features. First, the economic value of these services depends upon the characteristics of the natural resource system. The characteristics determining value can be affected by air and water pollution and by resource management decisions about such things as the rates of harvest of timber and fish, the extraction of minerals and petroleum, and the allocation of water flows between diversionary uses and various in-stream uses. Knowledge of the values of these services may be important for a variety of resource management decisions.

The second important feature is that access to the resource for recreation is usually not allocated through markets. Rather, access is typically open to all comers at a zero or nominal entrance fee that bears no relationship to the cost of providing access. Moreover, there is little or no variation in these access prices over time or across sites to provide data for econometric estimation of demand functions.

At first blush, the lack of direct access fees would seem to suggest little hope for recovering the underlying demand for recreation and the implied value of the services recreational sites provide. Fortunately, Hotelling (1947), in an unpublished letter to the U.S. Department of the Interior, suggested otherwise, noting that each individual visit to a recreation site involves an implicit transaction in which the cost of traveling to the site is incurred in return for access to the site. These travel costs include both explicit costs (e.g., in the form of gasoline, tolls, etc.) and the implicit opportunity costs of time (both traveling to the site and time on-site). Different individuals will face different travel costs to any site, and one individual will face different travel costs for the different sites he or she might visit. The responses of people to this variation in the implicit prices of visits are the basis for estimating the demand for recreation and the values of recreation sites and changes in site quality.

Travel cost models of recreation demand (or simply "recreation demand" models) have evolved considerably in the 65 years since Hotelling's original insight.

Early models relied on zonal data, relating total trips per capita in a region to an average measure of travel cost. As micro-level data became available, analysts became more concerned with the microeconomic foundations of recreation demand models (e.g., the role of substitute sites and the opportunity cost of time), as well as with econometric issues (e.g., the count nature of trip demand and the prevalence of corner solutions). Indeed, as Phaneuf and Smith (2005, 673) note, "[t]oday, economic analyses of recreation choices are among the most advanced examples of the microeconometric model of consumer behavior in economics." The resulting recreation demand models have come to play a central role in nonmarket valuation, particularly in terms of informing regulators in setting environmental policy, ex post cost benefit analysis, and in natural resource damage assessment cases.

In the first section of this chapter, a generic recreation demand model is developed to illustrate the key features of this valuation technique. The second section then describes the nature of the data typically available to analysts (i.e., seasonal trip counts to one or more recreational sites). The features and limitations of such data, in many ways, drive the various recreation demand models that have emerged in the literature. The third section describes single-site models. While such models are relatively rare these days, the models provide a useful starting point, highlighting the importance of controlling for substitute sites and the challenges associated with modeling count data with frequent corner solutions. The fourth section then provides the more general approaches for characterizing systems of recreation demand equations. The fifth section highlights some of the prominent issues in modeling recreation demand, and the sixth section provides summary and conclusions. Space constraints, as always, make it difficult to go into detail in some areas of recreation demand modeling, especially in terms of the myriad of econometric issues. Haab and McConnell (2002, chs. 6–8) provided an excellent treatment of econometric issues, while Phaneuf and Smith (2005) provided additional perspective on the evolution and state of the art in recreation demand modeling.

The Generic Recreation Demand Model

A wide range of modeling frameworks has emerged over the past 40 years seeking to characterize recreation demand, and the implied demand for the associated environmental amenities. In some instances, the starting point is the specification of Marshallian trip demand functions, while others begin with a representation of individual preferences for both trips and site amenities in the form of a direct or indirect utility function. Historically, approaches beginning with a demand function specification focused almost exclusively on the quantity aspect of the consumer decision (how many trips to take over a given time period), while those beginning with a utility function focused on the discrete site selection aspect (which recreation site provides the best combination of price and site characteristics). See Freeman (2003) for an exposition of this dichotomy.

However, over time, researchers have developed models that combine information from both margins (which site to visit and how often to take trips) to infer the value of recreation sites and their attributes from behavior. Thus, the dichotomy between quantity and site selection model approaches is less useful than it once was—at the heart of all of the models is the fundamental notion that individuals choose where and how often to recreate based on the attributes of the available sites and the cost of accessing the site. To the extent that individuals travel great distances to visit a site, incurring costs in the form of both time and money, they reveal information regarding the value they place in the amenities of the chosen site. A simple recreation demand model, adapted from McConnell (1985), helps to illustrate the basic elements of, and issues associated with, recreation demand.

A Simple Conceptual Model

Assume, for now, that there is only one site available and that all visits have the same duration. While relatively few empirical studies currently employ single-site models, this presentation provides an important foundation for more realistic multiple-site approaches. Modeling of the choice of the length of a visit to a site and of the choice of which sites to visit when there are alternatives will be taken up in later sections of this chapter. Furthermore, assume that the individual's utility depends on the total time spent at the site, the quality of the site, and the quantity of a numeraire. With the duration of a visit fixed for simplicity, the time on site can be represented by the number of visits. The individual solves the following utility maximization problem:

$$\max_{z,x} u(z,x,q), \tag{9.1}$$

subject to the twin constraints of monetary and time budgets:

$$M + w \cdot t_w = z + p_x \cdot x \tag{9.2}$$

and

$$t^* = t_w + (t_1 + t_2)x \ , \tag{9.3}$$

where

z = the quantity of the numeraire whose price is one,
x = number of visits to the recreation site,
q = environmental quality at the site,
M = exogenous income,
w = wage rate,
p_x = monetary cost of a trip,
t^* = total discretionary time,
t_w = hours worked,
t_1 = round-trip travel time, and
t_2 = time spent on site.

Assume that x and q are complements in the utility function. As discussed in Chapter 4, this means that the number of visits will be an increasing function of the site's environmental quality. The time constraint reflects the fact that both travel to the site and time spent on the site take time away from other activities. Thus, there is an opportunity cost to the time spent in the recreation activity. Assume also that the individual is free to choose the amount of time spent at work and that work does not convey utility (or disutility) directly. Thus, the opportunity cost of time is the wage rate. The nature of the time constraint and the opportunity cost of time are discussed in greater detail in a later section. Finally, assume that the monetary cost of a trip to the site has two components: the admission fee, f, which could be zero, and the monetary cost of travel. This cost is $p_d \cdot d$, where p_d is the per-mile cost of travel and d is the round trip distance to the site.

Substituting the time constraint (9.3) into the monetary budget constraint (9.2) yields

$$M + w \cdot t^* = z + c \cdot x\,, \tag{9.4}$$

where c is the full price of a visit given by

$$\begin{aligned} c &= p_x + w \cdot (t_1 + t_2) \\ &= f + p_d \cdot d + w \cdot (t_1 + t_2). \end{aligned} \tag{9.5}$$

As equation (9.5) makes clear, the full price of a visit consists of four components: the admission fee, the monetary cost of travel to the site, the time cost of travel to the site, and the cost of time spent at the site. On the assumption that individuals are free to choose the number of hours worked at a given wage rate, the two time costs are valued at the wage rate. In a more realistic model with income and payroll taxes, time would be valued at the after-tax wage rate.

Maximizing equation (9.1) subject to the constraint of equation (9.4) will yield the individual's demand function for visits:

$$x = x(c, M, q)\,. \tag{9.6}$$

If all individuals spend the same amount of time at the site and have the same wage, then this component of the price of a visit is the same for all individuals. Given these assumptions, the data on rates of visitation, travel costs, and variation in entry fees (if any) can be used to estimate the coefficient on c in a travel cost-visitation function.

There are a number of key assumptions underlying this basic version of the travel cost model:

- First, it is assumed that the wage rate is the relevant opportunity cost of time. In a provocative article, Alan Randall (1994, 88, 90) argues that for several reasons "travel cost is inherently unobservable." One of the more compelling elements of his argument concerns the difficulties in defining and measuring the opportunity cost of time spent in travel, which is characterized by Randall as "an empirical mystery." The fundamental issue of valuing time is returned to later in this chapter.

- Second, it is assumed that all visits entail the same amount of time spent on the site. This assumption plays two important roles in the simple form of the model: it makes it possible to measure site usage by the scalar x, the number of visits; and it makes the full price of a visit, c, a parameter to the individual. If the individual chooses the amount of time of each visit, then c is an endogenous variable. Modeling the choice of time spent on site is discussed in the section on time below.
- Third, it is assumed that there is no utility or disutility derived from the time spent traveling to the site. If part of the trip involves the pleasures of driving through scenic countryside, then travel cost is overestimated by equation (9.5). Conversely, screaming children in the back seat of the car can raise the full travel cost beyond c. The way various uses of time affect utility and the shadow value of time is discussed in the section on time below.
- Fourth, it is assumed that each trip to the site is for the sole purpose of visiting the site. If the purpose of the trip is to visit two or more sites or to visit a relative en route, then at least part of the travel cost would be a joint cost that cannot be uniquely allocated among different purposes. If an alternative destination can be identified as the primary purpose of the trip, then the relevant cost of the visit to the recreation site is the incremental cost of adding the site visit to the trip given the trip to the primary destination. Parsons (2003) discussed this and other approaches to dealing with multiple destination trips.
- Fifth, it is assumed that there are no alternative recreation sites available to these individuals. If there are other sites available, then it is likely that the number of visits that an individual makes to the site in question will depend not only on its implicit price but also on the implicit prices of any substitute sites in the region. Omitting the price of a substitute site will bias the estimates of both the intercept term and the travel cost parameter (Caulkins, Bishop, and Bouwes 1985; Kling 1989; McKean and Revier 1990). The sign of the bias on the own-price coefficient depends on the correlation between it and the omitted substitute price variables, which in turn depends upon the spatial distribution of the population relative to the available sites.
- Sixth, it is assumed that the individual's choice of where to live (which is one determinant of the cost of a trip to a recreation site) is independent of preferences for recreation visits. If people choose residential locations so as to be near preferred recreation sites, then the price of a visit is endogenous. Parsons (1991) has suggested an instrumental variables approach that may avoid the bias that such choices would otherwise impart to the estimation of trip demand functions.

Beyond the assumptions underlying the generic single-site model in equation (9.6), an important practical limitation of the single-site model is that it provides little basis for identifying the marginal impact of changing site attributes (i.e., changes in q). Most recreation demand databases are cross-sectional in nature, providing no variation in measured site attributes. To the extent that q represents *perceived* site characteristics (such as expected fish catch rates or a subjective measure of overall water quality), variation in q can be found in cross-sections. However, relying upon these subjective measures creates two further complications. First, the subjective measures themselves must now be elicited from individuals in the sample. Second, valuing changes in *physical* site attributes requires understanding how changes in the physical attributes translate into changes in *perceived* site attributes.

Analysis of the demand for a single site would be appropriate if the researcher is interested in valuing the availability of that single site as long as information on the price and quality of substitute sites is appropriately included in the specification. However, many policy-relevant questions involve changes in the value of a set of sites due to changes in the number and availability of sites, or to changes in the qualities of these sites. In such cases, the interactions and the substitution effects among sites must be modeled explicitly. This calls for some form of multisite model.

Multisite models are estimated as systems of demand equations. For example, for each site j one might specify a demand equation of the following form:

$$x_j = x_j\left(c_j, \boldsymbol{C}_{-j}, M, q_j, \boldsymbol{Q}_{-j}\right), \quad j = 1,\ldots,J\,, \tag{9.7}$$

where x_j is the number of visits to the site, c_j is the full price of a visit to j, $\boldsymbol{C}_{-j} = \left(c_1,\ldots,c_{j-1},c_{j+1},\ldots,c_J\right)$ denotes the set of substitute prices for visits to other sites, q_j denotes the quality attribute for site j, and $\boldsymbol{Q}_{-j}$ denotes the vector of attributes for the other sites. As the number of sites and demand equations increases, multisite models can become cumbersome and difficult to estimate. However, over the past 25 years, a series of multisite models have been developed that, together with advances in computational capabilities, make it possible to characterize recreation demand when there are well over 50 alternatives in the individual's choice set (e.g., von Haefen, Phaneuf, and Parsons 2004; Murdock 2006; Herriges, Phaneuf, and Tobias 2008). These multisite models, and their limitations, are described in detail below.

Data Challenges and Competing Perspectives

As with many areas of applied economics, the evolution of recreation demand modeling reflects, in part, changes in the nature of the data available to analysts, as well as improvements in modeling and computing power. Early efforts to model recreation demand (e.g., Clawson 1959; Clawson and Knetsch 1966) employed aggregate or market-level data, obscuring the role that individual attributes and preference heterogeneity might play in the demand for recreation. The usual practice these days is to rely upon individual or household-level information on recreation activity aggregated over the course of a season. For example, a survey might be used to elicit the total numbers of trips taken by a sample of individuals to each of a series of sites during the course of a year, along with socio-demographic characteristics of each survey respondent. While this provides a rich data set for use in modeling recreation demand, it also presents a number of challenges for the analyst seeking to reflect all of the unique features of the data into a theoretically consistent and empirically manageable framework. Some of the more challenging features of the data include:

- *Frequent corner solutions*: In most applications, a large portion of the population chooses to not visit any or all of the alternatives in the choice set. This creates difficulties in data gathering, in that larger sample sizes will be needed in order to find site users. It also creates challenges in terms of the underlying economic and econometric specifications, requiring that the modeling framework allow for frequent zeros to avoid biased welfare estimates.
- *Count data*: The trip data obtained in recreation demand surveys takes the form of nonnegative integers. Traditional continuous demand models, solving the optimization problem in equations (9.1) through (9.3) using standard first-order conditions, will at best only approximate what is essentially the solution to a discrete choice problem.
- *Seasonal trip aggregates*: As noted above, recreation surveys typically elicit counts of the numbers of trips to a series of sites over the course of a season. As such, the data preclude modeling the roles that temporally varying factors (such as weather) play in influencing recreation choices, since it is not known when individual trips are taken. It also precludes modeling any dynamics associated with recreational choices, such as habit formation or variety seeking, since the sequence in which sites are visited is not known. While these problems can be avoided by gathering diary data (e.g., day-by-day or week-by-week records of trip activity), such databases are rare (Provencher and Bishop 1997; Moeltner and Englin 2004), because they are costly to administer and they typically suffer from sample attrition over time.
- *Limited information on site characteristics*: One advantage in recreation demand modeling is that the data exhibit substantial variation in the price variable (i.e., the travel cost), since each individual faces a different set of travel costs due to their differing proximities to the sites being modeled. Unfortunately, this is typically offset by limited information regarding the attributes of the individual sites. Most studies have available only a few site attributes (e.g., fish catch rates or some measure of water quality). As a result, recreation demand models can suffer from omitted variables bias—a problem that has received increased attention in recent years (e.g., Murdock 2006).

While many of the difficulties above can be individually (or in some cases jointly) addressed using econometric or statistical techniques, the challenge is to do so in a manner that still allows the analyst to impute the welfare implications of changing site availability or attributes. The next two sections review the single- and multiple-site models that have emerged in the recreation demand literature over the past

40 years. Each model focuses on a subset of the above issues, abstracting in each case from the remaining concerns in order to keep the model tractable. In many ways, it is like the parable of the blind men and the elephant, each encountering a different part of the elephant and perceiving it to be a different object (the ear suggesting the elephant is a fan, the leg suggesting that it is a pillar, etc.). Each approach to modeling recreation demand deals with aspects of the recreation demand process, but may not show the whole picture.

Single-Site Models

As noted above, most empirical applications of recreation demand include multiple sites. Further, single-site recreation demand specifications along the lines of (9.6) are rarely estimated as most recreational opportunities face competition from substitute sites. It is more common to estimate demand for a site using some version of (9.7), for example

$$x_1 = f_1\left(c_1, \boldsymbol{C}_{-1}, M, q_1, \boldsymbol{Q}_{-1}; \beta_1\right) + \varepsilon_1, \tag{9.8}$$

where $f_1(\cdot)$ denotes the specific functional form chosen to represent how the demand for trips to site 1 depends upon the full cost of visiting that site (c_1), the vector of costs of visiting potential substitute sites ($\boldsymbol{C}_{-1}$), income (M), the quality attribute for site 1 (q_1), and the attributes of substitute sites ($\boldsymbol{Q}_{-1}$). β_1 denotes the vector of parameters to be estimated and ε_1 denotes the random error term. The random term's specific role has historically been left vague, capturing a myriad of possible factors, including measurement or specification errors by the analyst and/or optimization errors by the individual consumer. The problem with the practice of simply "tacking on" error terms in this fashion is that how the error term is used in subsequent welfare analysis depends critically on what type of error it reflects. This issue is revisited below.

A second general issue with single-site models is that they are typically plagued by little, if any, variation in each site's attributes (i.e., the q_j). Most recreation demand databases are cross-sectional in nature, providing only a single observation on the attributes of each site in the choice set. For example, with data on the number of trips to lake 1 during the course of a year, q_j might measure the annual mean water quality at lake j. Although lake quality will typically vary over the course of the year, it is not possible to directly link this variation to observed demand, since it is not known when individual trips are taken.

Without variation in the site attributes across individuals, their impact on trip demand cannot be identified in a single-site model, leading most analysts to employ a simpler functional representation:

$$x_1 = f_1\left(c_1, \boldsymbol{C}_{-1}, M; \beta_1\right) + \varepsilon_1. \tag{9.9}$$

In the following subsection, two competing single-site models are described: censored regression models and count data models. Unless otherwise indicated, it is assumed that the unit of observation is a randomly selected individual from the target population. Additional details regarding the estimation and interpretation of these models, as well as the impact of alternative sampling approaches, can be found in Haab and McConnell (2002, ch. 7).

Censored Regression Models

As noted in the data section, a common feature of recreation demand data is the nonnegative nature of trip demand, with a large number of corner solutions (i.e., many individuals choose not to visit the site being modeled). A popular framework used to capture these data characteristics is the Tobit model, a special case in the more general class of censored regression models. The model in equation (9.9) is replaced by

$$x_{i1} = \begin{cases} f_1(c_{i1}, \boldsymbol{C}_{-i1}, M_i; \beta_1) + \varepsilon_{i1} & x_{i1}^* > 0 \\ 0 & x_{i1}^* \leq 0, \end{cases} \tag{9.10}$$

where the subscript i is introduced to denote the observation for a given individual ($i = 1, \ldots, N$) and

$$x_{i1}^* = f_1\left(c_{i1}, \boldsymbol{C}_{-i1}, M_i; \beta_1\right) + \varepsilon_{i1} \tag{9.11}$$

is a latent (unobserved) variable determining participation, sometimes referred to as *potential* demand. If x_{i1}^* is positive, the individual participates in recreation and observed demand equals potential demand (i.e., $x_{i1} = x_{i1}^*$). If, however, potential demand is negative (or zero), the individual chooses not to recreate and observed demand is zero. Thus, observed demand is a censored version of potential demand, censoring potential demand from below at zero.

Estimating the underlying demand relationship in (9.10) requires specifying not only the functional form for $f_1(\cdot)$, but also the underlying distribution of the error term ε_{i1}. Typically it is assumed that error terms are independent and identically distributed (iid) normal random variables (i.e., $\varepsilon_{i1} \sim \text{iid}\, N\left(0, \sigma^2\right)$). Given this assumption, the parameters $\left(\beta_1, \sigma\right)$ can be estimated using maximum likelihood techniques or Heckman's (1976) two-step estimation procedure. Interpreting the coefficients of the model becomes more complex, even when $f_1(\cdot)$ is linear. When the error term is normally distributed, expected demand becomes

$$\begin{aligned} E\left(x_{i1}\right) &= \Pr\left(x_{i1}^* > 0\right) E\left(x_{i1}^* \mid x_{i1}^* > 0\right) \\ &= \Phi\left[\frac{f_1\left(c_{i1}, \boldsymbol{C}_{-i1}, M_i; \beta_1\right)}{\sigma}\right] f_1\left(c_{i1}, \boldsymbol{C}_{-i1}, M_i; \beta_1\right) \\ &\quad + \sigma \phi\left[\frac{f_1\left(c_{i1}, \boldsymbol{C}_{-i1}, M_i; \beta_1\right)}{\sigma}\right], \end{aligned} \tag{9.12}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal cumulative distribution function (cdf) and probability density function (pdf) respectively.

Welfare analysis using the Tobit model is typically conducted on the basis of consumer surplus measures. One issue alluded to above is how the error term (ε_{i1}) in the model is treated during these calculations. Bockstael and Strand (1987) noted that, if the error term represents measurement error, expected demand along the lines of (9.12) should be used in computing consumer surplus (essentially integrating out the unknown measurement error). If, on the other hand, the error term captures potential omitted variables or specification errors, then they argued that observed demand (x_{i1}) is the best predictor of demand to use in computing consumer surplus. Haab and McConnell (2002, 163) provided a numerical example in which the two approaches can yield welfare estimates that differ by a factor of five, making the interpretation of the error term a potentially important factor in conducting welfare analysis.

Finally, while the Tobit model is a popular form of the censored regression models, it does impose considerable structure on the relationship between the participation decision (i.e., decisions at the *extensive* margin) and the numbers of trips taken once participation is chosen (i.e., decisions at the *intensive* margin). Specifically, the same variables and error terms drive both decisions and the parameters are assumed to be the same. Bockstael et al. (1990) provide a discussion of alternative frameworks that relax these restrictions.

Count Data Models

An obvious limitation of censored regression models is that they do not explicitly reflect the discrete count nature of most trip data. When the counts themselves are typically large (e.g., in the twenties and thirties), a continuous approximation to these discrete count outcomes is likely to perform reasonably well. However, in many recreation demand settings, a large portion of the sample will take few, if any, trips to the site being modeled and a continuous approximation will not perform as well. This has led many analysts (e.g., Hellerstein and Mendelsohn 1993; Smith 1988) to turn to count data regression models, which explicitly limit the dependent variable to nonnegative integer values. A general discussion of count data models can be found in Cameron and Trivedi (1998), with discussion focused on the recreation demand context found in Haab and McConnell (2002, sec. 7.4).

The most basic of the count models is the Poisson regression model, with the conditional probability density function for trips x_{i1} given by

$$\Pr(x_{x1} = k \mid \boldsymbol{Z}_i) = \begin{cases} \dfrac{e^{-\lambda_i}\lambda_i^k}{k!} & k = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{9.13}$$

where $\boldsymbol{Z}_i$ denotes the factors thought to impact the number of trips taken by individual i, and

$$\lambda_i = \lambda(\boldsymbol{Z}_i;\beta) = E(x_{i1}|\boldsymbol{Z}_i) = \mathrm{Var}(x_{i1}|\boldsymbol{Z}_i) \tag{9.14}$$

denotes both the conditional mean and conditional variance of the count variable x_{i1}. For example, in the simplest version of the single-site model along the lines of (9.9), one might specify $\boldsymbol{Z}_i = \left(c_{i1}, \boldsymbol{C}_{-i1}, M_i\right)$. It is typically also assumed that $\lambda\left(\boldsymbol{Z}_i;\beta\right)$ has a linear exponential form. In other words,

$$\lambda\left(\boldsymbol{Z}_i;\beta\right) = \exp\left(\boldsymbol{Z}_i\beta\right). \tag{9.15}$$

This structure insures that (a) mean trips are positive and (b) the log-likelihood function used in maximum likelihood estimation is globally concave in the parameters β, which in turn eases estimation by insuring a unique optimum.

The simplicity of the Poisson regression model (with a linear exponential mean) carries with it a number of advantages. First, as already noted, unlike the censored regression model, the framework explicitly reflects the count nature of the trip data. Second, while the model is non-linear, interpretation of the parameters is relatively straightforward. Specifically, consider a change in z_{ik}, the *k*th element of $\mathbf{Z}_i$. The marginal effect of such a change on expected trips is given by

$$\frac{\partial E\left(x_{i1}\middle|\boldsymbol{Z}_i\right)}{\partial z_{ik}} = \beta_k \exp\left(\boldsymbol{Z}_i\beta\right) = \beta_k E\left(x_{i1}\middle|\boldsymbol{Z}_i\right). \tag{9.16}$$

Thus, the marginal impact of any factor is proportional to the expected number of trips, with the factor of proportionality being the corresponding parameter. Moreover, the maximum likelihood estimation of the Poisson regression model is mean fitting, with

$$\overline{x}_1 \equiv \tfrac{1}{N}\sum_{i=1}^{N} x_{i1} = \tfrac{1}{N}\sum_{i=1}^{N}\exp\left(\mathbf{Z}_i\hat{\beta}\right), \tag{9.17}$$

where $\hat{\beta}$ denotes the maximum likelihood parameter estimates. As a result, the mean marginal effect for the sample used in estimation reduces to $\hat{\beta}_k\overline{x}_1$.

Third, welfare analysis, typically conducted on the basis of consumer surplus, is straightforward. In particular, as Haab and McConnell (2002, 167) demonstrate, the welfare loss to individual *i* from closure of a site, obtained by integrating under the expected demand function in (9.15), is given by

$$CS_{1i} = \frac{1}{-\beta_{1c}}\exp\left(\mathbf{Z}_i\beta\right), \tag{9.18}$$

where β_{1c} denotes the coefficient on travel cost to site 1. Again, employing (9.17) the average welfare loss from the closure of site 1 reduces to $\overline{CS}_1 = -\overline{x}_1 \big/ \hat{\beta}_{1c}$.

The advantages of the Poisson regression model, however, do not come without a cost. In particular, the Poisson specification assumes, as noted in (9.14), that the conditional mean of trips is equal to the conditional variance of trips, a property known as *equidispersion*. Unfortunately, this assumption is frequently rejected in applications, with it often being the case that *overdispersion* holds (i.e., $\mathrm{Var}\left(x_{i1}\middle|\mathbf{Z}_i\right) > E\left(x_{i1}\middle|\mathbf{Z}_i\right)$). If it remains the case that $E\left(x_{i1}\middle|\mathbf{Z}_i\right) = \exp\left(\mathbf{Z}_i\beta\right)$, then maximum likelihood estimates from the Poisson model will still be consistent, but the estimated standard errors will need to be corrected. Alternatively, one can replace the Poisson specification with a more general count data model that does not impose *equidispersion*, such as the Negative Binomial or zero-inflated Poisson models (see, for example, Haab and McConnell 1996 and 2002, sec. 7.4.2).

Multiple-Site Models

While single-site models continue to be used, they are relatively rare, in large part because they are of limited use in policy analysis. In a cross-sectional setting, where one observes trips to a single-site given a fixed set of available site attributes, the impact of those site attributes cannot be isolated. Analysts are typically limited to estimating models of the sort identified in (9.9). While these models can be used to estimate the welfare impact of site closures or changes in access fees, they provide no information on the welfare implications of changing site amenities (e.g., improved water quality). Multisite models, on the other hand, by comparing visitation patterns to a series of available sites with differing site attributes and differing travel costs, provide the basis for identifying the marginal impact of those site attributes on individual well-being. In this section, four competing multisite models are reviewed. Additional details regarding the econometric issues associated with these models can be found in Haab and McConnell (2002, ch. 8).

The Linked Model

The introduction of multiple sites to the recreation demand problem significantly complicates modeling. Even in the simplest of settings in which trips are of fixed length, the analyst faces the task of characterizing the integer-valued numbers of trips to each of J available sites over some time horizon, subject to both budget and time constraints. The earliest efforts (e.g., Hanemann 1978; Feenberg and Mills 1980; Caulkins, Bishop, and Bouwes 1986) focused on the choice among recreational sites *conditional* on taking a trip, the so-called *site selection* problem. This abstracts from the total numbers of trips taken, the so-called *participation* problem. The linked model, originally introduced by Bockstael, Hanemann, and Kling (1987), provided a way of integrating these two problems into a two-stage model, with the participation model characterizing total trips as a function of indices measuring the overall appeal of the available sites, with the latter estimated as a part of the site selection model.

Site Selection

Site selection models represent an application of discrete choice analysis in which consumers are assumed to select one alternative from a finite choice set. Much of this literature is couched in terms of the hypothesis of random utility maximization (RUM) (McFadden 1974, 1981). Within the RUM construct, an individual i is assumed to receive utility from choosing alternative j, represented by the conditional utility function

$$u_{ij} = u\left(p_{ij}, \boldsymbol{Z}_{ij}, M_i\right), \quad j = 1, \ldots, \mathcal{J}, \tag{9.19}$$

where p_{ij} denotes the cost of alternative j to individual i, $\boldsymbol{Z}_{ij}$ denotes the attributes of the individual and/or the alternatives that impact the individual's conditional utility, and M_i denotes the individual's income. In the context of site selection, this conditional utility is typically assumed to take the form

$$u_{ij} = u\left(M_i - c_{ij}, \tilde{\boldsymbol{Q}}_j, \tilde{\boldsymbol{S}}_i\right), \tag{9.20}$$

where c_{ij} denotes the travel cost for individual i in visiting site j (so that $M_i - c_{ij}$ denotes the individual's residual income after purchasing a visit to site j), $\tilde{\boldsymbol{Q}}_j$ denotes the attributes of site j, and $\tilde{\boldsymbol{S}}_i$ denotes the attributes of the individual. The RUM hypothesis assumes that individual i simply chooses the alternative that maximizes his/her utility. That is, alternative j is chosen (denoted by $y_{ij} = 1$, as opposed to $y_{ij} = 0$ when it is not chosen) if

$$u\left(M_i - c_{ij}, \tilde{\boldsymbol{Q}}_j, \tilde{\boldsymbol{S}}_i\right) > u\left(M_i - c_{ik}, \tilde{\boldsymbol{Q}}_k, \tilde{\boldsymbol{S}}_i\right) \;\forall\; k \neq j \;. \tag{9.21}$$

It is important to emphasize that there is nothing random in this choice process from the individual's perspective, at least at the time the actual choice is made. They are assumed to have well-defined preferences over the alternatives in the choice set and to be able to identify the alternative yielding the maximum utility.[1] The randomness in the RUM framework is from the perspective of the analyst, who does not know all of the factors impacting the individual's decision or the correct functional relationship among the factors. Instead, the researcher represents the conditional utility function using the form

$$u_{ij} = v\left(M_i - c_{ij}, \boldsymbol{Q}_j, \boldsymbol{S}_i\right) + \varepsilon_{ij}, \tag{9.22}$$

where $\boldsymbol{Q}_j$ denotes the observable site characteristics (versus the full set of characteristics $\tilde{\boldsymbol{Q}}_j$) and $\boldsymbol{S}_i$ denotes the observable set of individual characteristics (versus the full set $\tilde{\mathbf{S}}_i$). The error captures the effect of both unobservables, those factors in $\left(\tilde{\boldsymbol{Q}}_j, \tilde{\boldsymbol{S}}_i\right)$ but not in $\left(\boldsymbol{Q}_j, \boldsymbol{S}_i\right)$, and model misspecification, since

$$\varepsilon_{ij} \equiv u\left(M_i - c_{ij}, \tilde{\boldsymbol{Q}}_j, \tilde{\boldsymbol{S}}_i\right) - v\left(M_i - c_{ij}, \boldsymbol{Q}_j, \boldsymbol{S}_i\right) \;. \tag{9.23}$$

With this limited information, the analyst can only specify the probability that an individual will choose a given alternative with

$$\begin{aligned} \Pr\left(y_{ij} = 1 \middle| \text{Trip}\right) &= \Pr\left[v_{ij} + \varepsilon_{ij} > v_{ik} + \varepsilon_{ik} \;\forall k \neq j\right] \\ &= \Pr\left[\Delta v_{ijk} > \Delta \varepsilon_{ijk}\right], \end{aligned} \tag{9.24}$$

1 If the individual has any uncertainty over the alternatives in the choice set, he/she is assumed to be able to integrate that uncertainty into an overall measure of utility (e.g., forming an expected utility to be used in comparing alternatives).

where $v_{ij} \equiv v\left(M_i - c_{ij}, \boldsymbol{Q}_j, \boldsymbol{S}_i\right)$, $\Delta v_{ijk} \equiv v_{ij} - v_{ik}$, and $\Delta\varepsilon_{ijk} \equiv \varepsilon_{ik} - \varepsilon_{ij}$. Note that the probability in (9.24) is conditional on the individual choosing to take a trip. Different discrete choice models result, depending upon the specification of the error terms ε_{ij}. If, for example, it is assumed that the error terms are independently and identically distributed with a Type I Extreme Value distribution, then a logit model results with

$$P_{ij} \equiv \Pr\left(y_{ij} = 1 \middle| \text{Trip}\right) = \frac{\exp\left(v_{ij}\right)}{\sum_{k=1}^{J} \exp\left(v_{ik}\right)} . \tag{9.25}$$

Alternatively, if $\boldsymbol{\varepsilon}_{i\bullet} \equiv \left(\varepsilon_{i1}, \ldots, \varepsilon_{iJ}\right)$ is drawn from a generalized extreme value (GEV) distribution, then a nested logit model results.

Welfare analysis is typically conducted on the basis of compensating variation. Specifically, consider a change in travel cost, site attributes and/or the set of alternatives available to the individual from $\left(\boldsymbol{C}_i^0, \boldsymbol{Q}^0, J^0\right)$ to $\left(\boldsymbol{C}_i^1, \boldsymbol{Q}^1, J^1\right)$ where $\boldsymbol{C}_i^s = \left(c_{i1}^s, \ldots, c_{iJ}^s\right)$ and $\boldsymbol{Q}^s = \left(\boldsymbol{Q}_1^s, \ldots, \boldsymbol{Q}_J^s\right)$ for $s = 0,1$. The compensating variation associated with this change is implicitly defined by

$$\begin{aligned} &\operatorname*{Max}_{j \in J^1} \left[v\left(M_i - c_{ij}^1 - CV_{|\text{Trip}}, \boldsymbol{Q}_j^1, \boldsymbol{S}_i\right) + \varepsilon_{ij}\right] \\ &= \operatorname*{Max}_{j \in J^0} \left[v\left(M_i - c_{ij}^0, \boldsymbol{Q}_j^0, \boldsymbol{S}_i\right) + \varepsilon_{ij}\right], \end{aligned} \tag{9.26}$$

where $CV_{|\text{Trip}} = CV_{|\text{Trip}}\left(M_i - c_i^0, M_i - c_i^1, \boldsymbol{Q}^0, \boldsymbol{Q}^1, J^0, J^1, \boldsymbol{\varepsilon}_{i\bullet}\right)$. The compensating variation represents the amount that equates the individual's utility before and after the change, allowing for the possibility that the individual may change their preferred alternative. As the notation indicates, the compensating variation is itself a random variable from the perspective of the analyst, since it depends on $\boldsymbol{\varepsilon}_{i\bullet}$. It is important to note that, in the context of the site selection model, the compensating variation $CV_{|\text{Trip}}$ is conditional on taking a trip. The unconditional welfare impact will be potentially smaller, since the individual has the option of simply staying at home.

If, as is typically the case in the literature, it is assumed that the marginal utility of income is constant, then

$$v_{ij}^s \equiv v\left(M_i - c_{ij}^s, \boldsymbol{Q}_j^s, \boldsymbol{S}_i\right) = \beta_M\left(M_i - c_{ij}^s\right) + \widehat{v}\left(\boldsymbol{Q}_j^s, \boldsymbol{S}_i\right) , \tag{9.27}$$

where β_M denotes the marginal utility of income and $\widehat{v}\left(\boldsymbol{Q}_j^s, \boldsymbol{S}_i\right)$ denotes the remaining terms' impacting site utility. One can then explicitly solve for $CV_{|\text{Trip}}$ in (9.26), with

$$\begin{aligned} CV_{|\text{Trip}} = \frac{1}{\beta_M}\Bigg\{&\operatorname*{Max}_{j \in J^1} \left[v\left(M_i - c_{ij}^1, \boldsymbol{Q}_j^1, \boldsymbol{S}_i\right) + \varepsilon_{ij}\right] - \\ &\operatorname*{Max}_{j \in J^0} \left[v\left(M_i - c_{ij}^0, \boldsymbol{Q}_j^0, \boldsymbol{S}_i\right) + \varepsilon_{ij}\right]\Bigg\}. \end{aligned} \tag{9.28}$$

The term in curly brackets represents the change in utility, and dividing by the marginal utility of income monetizes this change.

In the case of the logit model, computing the mean value of $CV_{|\mathrm{Trip}}$ becomes particularly straightforward, since

$$\begin{aligned} E\left\{\underset{j\in\mathcal{J}^s}{\mathrm{Max}}\left[v\left(M_i - c_{ij}^s, \boldsymbol{Q}_j^s, \boldsymbol{S}_i\right) + \varepsilon_{ij}\right]\right\} &= \ln\left[\sum_{j\in\mathcal{J}^s}\exp\left(v_{ij}^s\right)\right] + K \quad s = 0,1 \\ &= I\left(\boldsymbol{C}_i^s, M_i, \boldsymbol{Q}^s, \boldsymbol{S}_i\right) + K \end{aligned} \tag{9.29}$$

where K is Euler's constant and $v_{ij}^s \equiv v\left(M_i - c_{ij}^s, \boldsymbol{Q}_j^s, \boldsymbol{S}_i\right)$. A similar expression exists for nested logit models (see, for example, McFadden 1984; Morey 1999). The term $I\left(M_i, \boldsymbol{C}_i^s, \boldsymbol{Q}^s, \boldsymbol{S}_i\right)$ is referred to as the *inclusive value*. Using this result along with (9.28) yields

$$\overline{CV}_{|\mathrm{Trip}} = E\left(CV_{|\mathrm{Trip}}\right) = \frac{1}{\beta_M}\left\{\ln\left[\sum_{j\in\mathcal{J}^1}\exp\left(v_{ij}^1\right)\right] - \ln\left[\sum_{j\in\mathcal{J}^0}\exp\left(v_{ij}^0\right)\right]\right\}. \tag{9.30}$$

The term in curly brackets represents the change in expected utility. An interesting special case arises when considering the welfare impact from a site closure, say site 1, where

$$\begin{aligned} \overline{CV}_{1|\mathrm{Trip}} &= \frac{1}{\beta_M}\left\{\ln\left[\sum_{j=2}^{\mathcal{J}}\exp\left(v_{ij}^0\right)\right] - \ln\left[\sum_{j=1}^{\mathcal{J}}\exp\left(v_{ij}^0\right)\right]\right\} \\ &= \frac{1}{\beta_M}\ln\left(1 - P_{i1}\right) \approx \frac{-P_{i1}}{\beta_M}. \end{aligned} \tag{9.31}$$

As one would expect, closing a site represents a loss, with the loss increasing as the site represents a more popular alternative.

The simple functional form for the logit model probabilities in (9.25) makes estimation, typically via maximum likelihood, relatively easy. Moreover, welfare analysis is also straightforward using the compensating variation formula in (9.30). However, these simplifications come with a cost. The logit model imposes the well-known Independence of Irrelevant Alternatives (IIA) assumption, which means that the relative choice probabilities for any two alternatives (i.e., P_{ij}/P_{ik}) depend only on the attributes of the two alternatives and not on any of the other alternatives available to the individual. This assumption is often rejected in practice (Kling and Thomson 1996).

Another way to view the restrictiveness of the logit specification is to consider what it says about the correlation among the error terms (i.e., the ε_{ij}). Specifically, it assumes that the error terms are independently and identically distributed. Yet the error terms by construction represent those unobservable factors influencing the choices individuals make. In most applications, particularly those with limited information about the individuals and/or the alternative characteristics, it seems unlikely that the unobservables will be uncorrelated across choice alternatives. For example, in a model of lake recreation, one unobservable might be whether the individual owns a powerboat or not. This unobservable is likely to be a part of the error term for all the lakes in which power boating is allowed, inducing correlation among the associated error terms.

Over the past 20 years, advances in computational capabilities have enabled analysts to employ a much wider range of error specifications, allowing for complex patterns of correlation among the error terms. These models include mixed logit, multivariate probit and latent class (or finite mixture) models. Train (2009) provides an excellent general discussion of these models, while Herriges and Phaneuf (2002) provide a discussion in the context of recreation demand. While these models complicate both estimation and welfare calculations, the underlying process is similar, with welfare analysis conducted on the basis of Hicksian welfare measures.

Participation

The participation portion of the linked model focuses on the total number of trips, with a generic representation,

$$x_i = g\left(\boldsymbol{L}_i, \boldsymbol{Z}_i, M_i\right) + \eta_i \ , \tag{9.32}$$

where $x_i = \sum_{j=1}^{J} x_{ij}$ denotes the total number of trips taken by individual i, $\boldsymbol{L}_i$ denotes a vector of variables that link the participation equation to the site selection equation, $\boldsymbol{Z}_i$ denotes a set of other variables thought to influence the number of trips taken by an individual, and η_i is a random error term.

In the first linked model by Bockstael, Hanemann, and Kling (1987), BHK hereafter, the linking variable was the inclusive value term $I\left(\boldsymbol{C}_i, M_i, \boldsymbol{Q}, \boldsymbol{S}_i\right)$ implicitly defined in (9.29). Thus, the number of trips taken by an individual depends on the expected utility derived from taking a trip. BHK suggested that the welfare impact of a policy change could then be computed as the expected compensating variation of the change per trip, calculated using the site selection model's equation (9.30), multiplied by the number of trips (either before or after the change).

There have been a number of subsequent variations on the linked model proposed in the literature. Parsons and Kealy (1995) and Feather, Hellerstein, and Tomasi (1995) proposed using as linking variables, not the inclusive value function, but rather "average price" and "average quality" variables, where the weights in computing each average are the site selection probabilities predicted by the site selection model. Thus, (9.32) becomes

$$x_i = g\left(\bar{c}_i, \bar{\boldsymbol{Q}}_i, \boldsymbol{Z}_i, M_i\right) + \eta_i \,, \tag{9.33}$$

where

$$\bar{c}_i = \sum_{j=1}^{J} \hat{P}_{ij|\text{Trip}} c_{ij} \tag{9.34}$$

$$\bar{\boldsymbol{Q}}_i = \sum_{j=1}^{J} \hat{P}_{ij|\text{Trip}} \boldsymbol{Q}_j \tag{9.35}$$

and $\hat{P}_{ij|\text{Trip}}$ denotes the fitted site selection probabilities.

Finally, Hausman, Leonard, and McFadden (1995) suggested a slight variant on the BHK approach, replacing the inclusive value with its monetized counterpart (i.e., $p = I\left(M_i, \boldsymbol{C}_i, \boldsymbol{Q}, \boldsymbol{S}_i\right) / \beta_M$), arguing that it can formally be viewed as a price index for recreation demand. The authors claimed that the resulting model is utility-theoretic and consistent with a two-stage budgeting process. Unfortunately, this claim has subsequently been shown to not be true in general (e.g., Smith 1997). Instead, as Parsons and Kealy (1995, 360) suggested, the linked model should not be viewed as "derived from a single overall utility maximization problem," but rather as an approximation to the underlying optimization.

The Repeated RUM Model

While the linked model represented a significant step forward in recreation demand analysis, addressing both the site selection and participation decisions, its lack of an underlying unified utility theoretic framework forces the analyst to rely on approximations when conducting welfare analysis. Moreover, by dividing the two interrelated decisions into distinct econometric tasks, efficiency is likely to be sacrificed.

Morey, Rowe, and Watson (1993) proposed an approach based on the simple concept of adding one alternative to the choice set found in the standard site selection model, namely the alternative of staying at home, and repeating this choice over a series of T choice occasions. For example, in the context of annual recreational usage, T might be 52, allowing for the individual to take a trip (or stay at home) during each week of the year. The resulting repeated RUM model has become the workhorse of the recreation demand literature, providing a unified framework for modeling both site selection and participation and an internally consistent basis for welfare analysis.

The starting point in the repeated RUM model is similar to the site selection model described in the previous subsection. In this case, individual i is assumed to receive utility $\tilde{u}_{ijt}$ from choosing alternative j on choice occasion t, where a traditional specification might take the form

$$\tilde{u}_{ijt} = \begin{cases} \beta_M M_i + \delta \boldsymbol{S}_i + \varepsilon_{ijt} & j = 0 \\ \beta_M \left(M_i - c_{ij}\right) + \gamma \boldsymbol{Q}_j + \varepsilon_{ijt} & j = 1, \ldots, J. \end{cases} \tag{9.36}$$

In this case, site attributes ($\boldsymbol{Q}_j$) impact the utility received from visiting the corresponding recreational site. Individual attributes ($\boldsymbol{S}_i$), such as age, gender, and education, impact the individual's propensity to stay at home by increasing or decreasing $\tilde{u}_{i0t}$ relative to all the other $\tilde{u}_{ijt}$, but do not impact the relative appeal of the recreational sites (i.e., $\tilde{u}_{ijt}$ versus $\tilde{u}_{ikt}$, $j, k = 1, \ldots, J$). One can generalize the specification in (9.36) to allow for the latter effects by introducing interaction terms between $\boldsymbol{S}_i$ and $\boldsymbol{Q}_j$ into the trip utilities (i.e., $\tilde{u}_{ijt}, j = 1,\ldots,J$).

The specification in (9.36) assumes that the marginal utility of income is constant at β_M. Since $\beta_M M_i$ appears in all of the choice utilities, it has no impact on their relative values and, hence, has no impact on the choices made. Consequently, an equivalent representation of consumer utility is

$$u_{ijt} = \begin{cases} \delta \boldsymbol{S}_i + \varepsilon_{ijt} & j = 0 \\ \beta_M c_{ij} + \gamma \boldsymbol{Q}_j + \varepsilon_{ijt} & j = 1,\ldots,J. \end{cases} \tag{9.37}$$

As in the case of the site selection model, a key step in completing the model specification is the choice of the error distribution. Assuming that the error terms are iid Extreme Value will yield logit probabilities analogous to those in (9.25). The problem is that this assumes the error terms are uncorrelated, a restriction that is unlikely to hold, particularly among the trip alternatives. A more common specification is to assume that the error vector $\boldsymbol{\varepsilon}_{i\bullet t} \equiv \left(\varepsilon_{i0t}, \varepsilon_{i1t}, \ldots, \varepsilon_{iJt}\right)$ is drawn from a Generalized Extreme Value (GEV) distribution yielding a nested logit model with the trip alternatives belonging to a single nest. Intuitively, this nesting structure implies that the trip alternatives are more similar to each other than to the stay-at-home option. Statistically, the assumption implies that the ε_{ijt} (j = 1,…,J) are correlated with each other, but not with ε_{i0t}. Formally, the resulting choice probabilities take the form

$$P_{ijt} = \Pr\left(y_{ijt} = 1\right) = \Pr\left(y_{ijt} = 1 \middle| y_{i0t} = 0\right) \Pr\left(y_{i0t} = 0\right) = P_{ij|\text{Trip}} P_{\text{Trip}}, \tag{9.38}$$

where

$$P_{ij|\text{Trip}} = \frac{\exp\left(\dfrac{v_{ij}}{\theta}\right)}{\sum_{j=1}^{J} \exp\left(\dfrac{v_{ik}}{\theta}\right)} \tag{9.39}$$

denotes the probability of visiting site j conditional on choosing to take a trip (i.e., to not stay at home),

$$P_{\text{Trip}} = \frac{\left[\sum_{j=1}^{J} \exp\left(\dfrac{v_{ij}}{\theta}\right)\right]^{\theta}}{\exp\left(v_{i0}\right) + \left[\sum_{j=1}^{J} \exp\left(\dfrac{v_{ij}}{\theta}\right)\right]^{\theta}} \tag{9.40}$$

denotes the probability of taking a trip, and

$$v_{ij} = \begin{cases} \delta \boldsymbol{S}_i & j = 0 \\ -\beta_M c_{ij} + \gamma \boldsymbol{Q}_j & j = 1,\ldots,J. \end{cases} \tag{9.41}$$

The parameter θ in equations (9.39) and (9.40) is known as the *dissimilarity* coefficient and is restricted to lie in the unit interval, specifically $\theta \in (0,1]$ for consistency with the RUM model. The smaller θ becomes, the more *similar* the trip alternatives become (with greater correlation among their error terms). At the other extreme, with $\theta = 1$ the model reduces to a standard logit model.

Notice that the choice probabilities in (9.38) do not vary by choice occasion. This is essentially required by the data, since the analyst does not typically know when individual trips are taken, so that time-dependent factors cannot be accounted for. Also, notice that equation (9.39) corresponds to the site selection model (9.25) discussed above. With the error terms assumed to be iid, the contribution of individual i to the likelihood function becomes

$$L_i = \prod_{j=0}^{J} P_{ij}^{n_{ij}}, \tag{9.42}$$

where $n_{ij} = \sum_{t=1}^{T} y_{ijt}$ denotes the number of times that individual i chose alternative j across the T choice occasions.

Welfare analysis using the repeated RUM model is straightforward. Consider a policy change from $\left(\boldsymbol{C}_i^0, \boldsymbol{Q}_j^0, J^0\right)$ to $\left(\boldsymbol{C}_i^1, \boldsymbol{Q}_j^1, J^1\right)$. For each choice occasion, the compensating variation for the change (CV_{CO}) is calculated just as it was for the site selection model, only now the choice set includes the option of staying at home. Specifically,

$$\begin{aligned} &\underset{j\in J^1}{\text{Max}}\left[v\left(M_i - c_{ij}^1 - CV_{CO}, \boldsymbol{Q}_j^1, \boldsymbol{S}_i\right) + \varepsilon_{ijt}\right] \\ &= \underset{j\in J^0}{\text{Max}}\left[v\left(M_i - c_{ij}^0, \boldsymbol{Q}_j^0, \boldsymbol{S}_i\right) + \varepsilon_{ijt}\right]. \end{aligned} \tag{9.43}$$

If the nested logit model is assumed, as in (9.38) through (9.40), then the mean compensating variation per choice occasion has a convenient closed form expression:

$$\overline{CV}_{CO} = E\left(CV_{CO}\right) = \frac{1}{\beta_M}\left(\begin{matrix} \ln\left\{\exp\left(v_{i0}^1\right) + \left[\sum_{j\in J^1}\exp\left(\frac{v_{ij}^1}{\theta}\right)\right]^{\theta}\right\} - \\ \ln\left\{\exp\left(v_{i0}^0\right) + \left[\sum_{j\in J^0}\exp\left(\frac{v_{ij}^0}{\theta}\right)\right]^{\theta}\right\} \end{matrix}\right) \tag{9.44}$$

where v_{ij}^s denotes (9.41) evaluated under scenario s, $s = 0,1$. The expected compensating variation for the season as a whole is then simply $T \cdot \overline{CV}_{CO}$.

Consider the special case in which the policy scenario corresponds to the closure of site 1. As Haab and McConnell (2002) demonstrated in their equation (8.45), the corresponding expected compensating variation per choice occasion becomes

$$\overline{CV}_{1,CO} = \frac{1}{\beta_M} \ln\left[1 - P_{\text{Trip}} + \left(1 - P_{1|\text{Trip}}\right)^{\theta} P_{\text{Trip}}\right]. \tag{9.45}$$

Comparing this to the expected compensating variation from the site selection model in equation (9.31), one can show that

$$\overline{CV}_{1,CO} = \frac{1}{\beta_M} \ln\left[1 - P_{\text{Trip}} + \exp\left(\beta_M \overline{CV}_{1|\text{Trip}}\right) P_{\text{Trip}}\right]. \tag{9.46}$$

Clearly, $\overline{CV}_{1,CO} \to \overline{CV}_{1|\text{Trip}}$ as $P_{\text{Trip}} \to 1$, but in general $\left|\overline{CV}_{1,CO}\right| < \left|\overline{CV}_{1|\text{Trip}}\right|$ (i.e., the site selection model will overstate the unconditional welfare impact, ignoring the option to substitute to the no-trip option). At the same time, using (9.45), $\overline{CV}_{1,CO} \to 0$ as $\theta \to 0$. The intuition here is that as $\theta \to 0$, the various trip alternatives are becoming more similar (i.e., better substitutes). Losing one of the sites has little impact in this situation since the individual can readily substitute to another similar site.

Since its introduction 20 years ago, a variety of issues have emerged with the basic repeated RUM model. First, the model assumes that there are a fixed number of choice occasions (T). The choice of T is ad hoc, though empirically the resulting welfare calculations have been found not to be particularly sensitive to changes in its value (Parsons, Jakus, and Tomasi 1999; Shaw and Ozog 1999). Typically, T is chosen to accommodate the largest number of trips taken by individuals in the sample or some reasonable upper limit on the numbers of trips.

Second, the basic nested logit model described above assumes that the unobservable factors influencing where the individual chooses to recreate, captured by the error term vector $\varepsilon_{i\bullet t}$, are independent over choice occasions, which seems unlikely. One generalization that relaxes this assumption is the mixed logit framework, which replaces (9.41) with

$$v_{ij}\left(\zeta_i\right) = \begin{cases} \delta_i \boldsymbol{S}_i & j = 0 \\ \beta_{Mi} c_{ij} + \gamma_i \boldsymbol{Q}_j & j = 1, \ldots, J. \end{cases} \tag{9.47}$$

where $\zeta_i = \left(\beta_{Mi}, \delta_i, \gamma_i\right)$ denotes the parameters of the model, which are now assumed to be individual specific and drawn from some specified distribution $f(\zeta)$. The choice probabilities, as well as their contribution to the likelihood function, become conditional on the individual's realization of ζ_i. Instead of (9.42), the unconditional likelihood function becomes

$$L_i = \int L_i(\zeta) f(\zeta) d\zeta = \int \left\{ \prod_{j=0}^{J} \left[P_{ij}(\zeta)\right]^{n_{ij}} \right\} f(\zeta) d\zeta \ . \tag{9.48}$$

The introduction of random parameters induces correlation over the choice occasions. The parameters can be drawn from either a continuous or finite mixing distribution. In the former case, the integration in (9.48) is usually conducted using simulation techniques. In the latter case, also known as the latent class model, E-M algorithms are often used for estimation. E-M algorithms are iterative procedures that have been found to greatly simplify estimation in this context. See Train (2009) for additional discussion of these models and their estimation.

Finally, in recent years, considerable attention has been paid to concerns about omitted variable bias. In most applications, relatively few attributes of the sites are observed, such as fish catch rates (Morey, Rowe, and Watson 1993), fish toxin levels (Phaneuf, Kling, and Herriges 2000), or dummy variable indicators capturing different levels of water quality (Parsons, Helm, and Bondelid 2003), although there are some notable exceptions (Egan et al. 2009). The risk in this setting is that unobserved site attributes may be correlated with the observed attributes or travel costs (or both), leading to omitted variables bias for the estimated parameters, and biasing any subsequent welfare calculations. One solution to this problem, described in Murdock (2006), is to introduce alternative specific constants into the model, replacing (9.41) with

$$v_{ij} = \begin{cases} \delta_i \boldsymbol{S}_i & j = 0 \\ \alpha_j - \beta_M c_{ij} & j = 1, \ldots, \mathcal{J}, \end{cases} \tag{9.49}$$

where

$$\alpha_j = \gamma \boldsymbol{Q}_j + \xi_j \ , \tag{9.50}$$

and ξ_j denotes the unobserved attributes of site j. One can no longer include $\gamma \boldsymbol{Q}_j$ directly in (9.49), since it will be perfectly collinear with the alternative specific constants. Instead, in the first stage of estimation, the parameters $(\alpha_1, \ldots, \alpha_{\mathcal{J}}, \beta_M, \gamma)$ are estimated, typically via maximum likelihood. The γ's are recovered in a second stage by regressing the estimated alternative specific constants on observable site attributes using (9.50). The downside of this approach is that, while it avoids problems stemming from omitted variable bias, it complicates estimation, since there are now $\mathcal{J}$ alternative specific constants to estimate along with β and γ. This can be challenging when there are a large number of alternatives in the choice set. Murdock (2006) provided an iterative approach using a contraction mapping technique, which simplifies estimation when a logit model is being used. Abidoye, Herriges, and Tobias (2012) provided an alternative technique drawing on Bayesian simulation tools.

Multivariate Count Data Models

An alternative to the repeated RUM model is to generalize the single-site count data model to allow for a system of counts. There have been several efforts along these lines in the recreation demand literature, though they remain relatively rare. The paper by Ozuna and Gomez (1994) represents one of the earliest applications. The authors estimated a two-equation system of counts using what they refer to as the SUPREME model. Formally, they specified

$$\begin{aligned} x_{i1} &= x_{i1}^{*} + \omega_i \\ x_{i2} &= x_{i2}^{*} + \omega_i \end{aligned} \tag{9.51}$$

$x_{ij}^{*} \sim \text{Poisson}\left(\lambda_{ij}\right)$, where $\lambda_{ij} = \exp\left(\boldsymbol{Z}_{ij}\beta_{j}\right)$ takes the usual linear exponential form introduced for the single-site count data model, and $\omega_{i} \sim \text{Poisson}\left(\lambda_{\omega}\right)$, where λ_{ω} is a constant to be estimated. Since the sum of two Poisson distributions also follows a Poisson distribution, the x_{ij} follow Poisson distributions with conditional means equal to $\lambda_{ij} + \lambda_{\omega}$. Moreover, the counts are now correlated through the shared ω_{i}. The advantage of the SUPREME model is that it allows for correlated counts. On the other hand, the correlation is restricted to be positive, which may not always be the case, and it has generally been limited to two-equation systems, which restricts its usefulness in practice.

Englin, Boxall, and Watson (1998) took another tack. They estimated a system of four count equations in which each count follows a Poisson distribution, with a linear exponential mean, starting with

$$E\left(x_{ij}\right) = \exp\left(\alpha_{j} + \sum_{k=1}^{J} \beta_{jk} c_{ik} + \gamma_{j} M_{i}\right). \tag{9.52}$$

They showed that imposing integrability restrictions onto this demand system requires that there be (a) no cross-price effects (i.e., $\beta_{jk} = 0,\ j \neq k$) and (b) common income effect (i.e., $\gamma_{j} = \gamma,\ \forall j$). Furthermore, the intercepts are restricted to satisfy

$$\alpha_{j} = \alpha_{k}\left(\frac{\beta_{jj}}{\beta_{kk}}\right). \tag{9.53}$$

While this results in a utility theoretic system of Poisson demand equations, the use of independent Poisson distributions ignores potential correlation among the count variables.

In recent years, more general count systems have emerged, blending these two approaches. Shonkwiler (1999), for example, used the basic structure in Englin, Boxall, and Watson (1998), but introduced mixing distributions analogous to the mixed logit approach to allow for a general pattern of correlation among the four counts in his recreation demand system. Herriges, Phaneuf, and Tobias (2008) developed a related model, though they used Bayesian simulation tools to allow for a much larger system of counts, with 29 recreational sites.

The Kuhn–Tucker Model

A conceptual limitation of the system of counts approach to modeling recreation demand is that it is agnostic regarding the source of variation in the numbers of trips taken by households, even once one conditions on observable individual and site characteristics. Yet, as noted earlier, how one should treat a model's error terms in conducting welfare analysis depends upon the source of the error term itself (e.g., whether it is due to measurement error versus unobserved source of preference heterogeneity). Moreover, welfare analysis using count data models is generally couched in terms of consumer surplus, with the accompanying limitations of this welfare measure. While the repeated RUM framework avoids both of these problems, it is not without its own limitation. First, the model is based on an artificial construct (i.e., that individuals face a fixed number of choice occasions during which they choose among the available alternatives in the choice set). There is evidence in the literature that welfare estimates are not highly sensitive to the number of choice occasions used, though this research assumes that the number of choice occasions is the same for all individuals. Second, the repeated RUM model, as it has been employed in the literature, imposes weak complementarity, with the utility the individual receives from choosing an alternative (or set of alternatives) during the course of a season depending only on the attributes of the chosen sites.

An alternative modeling approach that avoids all of these issues is the corner solutions, or Kuhn–Tucker (KT), model originally developed by Wales and Woodland (1983) and Hanemann (1978), and first applied to recreation demand by Phaneuf et al. (2000). The KT model takes a direct approach for dealing with the frequent number of corner solutions observed in recreation demand applications by starting with the underlying consumer utility maximization problem subject to standard budget and non-negativity constraints. Corner solutions arise naturally due to the familiar Kuhn–Tucker conditions. The error terms are assumed to stem from unobservable factors altering consumer preferences, imbedded in the individual's direct utility function, and carry through to the implied demand equations. As such, the Kuhn–Tucker strategy provides a unified and internally consistent framework for characterizing the occurrence of corner solutions. Moreover, like the repeated RUM model, welfare analysis can be conducted on the basis of compensating variation.

The starting point in the KT model is the specification of the consumer's utility function. In particular, with J available recreational sites, it is assumed that individuals solve

$$\operatorname*{Max}_{\boldsymbol{X},z}\ u(z,\boldsymbol{X},\boldsymbol{Q},\varepsilon)\ , \tag{9.54}$$

subject to

$$\boldsymbol{P}'\boldsymbol{X} + z \leq M\ , \tag{9.55}$$

and

$$\boldsymbol{X} \geq \boldsymbol{0} \text{ and } z \geq 0\ , \tag{9.56}$$

where $u(z,\boldsymbol{X},\boldsymbol{Q},\boldsymbol{\varepsilon})$ is assumed to be a quasi-concave, increasing, and differentiable function of $(\boldsymbol{X},\boldsymbol{z})$, $\boldsymbol{X} = (x_1,\ldots,x_J)$ is the vector of trips taken to the J sites in the choice set, z is a numeraire good, $\boldsymbol{Q} = (q_1,\ldots,q_J)$ is a vector of site attributes, and $\boldsymbol{P} = (p_1,\ldots,p_J)$ is the vector of travel costs to the J sites.[2] The error term vector $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_J)$ captures unobserved factors that influence consumer preferences. These factors are assumed known by the decision-maker, but not the analyst, much

2 For simplicity, each site is characterized here using a single-site attribute, though the KT model more generally would allow for a vector of site attributes.

like in the RUM framework. The first-order necessary and sufficient conditions for utility maximization then become

$$
\begin{aligned}
& u_j(z,\boldsymbol{X},\boldsymbol{Q},\varepsilon) \equiv \frac{\partial u(z,\boldsymbol{X},\boldsymbol{Q},\varepsilon)}{\partial x_j} \leq \lambda p_j \quad j=1,\ldots,J \\
& x_j \geq 0 \\
& x_j\left[u_j(z,\boldsymbol{X},\boldsymbol{Q},\varepsilon)-\lambda p_j\right]=0
\end{aligned}
\tag{9.57}
$$

$$
\begin{aligned}
& u_z(z,\boldsymbol{X},\boldsymbol{Q},\varepsilon) \equiv \frac{\partial u(z,\boldsymbol{X},\boldsymbol{Q},\varepsilon)}{\partial z} \leq \lambda \quad j=1,\ldots,J \\
& z \geq 0 \\
& z\left[u_z(z,\boldsymbol{X},\boldsymbol{Q},\varepsilon)-\lambda\right]=0
\end{aligned}
\tag{9.58}
$$

and

$$
\begin{aligned}
& \boldsymbol{P}'\boldsymbol{X}+z \leq M \\
& \lambda \geq 0 \\
& \lambda[M-\boldsymbol{P}'\boldsymbol{X}-z]=0
\end{aligned}
\tag{9.59}
$$

where λ denotes the marginal utility of income.

Assuming that the numeraire good is essential, a realistic assumption in the context of modeling recreation demand, equation (9.58) implies that $\lambda = u_z(z,\boldsymbol{X},\boldsymbol{Q},\varepsilon)$ and $z = M-\boldsymbol{P}'\boldsymbol{X}$. Substituting these results into (9.57) yields a revised form of the first order conditions:

$$
\begin{aligned}
& \delta_j(M-\boldsymbol{P}'\boldsymbol{X},\boldsymbol{X},\boldsymbol{Q},\varepsilon) \equiv \frac{u_j(M-\boldsymbol{P}'\boldsymbol{X},\boldsymbol{X},\boldsymbol{Q},\varepsilon)}{u_z(M-\boldsymbol{P}'\boldsymbol{X},\boldsymbol{X},\boldsymbol{Q},\varepsilon)} \leq p_j \quad j=1,\ldots,J \\
& x_j \geq 0 \\
& x_j\left[\delta_j(M-\boldsymbol{P}'\boldsymbol{X},\boldsymbol{X},\boldsymbol{Q},\varepsilon)-p_j\right]=0.
\end{aligned}
\tag{9.60}
$$

The term $\delta_j(M-\boldsymbol{P}'\boldsymbol{X},\boldsymbol{X},\boldsymbol{Q},\varepsilon)$ denotes the individual's marginal willingness to pay (MWTP) for site *j*. If the individual chooses to visit site *j* ($x_j > 0$), they choose a level of usage that equates their MWTP to the corresponding travel cost. If, however, the associated travel cost exceeds their MWTP at any quantity level, then they choose to not visit the site (i.e., $x_j = 0$).

The final conceptual step in the KT model is in terms of the error terms. Specifically, it is assumed that each ε_j is uniquely related to a corresponding site, with

$$
\frac{\partial u_j}{\partial \varepsilon_k}\begin{cases} >0 & k=j \\ =0 & k\neq j \end{cases}
\tag{9.61}
$$

and $\partial u_z / \partial \varepsilon_j = 0\ \forall j$. This implies that both the marginal utilities and the marginal WTP for a site are strictly increasing functions of only one error term. In particular, $u_j\left(M - \boldsymbol{P}'\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{Q}, \boldsymbol{\varepsilon}\right) = \tilde{u}_j\left(M - \boldsymbol{P}'\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{Q}, \varepsilon_j\right)$ and $\delta_j\left(M - \boldsymbol{P}'\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{Q}, \boldsymbol{\varepsilon}\right) = \tilde{\delta}_j\left(M - \boldsymbol{P}'\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{Q}, \varepsilon_j\right)$. The benefit of this assumption is that one can now draw implications for the error terms in the model that can be used in estimation. Specifically, let $g_j = g_j\left(M - \boldsymbol{P}'\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{Q}\right)$ be defined implicitly as the level of the error term that would just induce the individual to want to start visiting site *j*; in other words,

$$\tilde{\delta}_j\left(M - \boldsymbol{P}'\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{Q}, g_j\right) = p_j. \tag{9.62}$$

Then the first order conditions in (9.60) imply that

$$\varepsilon_j \begin{cases} = g_j\left(M - \boldsymbol{P}'\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{Q}\right) & x_j > 0 \\ \leq g_j\left(M - \boldsymbol{P}'\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{Q}\right) & x_j = 0. \end{cases} \tag{9.63}$$

Thus, for sites that are visited, the observed consumption levels pin down the associated error terms (i.e., the unobservables), whereas for the sites not visited one can only bound the corresponding error terms. This information can in turn be used to derive the likelihood function required to estimate the parameters of the individual's utility function via maximum likelihood. Note that this is essentially what happens in the Tobit model described above for single-site models. The difference here is the structural underpinnings of the model.

Implementation of the KT model to date has been limited to relatively simple functional forms for the utility function $u\left(z, \boldsymbol{X}, \boldsymbol{Q}, \varepsilon\right)$ and distributional assumptions for the error terms (i.e., $\boldsymbol{\varepsilon}$). The most commonly used utility functions are variations on the linear expenditure system, with

$$u\left(z, \boldsymbol{X}, \boldsymbol{Q}, \boldsymbol{\varepsilon}\right) = \Psi_j \ln\left(x_j \omega_j + \theta_j\right) + \ln\left(z\right), \tag{9.64}$$

where $\Psi_j = \exp\left(\sum_{k=1}^{K} \upsilon_k q_{jk} + \varepsilon_j\right)$ and $\omega_j = \exp\left(\sum_{k=1}^{K} \zeta_k q_{jk}\right)$ potentially depend upon observable site characteristics (q_{jk}). The appeal of the linear expenditure system is that one can explicitly solve for the implied conditional demand equations and the functional form for the boundary variable $g_j\left(M - \boldsymbol{P}'\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{Q}\right)$. The error vector $\boldsymbol{\varepsilon} = \left(\varepsilon_1, \ldots, \varepsilon_J\right)$ is typically assumed to be either composed of iid extreme value terms or drawn from a GEV distribution (allowing for nesting of similar sites). Even with these assumptions, estimation and welfare analysis using the KT model can be challenging. Initial efforts (Phaneuf, Kling, and Herriges 2000) were limited to relatively few sites. Subsequently, von Haefen, Phaneuf, and Parsons (2004) developed estimation techniques allowing for much larger choice sets (i.e., in excess of 60), and generalizations of the linear expenditure system for use with the KT model (von Haefen 2007). Bhattacharya (2010) developed and implemented a panel version of the model. Even so, the KT model remains relatively rare in recreation demand models, despite the unified framework that it provides for modeling both the intensive and extensive margins of recreation demand. One of the limitations of the KT model, beyond its computational burden, is that it does not explicitly account for the count nature of recreation demand, but instead treats demand as a continuous variable.

Ongoing Issues

There are a myriad of ongoing issues in the modeling of recreation demand. This section of the chapter highlights some of the more important issues and describes the associated research efforts to date.

Defining the Choice Set

In describing the various recreation demand models above, it was assumed that the analyst knows the choice set being used by the individual. In practice, this is rarely the case. Instead, the researcher must identify the set of activities that represent reasonable substitutes for the site choices being modeled and which it is reasonable to assume that individuals are aware of. The issue of choice set was originally associated with RUM models, but it applies equally to all multiple-site approaches since the researcher must identify the potential set of substitutes (or complements) for inclusion, regardless of whether the approach taken is a Kuhn–Tucker model, a linked approach, or any other modeling strategy. Errors can emerge from the choice set being too large or too small. For example, suppose there is an excellent fishing site that relatively few people are aware of. If the site is included in everyone's choice set, an estimated recreation demand model might imply that people do not value high quality fishing, since few individuals would have visited the site, when the fact of the matter is that they simply did not know that the site was an option. Alternatively, since older residents are more likely to have learned of the site, the analysis might be used to infer that only older people value such fishing opportunities. As this example suggests, the choice set itself is likely to be individual specific and dynamic, evolving over time as the individual visits and learns about surrounding recreational opportunities. Analogous problems have been considered in the context of career and college choice and the marketing literatures. Treating the choice set as exogenous and using the universal choice set (e.g., all of the lakes in Iowa for a study of Iowa recreation), as is often done in practice, is at best an approximation whose realism will depend upon the application at hand.

At the other extreme, excluding alternatives from a choice set can create problems by understating an individual's ability to substitute away from an alternative when conditions deteriorate, consequently overstating the welfare impact of those changes. Exclusions can also limit the analyst's ability to identify important marginal effects. For example, Peters, Adamowicz, and Boxall (1995) allowed survey respondents to identify alternatives that they "considered" choosing, narrowing the choice set used in estimation to this so-called "consideration set." The risk here is that potentially valuable information is lost.

Sites may not be considered precisely because of their specific environmental conditions. Excluding them from the model will hamper the analyst's ability to identify which environmental characters are valued. Parsons, Massey, and Tomasi (1999) explored an alternative approach in which the choice set is restricted to sites that the individual is familiar with. While this avoids some of the concerns regarding the use of "consideration sets," the approach does require additional data-gathering efforts. Haab and Hicks (1997), on the other hand, applied a variation of Manski's (1977) model of choice set generation which treats the choice set itself as endogenously generated.

A practical issue in defining the choice set is that the choice set can become so large as to become unmanageable from an econometric perspective. Several approaches have been suggested to alleviate this problem. Parsons and Hauber (1998) proposed limiting the choice set for each individual to alternatives within a fixed distance of the person's home, arguing that alternatives beyond a reasonable distance are effectively irrelevant, particularly when day trips are being modeled.

An alternative to eliminating sites from the choice set is to aggregate alternatives. Kaoru and Smith (1990) were the first to analyze the effects of aggregation on preference parameter estimation and welfare measurement in the context of recreation demand. Their work suggested that models with only a mild degree of site aggregation (i.e., 35 sites aggregated to 23 or 11 composite sites) performed relatively well in capturing site selection. The results, however, were not as promising in terms of subsequent welfare calculations. For example, the welfare impact from the closure of an aggregate site was understated by more than a factor of 2 using either site aggregation. The welfare gain from site quality improvements fared even worse, being understated by a factor of 5 when 11 composite sites were used.

Parsons and Needelman's (1992) subsequent paper identified two distinct sources of bias stemming from site aggregation, one linked to the number of sites being aggregated (the so-called size effect) and the other tied to the degree of heterogeneity among the sites being combined. Their empirical results suggest that aggregation tends to lead to significant bias in parameter estimates (except for the travel cost coefficient), particularly when large numbers of sites are aggregated. The authors suggest minimizing heterogeneity of sites within aggregates and controlling for the number of sites in the aggregate groups by including a size measure as a factor in the conditional utility of each aggregate site.

A series of subsequent papers have largely confirmed the findings in Parsons and Needelman (1992). For example, Kaoru, Smith, and Liu (1995) found that parameters on the price and size variables are consistent across the different aggregation schemes, but that other parameters are sensitive to aggregation. Feather and Lupi (1998) considered partial aggregation as a compromise between a fully disaggregated model and a traditional aggregate model, with the idea being to leave as disaggregate the most popular and policy relevant sites, aggregating only those sites that are of little interest from a policy perspective or that are only infrequently used. As one might expect, this tends to mitigate any aggregation bias. Parsons, Plantinga, and Boyle (2000) suggested a similar partial aggregation approach, treating sites of policy interest and their close substitutes as individual sites and aggregating other sites to some degree.

The Value of Time

Both travel to a site and the recreation activity itself take time, and time is scarce. Also, the time spent in any activity could be a separate argument in individuals' utility functions. For example, time at work could yield either positive or negative utility; and almost certainly, time spent at a recreation site has a positive marginal utility, at least up to some point. The opportunity cost of time must be included in a properly specified model of the demand for recreation visits. The appropriate shadow price of time will also depend on the alternative uses to which the time could be put and on the nature of the constraints on individual choice.

It has long been recognized that the time spent in travel to a site should be included as a component of travel cost for purposes of estimating the demand for visits in a travel cost model (for example, Cesario and Knetsch 1970) or the probability of selecting a site in an RUM model. From equation (9.5), travel cost is defined as the sum of the monetary and time costs of travel to the site. If the time component of costs is omitted in the estimation of the demand function for visits, the cost variable is biased downward. The result will be that the estimated parameter on cost or price will be biased upward. As a consequence, the estimated demand curve will be more elastic than the true demand curve, and the benefits of the recreation site (the area under the demand curve) will be underestimated.

The choice of the shadow price of time is critical to the estimation of the elasticity of demand for the site and the calculation of the value of the site. The choice of a high shadow price of time raises the importance of time cost in explaining visits as a function of distance. With a higher shadow price of time, the predicted reduction in the number of visits because of an increase in the entry fee is smaller, and the estimated demand curve is less elastic. Clearly, though, using a shadow price of time that is too high can have the opposite effect on estimates of the elasticity of site demand and value; that is, underestimates of site demand and overestimates of site value.

One of the key assumptions in the simple travel cost model described above is that individuals are free to choose the hours they work at a given wage and that this wage governs the tradeoff between work and leisure. Given this assumption, the individual maximizes utility by allocating time among alternative activities so as to equate the marginal values of time in these activities with the wage rate. Thus, the wage can be taken as an indicator of the shadow value or marginal opportunity cost of time.

One practical difficulty in using the wage rate in recreation demand studies is that surveys often provide data on family income rather than the hourly wage rate. If the wage is inferred by dividing family income by some estimate of hours worked (typically, 2,000 hours per year), measurement error is introduced.

There is also the fact that there is empirical evidence from other types of behavior that is inconsistent with the standard model of the labor–leisure tradeoff at the market wage. Cesario (1976) was the first to point this out in the context of the analysis of recreation demand. He cited evidence from choices of transportation mode by commuters that the revealed opportunity cost of time was perhaps only one-third of the market wage. For more recent evidence from stated preference analysis of transportation mode choice, see Calfee, Winston, and Stempki (2001). Following Cesario, the practice of using a shadow price of time of one-third of the wage became common in recreation demand analysis (Shaw and Feather 1999). Englin and Shonkwiler (1995a) used a latent variable approach in an effort to estimate the unobserved true cost of travel time (including time cost) as a function of observable indicator variables. In their application of the model to data on recreation visits to Lake Champlain, they found that travel time has an estimated value that is about 40 percent of the average wage of the sample.

The evidence of a lower shadow price of time could be explained either by institutional constraints on the choice of hours of work, such as the standard 40-hour workweek and monthly or annual salaries or by a richer theory of choice that recognized that time spent in various activities including recreation, traveling, and at work might convey utility (positive or negative) directly. Each possibility is considered in turn.

Suppose that all (or most) jobs offer a 40-hour workweek at a fixed weekly salary. Then the individual cannot equate values and costs of leisure time at the margin. Several authors have analyzed models in which there is no tradeoff between work and recreation time. Smith, Desvousges, and McGivney (1983) assumed that individuals could allocate a fixed nonrecreational time budget between work and other activities. However, the recreation time budget was fixed and exogenous. Individuals could only choose to allocate time between two or more sites. They found that the relevant price of a visit to a site was the sum of the money travel cost and a proportion of the total time per visit, $t_1 + t_2$ in (9.3). This proportionality factor depended upon the wage rate and the shadow prices attached to the constraints on money and time. Thus, the proportionality factor was not observable, but it could be estimated from the data.

Bockstael, Strand, and Hanemann (1987) presented a model that was flexible in that it could accommodate either of two different cases regarding the individual's labor-market equilibrium. These authors assumed that the duration of a visit was fixed. If the individual could alter hours of work at the margin, the relevant price was the sum of the money travel cost and the total time of the visit $\left(t_1 + t_2\right)$ valued at the marginal wage rate. If the individual could not make marginal adjustments in the number of hours worked, the money travel cost and time cost per visit entered the demand function for visits separately. The parameter for the time cost component was a function of parameters of the preference function. Bockstael, Strand, and Hanemann assumed a specific functional form for preferences, derived the alternative demand functions, and estimated the parameters from a data set that distinguished between those who could and those who could not alter their work time at the margin.

Using a different model, Feather and Shaw (1999) used individuals' responses to survey questions to classify them according to whether they were able to choose freely the number of hours they worked at a given wage, were required to work more hours than they wished at the given wage (were "overemployed"), or would prefer to work more hours at the given wage (were "underemployed"). They were then able to estimate a shadow wage function for their sample and to compute shadow wages for the individuals in each group. As the standard model predicts, the shadow wage for those able to freely choose their hours of work was not significantly different from the mean observed wage, while for those who were overemployed (underemployed), the shadow wage was greater (less) than the observed wage. Since the three groups each represented about one-third of the working sample, this result does not support the practice of using a shadow price of time of one-third the wage for all recreationists.

Another line of research questions whether the wage rate, however measured, is the appropriate shadow price of time even for those people who can freely trade off time and work at the going wage. One of the key assumptions in the standard model, in which the wage rate is made the appropriate shadow price, is that there are no time variables appearing directly or implicitly as arguments in the utility function. Suppose that the time spent at work has either a positive or a negative utility at the margin, then there will be a divergence between the shadow price or scarcity value of time in other activities and the wage rate as a measure of the opportunity cost of time at work. This basic idea can be traced back to Johnson (1966). Cesario (1976) first analyzed its implications in the context of the travel cost model of recreation demand. To illustrate this point, the simple travel cost model described earlier in this chapter is modified to include time at work as an argument in the utility function.

The constrained maximization problem becomes

$$\max_{z,x} u(z,x) + \lambda\left(M - wt_w - z - p_x x\right) + \mu\left[t^* - t_w - \left(t_1 + t_2\right)x\right], \tag{9.65}$$

where site quality q has been omitted for simplicity. Since μ is the marginal utility of time, μ/λ is the scarcity value or marginal willingness to pay for time. The relevant first-order conditions include

$$\frac{\partial u/\partial x}{\lambda} = p_x + \frac{\mu}{\lambda}\left(t_1 + t_2\right) \tag{9.66}$$

and

$$\frac{\partial u/\partial t_w}{\lambda} + w = \frac{\mu}{\lambda}. \tag{9.67}$$

Equation (9.66) says that the marginal willingness to pay for a visit must equal the full cost of a visit. The full cost is the sum of the monetary cost p_x and the time cost of the visit, where the time cost $t_1 + t_2$ should be valued at the scarcity value of time, μ/λ. However, equation (9.67) shows that the wage rate may not be a good measure of the scarcity value of time. If the marginal utility of the time spent working is negative, then the wage rate is an overestimate of the scarcity value of time because the wage is also compensating for the disutility of work. In other words, the opportunity cost of diverting an hour of time away from work is less than the wage rate.

In contrast to this suggestion, McConnell and Strand (1981) derived an alternative model that allowed for the estimation of the relationship between the scarcity value of time and the wage rate from the sample data. Suppose the demand for visits can be specified as a linear function of travel cost:

$$x = a + b_1\left[p_x + (\mu/\lambda)(t_1 + t_2)\right]. \tag{9.68}$$

First, let s be the ratio of the shadow price of time to the wage rate. Substituting $s \cdot w$ for μ/λ in equation 9.68, one can then estimate

$$x = a + b_1 \cdot p_x + b_2 \cdot w \cdot (t_1 + t_2). \tag{9.69}$$

Since $b_2 w = b_1 s w$, then $s = b_2 / b_1$. For McConnell and Strand's sample, s was about 0.6.

In the McConnell and Strand model, although the scarcity value of time depends in part on the wage rate, it depends also on the alternative use of time, specifically on the marginal utility of work. A richer model of time and choice would allow for other utility-yielding uses of time. For example, there could be other leisure activities; or time could be an input into the household production of utility-yielding goods.

In a more recent series of papers, Larson and Shaikh (2001, 2004) and Shaikh and Larson (2003) emphasized the importance of explicitly incorporating both budget and time constraints when specifying the functional form for recreation demand equations. Rather than a single Roy's identity linking Marshallian demands to the underlying indirect utility function, they demonstrate that two Roy's identities apply, which in turn implicitly link money and time prices. As the authors noted, "the most common practice in empirical recreation demand specification, which includes a time price (as a component of full price) without a time budget (money income only) … is inconsistent with two-constraint choice" (Shaikh and Larson 2003, 954). In their accompanying application, Shaikh and Larson used a functional representation for the marginal value of time that nests a number of commonly used specifications. The authors rejected the most frequently used specifications for the value of time (including a fixed fraction of the wage rate) and argued instead for the flexibility of an endogenous marginal value of time approach. Clearly, additional empirical research is needed to corroborate these results, couched in the context of a single-site application, and to suggest the key factors in characterizing the marginal value of time.

Finally, in their closing remarks, Shaikh and Larson (2003) suggested that it may be necessary to allow for multiple time constraints when modeling recreation demand, reflecting the nature of the recreation activity. Palmquist, Phaneuf, and Smith (2010) pursued this line of inquiry, suggesting that the marginal value of time increases for longer trips, since it is more difficult to "cobble-together" the necessary block of time for such trips. The authors used stated preference survey data to infer the marginal opportunity cost of time and, as expected, found that it increases with duration (from $27.89 for a 2-hour trip to $32.15 for an 8-hour trip). While these changes are relatively small, they highlight the complex nature of the marginal value of time.

Time on Site

As the models described above have shown, time spent on the recreation site is part of the cost of the recreation visit. Therefore, in principle, time on site should be included in the estimated demand function for recreation visits. However, there are some assumptions under which this is not necessary. As McConnell (1985) has shown, if all individuals in a sample choose visits of the same duration and if they all have the same opportunity cost of time, time cost on site becomes part of the constant term in the estimated equation. The assumption of identical duration of visits may be reasonable in some cases, for example, where day-trip activities predominate, or if separate equations are estimated for visits of different duration. However, the assumption of identical opportunity costs of time is probably not reasonable, especially given that most travel cost studies now are based on observations of individuals rather than aggregates of people grouped by distance zone.

A related issue is how to model the choice of the duration of a visit. McConnell (1992) has investigated the implications of incorporating the choice of time on site for the proper specification of the recreation demand function. He specified the following utility function:

$$u = u(z, x, t_2), \tag{9.70}$$

and assumed that it displays what he called joint weak complementarity between x and t_2. By this he means that if $x = 0$, the marginal utility of on-site time is zero, and if on-site time is zero, the marginal utility of a trip is zero. The individual must choose x, t_2, and z to maximize equation (9.70), subject to the constraint of equation (9.4), or

$$M + w \cdot t^* = z + c \cdot x, \tag{9.71}$$

where $c = f + p_d \cdot d + w \cdot (t_1 + t_2)$. Since c is not exogenous to the choice problem, it cannot be an argument in a Marshallian demand function. McConnell showed that it is still possible to define a Marshallian demand function based on the exogenous components of trip cost. First, define a new term:

$$c^* = f + p_d \cdot d + c \cdot t_1. \tag{9.72}$$

After maximizing equation (9.70) and solving for the demand functions, the indirect utility function can be obtained:

$$v = v\left(c^*, w, M\right). \tag{9.73}$$

McConnell then showed that Roy's identity can be employed to derive a Marshallian demand function that contains only exogenous variables:

$$x = x\left(c^*, w, M\right). \tag{9.74}$$

This expression can be estimated. For other efforts at modeling the determinants of time on site, see Smith, Desvousges, and McGivney (1983), Kealy and Bishop (1986), and Wilman (1987).

Sampling Issues

A practical problem that researchers frequently encounter in recreation demand studies is that only a small fraction of the population may visit a site of interest. In this setting, a random population survey can be expensive, requiring a relatively large sample in order to yield enough users to accurately estimate the parameters of interest. Screening surveys can be used to identify visitors to a site, who can then be followed up with a more detailed survey regarding their usage, but this approach can also be expensive and time consuming. An alternative procedure is to intercept users at the site in question, guaranteeing information regarding site users. Unfortunately, the process creates two problems: (a) truncation (with only users being surveyed) and (b) endogenous stratification (with avid users more likely to be intercepted on site). Both problems will tend to inflate the numbers of trips obtained from an on-site sample relative to those in the general population.

A number of authors have developed techniques designed to "undo" the effects of on-site sampling. Shaw (1988) developed a simple correction for both truncation and endogenous stratification in the case of a single-site Poisson count data model. Englin and Shonkwiler (1995b) subsequently extended Shaw's correction to the case of the Negative Binomial model. Egan and Herriges (2006) extended this approach to a multivariate count data model setting. The risk with all of these procedures is that they rely heavily on maintained distributional assumptions, particularly in terms of inferring the behavior and preferences of nonusers. This suggests that if an on-site sample is employed, efforts should be made to gather a representative sample of nonusers, which can be used to augment an on-site sample and discern any departures from maintained assumptions.

Choice Dynamics and General Equilibrium Concerns

The models of recreation demand that dominate the literature are largely static in nature, assuming a fixed (and often exogenous) choice set and few, if any, interactions among the trips taken by an individual over time. The need for this simplification is driven to a large extent by the nature of the data available to researchers, consisting primarily of total trip counts during a season. There are, however, obvious dynamic elements that are being abstracted from. As already noted, the choice set itself is surely dynamic in nature, evolving over time as individuals explore the available recreational opportunities and, in doing so, learn about sites and their attributes. This is a topic that has been explored in other settings, such as in the job and college search literatures, but has yet to be explored in the context of recreation demand. Perhaps the closest application along these lines is the paper by Provencher and Bishop (1997), in which the authors developed a dynamic discrete choice model using diary data. Although in their application the choice set is given, individuals learn about catch rates at individual sites through trips to the site.

A second dynamic element in recreation demand is the potential linkage between current and past trips, a phenomenon known as "state dependence." The notion is that an individual's past trips may influence their inclination to revisit the same sites in the future. "Habit formation," for example, might arise if the individual finds comfort in the familiar. Once they visit a given site, their inclination would be to return to the same site on subsequent choice occasions. Other individuals might exhibit "variety seeking" and would, ceteris paribus, avoid soon visiting the same site again. Moeltner and Englin (2004) provided one of the few applications examining the impact of both state dependence and time varying site attributes on recreation demand, with an application to the choice of ski resorts. Once again, the availability of diary data is key to their analysis. Interestingly, they found that, while habit formation tends to dominate variety seeking as the form of state dependence, time varying attributes (such as weather conditions) are also critical to predicting individual choice. Neither factor is controlled for in traditional repeated RUM models.

Finally, most recreation demand models treat site attributes as exogenous factors to which individuals respond. There is, however, a growing literature recognizing the fact that some site attributes are endogenously determined. One such attribute is site congestion. Congestion of a recreation site occurs when the number of users is so large that it diminishes the utility and therefore the willingness to pay of those users. The degree of crowding at a site can be considered as one of the site attributes that influence the quality of the recreation services each individual experiences and the utility each individual obtains from visiting the site. The presence of congestion at a recreation site has implications for the estimation of recreation demand and the measurement of the value of environmental quality change.

Congestion, like pollution, is a negative externality. It affects all users of a site. The earliest analyses of the problem of congested recreation facilities include Fisher and Krutilla (1972) and Cicchetti and Smith (1973). Freeman and Haveman (1977) developed a simple model of congestion to show the effects of congestion on demand for a site and on the prediction of recreation use. The model illustrates how the value of improving quality at a site is reduced if congestion is present, and how improving quality at one site can lead to additional benefits by reducing congestion at other sites. Other theoretical analyses can be found in McConnell and Duff (1976), McConnell (1980), and Smith (1981). Dorfman (1984) provided an elegant treatment of the optimal congestion and optimal pricing problems.

More recently, Timmins and Murdock (2007) drew on locational sorting models to estimate the disamenity impact of site congestion. Their analysis clearly showed that not only is congestion an important site characteristic, but also the failure to recognize its endogeneity can lead to significant bias in assessing both the marginal impact of congestion and the general equilibrium implications of policy scenarios. Papers by Schuhmann and Schwabe (2004) and Hicks, Horrace, and Schnier (forthcoming) have generalized this model, allowing for both heterogeneity in how individuals perceive the impact of congestion, and nonlinearities in the marginal impact of congestion. Phaneuf, Carbone, and Herriges (2009) developed a similar model using the Kuhn–Tucker framework.

Summary and Conclusions

The recreation demand literature has evolved substantially over the last 40 years. While much of the early literature was segmented, with one branch focused on modeling the total numbers of recreational trips taken (i.e., the participation decision) and the other concerned with where individuals recreated (i.e., the site selection literature), this distinction has largely disappeared. Recreation models, whether based on RUM, counts, or continuous demand specifications, are largely couched as systems of demand, controlling for substitution possibilities among the myriad of alternatives in the individual's choice set and allowing for frequent corner solutions. What distinguishes the models is which aspect of the consumer's optimization problem the model abstracts away from (and, hence, which part of the "elephant" it chooses to focus on). The repeated RUM model assumes a fixed and common number of choice occasions faced by decision-makers, whereas the Kuhn–Tucker model ignores the count nature of the data. Systems of counts lack an underlying utility maximization structure that consistently integrates the random nature of the counts themselves. Moreover, all of these models are constrained by the nature of the data typically available to analysts (i.e., seasonal counts of total trips to each of the sites in the choice set). This limits the ability of the analyst to allow for both temporal effects (e.g., due to weather) and inter-temporal effects (e.g., such as learning, habit formation, and variety seeking). While diary data would allow for more complex models, collecting such data faces significant challenges (both in terms of nonparticipation and attrition); and modeling the dynamics of individual behavior in these settings often requires strong assumptions (in terms of what the individual knows about their choices and when) in order to make the analysis tractable.

Despite the limitations of the various recreation demand models, they have become a mainstay in nonmarket valuation, informing both policymakers and natural resource damage assessment efforts. Their strength comes from the rich variation in travel costs across individuals, as well as their reliance on revealed preference. At the same time, ongoing research is clearly needed, particularly in terms of providing a richer characterization of the value of time and the choice set being used by the consumer. There is also a need to better understand both the choice of time on site (e.g., day trips versus weeklong vacations) and the impact of multiple-destination trips.

References

Abidoye, Babatunde O., Joseph A. Herriges, and Justin L. Tobias. 2012. "Controlling for Observed and Unobserved Site Characteristics in RUM Models of Recreation Demand." *American Journal of Agricultural Economics* 94(5): 1070–1093.

Bhattacharya, Subhra. 2010. "Three Essays in Economics of the Environment." Graduate Theses and Dissertations. Paper 11299. Available at: http://lib.dr.iastate.edu/etd/11299

Bockstael, Nancy E., and Ivar Strand. 1987. "The Effects of Common Sources of Regression Error on Benefit Estimates." *Land Economics* 63(1): 11–20.

Bockstael, Nancy E., W. Michael Hanemann, and Catherine L. Kling. 1987. "Estimating the Value of Water Quality Improvements in a Recreational Demand Framework." *Water Resources Research* 23(5): 951–960.

Bockstael, Nancy E., Ivar E. Strand, and W. Michael Hanemann. 1987. "Time and the Recreational Demand Model." *American Journal of Agricultural Economics* 69(2): 293–302.

Bockstael, Nancy E., Ivar E. Strand Jr., Kenneth E. McConnell, and Firuzeh Arsanjani. 1990. "Sample Selection Bias in the Estimation of Recreation Demand Functions: An Application to Sportfishing." *Land Economics* 66(1): 40–49.

Calfee, John, Clifford Winston, and Randolph Stempki. 2001. "Econometric Issues in Estimating Consumer Preferences from Stated Preference Data: A Case Study of the Value of Automobile Travel Time." *Review of Economics and Statistics* 83(4): 699–707.

Cameron, A. Colin, and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Caulkins, Peter P., Richard C. Bishop, and Nicolaas W. Bouwes Sr. 1985. "Omitted Cross-Price Variables in the Linear Travel Cost Model: Correcting Common Misperceptions." *Land Economics* 61(2): 182–187.

Caulkins, Peter P., Richard C. Bishop, and Nicolaas W. Bouwes Sr. 1986. "The Travel Cost Model for Lake Recreation: A Comparison of Two Methods for Incorporating Site Quality and Substitution Effects." *American Journal of Agricultural Economics* 68(2): 291–297.

Cesario, Frank J. 1976. "Value of Time and Recreation Benefit Studies." *Land Economics* 52(1): 32–41.

Cesario, Frank J., and Jack L. Knetsch. 1970. "Time Bias in Recreation Benefit Estimates." *Water Resources Research* 6(3): 700–704.

Cicchetti, Charles J., and V. Kerry Smith. 1973. "Congestion, Quality Deterioration and Optimal Use: Wilderness Recreation in the Spanish Peaks Primitive Area." *Social Science Research* 2(1):15–30.

Clawson, Marion. 1959. *Methods of Measuring the Demand for and Value of Outdoor Recreation*. Washington D.C.: Resources of the Future Reprint No. 10.

Clawson, Marion, and Jack L. Knetsch. 1966. *The Economics of Outdoor Recreation*. Baltimore: Johns Hopkins University Press.

Dorfman, Robert. 1984. "On Optimal Congestion." *Journal of Environmental Economics and Management* 11(2): 91–106.

Egan, Kevin, and Joseph A. Herriges. 2006. "Multivariate Count Data Regression Models with Individual Panel Data from On-Site Sampling." *Journal of Environmental Economics and Management* 52(2): 567–581.

Egan, Kevin, Joseph A. Herriges, Catherine L. Kling, and John Downing. 2009. "Recreation Demand Using Physical Measures of Water Quality." *American Journal of Agricultural Economics* 91:106–123.

Englin, Jeffrey, and J.S. Shonkwiler. 1995a. "Modeling Recreation Demand in the Presence of Unobservable Travel Costs: Toward a Travel Price Model." *Journal of Environmental Economics and Management* 29(3): 368–377.

Englin, Jeffrey, and J.S. Shonkwiler. 1995b. "Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand under Conditions of Endogenous Stratification and Truncation." *Review of Economics and Statistics* 77(1): 104–112.

Englin, Jeffrey, Peter Boxall, and David Watson. 1998. "Modeling Recreation Demand in a Poisson System of Equations: An Analysis of the Impact of International Exchange Rates." *American Journal of Agricultural Economics* 80(2): 255–263.

Feather, Peter, Daniel Hellerstein, and T. Tomasi. 1995. "A Discrete-Count Data Model of Recreation Demand." *Journal of Environmental Economic and Management* 29(2): 316–322.

Feather, Peter, and Frank Lupi. 1998. "Using Partial Aggregation to Reduce Bias in Random Utility Travel Cost Models." *Water Resources Research* 34(12): 3595–3603.

Feather, Peter, and W. Douglass Shaw. 1999. "Estimating the Cost of Leisure Time for Recreation Demand Models." *Journal of Environmental Economics and Management* 38(1): 49–65.

Feenberg, Daniel, and Edwin S. Mills. 1980. *Measuring the Benefits of Water Pollution Abatement.* New York: Academic Press.

Fisher, Anthony C., and John V. Krutilla. 1972. "Determination of Optimal Capacity of Resource-based Recreation Facilities." *Natural Resources Journal* 12(3): 417–444.

Freeman, A. Myrick, III. 2003. *The Measurement of Environmental and Resource Values: Theory and Methods.* 2d. ed. Washington, D.C.: Resources for the Future Press.

Freeman, A. Myrick, III, and Robert H. Haveman. 1977. "Congestion, Quality Deterioration, and Heterogeneous Tastes." *Journal of Public Economics* 8(2): 225–232.

Haab, Timothy C., and Robert L. Hicks. 1997. "Accounting for Choice Set Endogeneity in Random Utility Models of Site Choice." *Journal of Environmental Economics and Management* 34(2): 127–147.

Haab, Timothy C., and Kenneth E. McConnell. 1996. "Count Data Models and the Problems of Zeros in Recreation Demand Analysis." *American Journal of Agricultural Economics* 78(1): 89–102.

Haab, Timothy C., and Kenneth E. McConnell. 2002. *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation.* Cheltenham: Edward Elgar.

Hanemann, W. Michael. 1978. "A Methodological and Empirical Study of the Recreation Benefits from Water Quality Improvement." Ph.D. Dissertation, Harvard University, Cambridge, MA.

Hausman, Jerry A., Gregory K. Leonard, and Daniel McFadden. 1995. "A Utility-Consistent, Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses Due to Natural Resource Damage." *Journal of Public Economics* 56(1): 1–30.

Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5: 475–492.

Hellerstein, Daniel, and Robert Mendelsohn. 1993. "A Theoretical Foundation for Count Data Models." *American Journal of Agricultural Economics*. 75(3): 604–611.
Herriges, Joseph A., and Daniel J. Phaneuf. 2002. "Inducing Patterns Correlation and Substitution in Repeated Logit Model of Recreation Demand." *American Journal of Agricultural Economics* 84(4): 1076–1090.
Herriges, Joseph A., Daniel J. Phaneuf, and Justin L.Tobias. 2008. "Estimating Demand Systems when Outcomes are Correlated Count." *Journal of Econometrics* 147(2): 282–298.
Hicks, Robert L., William Horrace, and Kurt Schnier. Forthcoming. "Strategic Substitutes or Complements: The Game of Where to Fish?" *Journal of Econometrics*.
Hotelling, Harold. 1947. "Letter to the National Park Service." Reprinted in *An Economic Study of the Monetary Evaluation of Recreation in the National Park Service* (1949). US Department of the Interior, National Park Service and Recreation Planning Division, Washington, D.C.
Johnson, M. Bruce. 1966. "Travel Time and the Price of Leisure." *Western Economics Journal* 4(1): 135–145.
Kaoru, Yoshiaki, and V. Kerry Smith. 1990. " 'Black Mayonnaise' and Marine Recreation: Methodological Issues in Valuing a Cleanup." Marine Policy Center, Woods Hole Oceanographic Institution, Woods Hole, MA.
Kaoru, Yoshiaki, V. Kerry Smith, and Jin Long Liu. 1995. "Using Random Utility Models to Estimate the Value of Estuarine Resources." *American Journal of Agricultural Economics* 77(1): 141–151.
Kealy, Mary Jo, and Richard C. Bishop. 1986. "Theoretical and Empirical Specification Issues and Travel Cost Demand Studies." *American Journal of Agricultural Economics* 68(3): 660–667.
Kling, Catherine L. 1989. "A Note on the Welfare Effects of Omitting Substitute Prices and Qualities from Travel Cost Models." *Land Economics* 65: 290–296.
Kling, Catherine L., and Cynthia Thomson. 1996. "The Implications of Model Specification for Welfare Estimation in Nested Logit Models." *American Journal of Agricultural Economics* 78: 103–114.
Larson, Douglas M., and Sabina L. Shaikh. 2001. "Empirical Specification and Considerations for Two-Constraint Models of Recreation Demand." *American Journal of Agricultural Economics* 83(2): 428–440.
Larson, Douglas M., and Sabina L. Shaikh. 2004. "Recreation Demand Choices and Revealed Values of Leisure Time." *Economic Inquiry* 42(2): 264–278.
Manski, Charles F. 1977. "The Structure of Random Utility Models." *Theory and Decision* 8(3): 229–254.
McConnell, Kenneth E. 1980. "Valuing Congested Recreation Sites." *Journal of Environmental Economics and Management* 7(4): 389–394.
McConnell, Kenneth E. 1985. "The Economics of Outdoor Recreation." In *Handbook of Natural Resource and Energy Economics, vol. 1*, edited by Allen V. Kneese and James L. Sweeney. Amsterdam: North Holland.
McConnell, Kenneth E. 1992. "On-Site Time in the Demand for Recreation." *American Journal of Agricultural Economics* 74(4): 918–925.
McConnell, Kenneth E., and Virginia A. Duff. 1976. "Estimating Net Benefits of Outdoor Recreation Under Conditions of Excess Demand." *Journal of Environmental Economics and Management* 2(3): 224–230.
McConnell, Kenneth E., and Ivar Strand. 1981. "Measuring the Cost of Time in Recreation Demand Analysis: An Application to Sport Fishing." *American Journal of Agricultural Economics* 63(1): 153–156.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by Paul Zarembka, New York: Amsterdam, 105–142.

McFadden, Daniel. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Choice Data with Econometric Applications*, edited by Charles F. Manski and Daniel McFadden, Cambridge: MIT Press, 198–272.

McFadden, Daniel. 1984. "Econometric Analysis of Qualitative Response Models." In *Handbook of Econometrics, vol. 2.*, edited by Zvi. Griliches and Michael D. Intriligator. Amsterdam: North Holland, 1395–1457.

McKean, John R., and Charles F. Revier. 1990. "An Extension of: 'Omitted Cross-Price Variable Biases in the Linear Travel Cost Model: Correcting Common Misperceptions'." *Land Economics* 66(4): 431–436.

Moeltner, Klaus, and Jeffrey Englin. 2004. "Choice Behavior under Time-Variant Quality: State Dependence versus 'Playing-it-by-Ear' in Selecting Ski Resorts." *Journal of Business & Economic Statistics* 22(2): 214–224.

Morey, Edward R. 1999. "Two RUMs Uncloaked: Nested-Logit Models of Site Choice and Nested-Logit Models of Participation and Site Choice." In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice,* edited by Joseph A. Herriges and Catherine L. Kling. Cheltenham: Edward Elgar.

Morey, Edward R., Robert D. Rowe, and Michael Watson. 1993. "A Repeated Nested-Logit Model of Atlantic Salmon Fishing." *American Journal of Agricultural Economics* 75(3): 578–592.

Murdock, Jennifer. 2006. "Handling Unobserved Site Characteristics in Random Utility Models of Recreation Demand." *Journal of Environmental Economics and Management,* 51(1): 1–25.

Ozuna, Teofilo, and Irma Adriana Gomez. 1994. "Estimating a System of Recreation Demand Functions Using a Seemingly Unrelated Poisson Regression Approach." *Review of Economics and Statistics*. 76(2): 356–360.

Palmquist, Raymond B, Daniel J. Phaneuf, and V. Kerry Smith. 2010. "Short Run Constraints and the Increasing Marginal Value of Time in Recreation." *Environmental & Resource Economics* 46(1): 19–41.

Parsons, George R. 1991. "A Note on Choice of Residential Location in Travel Cost Demand Models." *Land Economics* 67(3): 360–364.

Parsons, George R. 2003. "The Travel Cost Model." In *A Primer on Nonmarket Valuation*, edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Dordrecht: Kluwer Academic Press.

Parsons, George R., and A. Brett Hauber. 1998. "Choice Set Boundaries in a Random Utility Model of Recreation Demand." *Land Economics* 74(1): 32–48.

Parsons, George R., and Mary Jo Kealy. 1995. "A Demand Theory for the Number of Trips in a Random Utility Model of Recreation." *Journal of Environmental Economics and Management* 29(3): 357–367.

Parsons, George, and Michael S. Needelman. 1992. "Site Aggregation in a Random Utility Model of Recreation." *Land Economics* 68(4): 418–433.

Parsons, George R., Eric C. Helm, and Tim Bondelid. 2003. "Measuring the Economic Benefits of Water Quality Improvements to Recreational Users in Six Northeastern States: An Application to the Random Utility Maximization Model." Working Paper, University of Delaware, July.

Parsons, George R., Paul M. Jakus, and Ted Tomasi. 1999. "A Comparison of Welfare Estimates from Four Models for Linking Seasonal Recreational Trips to Multinomial

Logit Models of Site Choice." *Journal of Environmental Economics and Management* 38(2): 143–157.

Parsons, George, D. Matthew Massey, and Ted Tomasi. 1999. "Familiar and Favorite Sites in a Random Utility Model of Beach Recreation." *Marine Resource Economics* 14(4): 299–315.

Parsons, George R, Andrew J. Plantinga, and Kevin J. Boyle. 2000. "Narrow Choice Sets in a Random Utility Model of Recreation Demand." *Land Economics* 76(1): 86–99.

Peters, T., Wictor Adamowicz, and Peter Boxall. 1995. "The Influence of Choice Set Consideration in Modeling the Benefits of Improved Water Quality." *Water Resources Research* 31(7): 1781–1787.

Phaneuf, Daniel J., and V. Kerry Smith. 2005. "Recreation Demand Models." In *Handbook of Environmental Economics*, edited by Karl-Göran Mäler and Jeffery R. Vincent. Amsterdam: Elsevier.

Phaneuf, Daniel J., Jared C. Carbone, and Joseph A. Herriges. 2009. "Non-Price Equilibria for Non-Marketed Goods." *Journal of Environmental Economics and Management* 57(1): 45–64.

Phaneuf, Daniel J., Catherine L. Kling, and Joseph A. Herriges. 2000. "Estimation and Welfare Calculations in a Generalized Corner Solution Model with an Application to Recreation Demand." *Review of Economics and Statistics* 82(1): 83–92.

Provencher, William, and Richard C. Bishop. 1997. "An Estimable Dynamic Model of Recreation Behavior with an Application to Great Lakes Angling." *Journal of Environmental Economics and Management* 33(2): 107–127.

Randall, Alan. 1994. "A Difficulty with the Travel Cost Method." *Land Economics* 70(1): 88–96.

Schuhmann, Peter W, and Kurt A. Schwabe 2004. "An Analysis of Congestions Measures and Heterogeneous Angler Preferences in a Random Utility Model of Recreational Fishing." *Environmental & Resource Economics* 27(4): 429–450.

Shaikh, Sabina L., and Douglas M. Larson. 2003. "A Two-Constraint Almost Ideal Demand Model of Recreation and Donations." *Review of Economics and Statistics* 85(4): 953–961.

Shaw, Daigee. 1988. "On-site Samples' Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Stratification." *Journal of Econometrics* 37(2): 211–223.

Shaw, W. Douglass, and Peter Feather. 1999. "Possibilities for Including the Opportunity Cost of Time in Recreation Demand Models." *Land Economics* 75(4): 592–602.

Shaw W. Douglass, and Michael T. Ozog. 1999. "Modeling Overnight Recreation Trip Choice: Application of a Repeated Nested Multinomial Logit Model." *Environmental and Resource Economics* 13(4): 397–414.

Shonkwiler, J.S. 1999. "Recreation Demand Systems for Multiple Site Count Data Travel Cost Models." In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*¸ edited by Joseph A. Herriges and Catherine L. Kling. Cheltenham: Edward Elgar.

Smith, V. Kerry. 1981. "Congestion, Travel Cost Recreational Demand Models, and Benefit Evaluation." *Journal of Environmental Economics and Management* 8(1): 92–96.

Smith, V. Kerry. 1988. "Selection and Recreation Demand." *American Journal of Agricultural Economics*. 70(1): 29–36.

Smith, V. Kerry. 1997. "Combining Discrete Choice and Count Data Models: A Comment." Mimeograph.

Smith, V. Kerry, William H. Desvousges, and Matthew P. McGivney. 1983. "The Opportunity Cost of Travel Time in Recreation Demand Models." *Land Economics* 59(3): 259–278.

Timmins, Christopher, and Jennifer Murdock. 2007. "A Revealed Preference Approach to the Measurement of Congestion in Travel Cost Models." *Journal of Environmental Economics and Management* 53(2): 230–249.

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation.* 2d ed. Cambridge: Cambridge University Press.

Von Haefen, Roger H. 2007. "Empirical Strategies for Incorporating Weak Complementarity into Consumer Demand Models." *Journal of Environmental Economics and Management* 54(1): 15–31.

Von Haefen, Roger H., Daniel J. Phaneuf, and George R. Parsons. 2004. "Estimation and Welfare Calculation with Large Demand Systems." *Journal of Business and Economic Statistics* 22(2): 194–205.

Wales, T.J., and A.D. Woodland. 1983. "Estimation of Consumer Demand with Binding Nonnegativity Constraints." *Journal of Econometrics* 21(3): 263–285.

Wilman, Elizabeth A. 1987. "A Simple Repackaging Model of Recreational Choices." *American Journal of Agricultural Economics* 69(3): 604–612.

Chapter 10

Property Value Models

Economists have documented the relationship between the prices of housing units and quantities of environmental amenities since before this relationship was recognized as an application of the theory of hedonic prices (for example, Ridker and Henning 1967). Indeed, examples of the statistical analysis of the linkage between farmland prices and the characteristics of the land can be found as early as 1922. See Colwell and Dilmore (1999) for a review. The past 35 years have seen an explosion of both theoretical and empirical studies of the monetary values of nonmarket amenities and disamenities based on hedonic price theory. It is now well accepted that housing price differentials do reflect differences in the quantities of various characteristics of housing and that these differentials have significance for applied welfare analysis. For example, Smith and Huang (1995) conducted a meta-analysis of hedonic studies of air pollution and housing prices. They reported finding 37 studies and more than 160 separate estimates of the effects of air quality on housing prices.

At the same time, much has also been learned about the limitations of standard hedonic property analysis. As will be discussed later, the bulk of the empirical literature has focused on estimating the hedonic price function itself (that is, the equilibrium relationship between a home's prices and its characteristics). This "first stage" in the hedonic model can be used, under certain conditions, to infer the *marginal* willingness-to-pay (MWTP) for individual housing characteristics, including environmental amenities. While such information is valuable, most environmental policy scenarios envision discrete changes for which marginal analysis is no longer sufficient. Hedonic analysis can be extended to value discrete changes, but such "second stage" hedonic analyses are rare in practice, requiring complex econometrics and/or strong a priori restrictions on preferences (see Palmquist 2005).

In part, responding to this limitation of traditional hedonic price analysis, the "equilibrium sorting" literature has emerged in the past decade as a competing paradigm for characterizing how amenities are capitalized into the value of a property. Rather than focusing on a largely reduced form characterization of the hedonic price function, the sorting literature emphasizes the sorting process itself; that is, how the distribution of both individual characteristics and preferences interact with the distribution of housing characteristics in a market to determine market prices. Having characterized the sorting process, an analyst can, in theory, examine both the partial and general equilibrium implications of a discrete change in the distribution of property amenities. Smith et al. (2004), for example, estimated the impact of the projected ozone reductions in the Los Angeles air basin stemming from the 1990 Clean Air Act Amendments. The authors found that estimates of these impacts, and their distribution in the population, changed substantively depending upon whether or not individuals are allowed to re-sort (that is, move) and property values are allowed to change in response to the discrete air quality improvements. Such changes are indicative of the long run, general equilibrium impact of an environmental policy. The sorting literature, of course, is not without its limitations, requiring strong assumptions regarding the structure and distribution of preferences.

This chapter provides a summary of the current state of knowledge regarding these two approaches to property value modeling: hedonic price and equilibrium sorting. The first section focuses on hedonic pricing, beginning with a brief review of the evolution of economic thinking about property prices and environmental amenities, followed by a detailed exposition of the hedonic property value model. This presentation includes discussions of: (a) problems in estimating the hedonic price function; (b) approaches to recovering information on preferences and the demands for characteristics from the hedonic price function; and (c) the measurement of welfare change. The second section turns to the more recent equilibrium sorting literature. The distinction is made between what Kuminoff, Smith, and Timmins (2013) refer to as the "pure characteristics" sorting models of Epple and Sieg (1999) (among others) and the random utility sorting models pioneered by Bayer, McMillan, and Reuben (2004).

Hedonic Pricing

Historical Background

The theory of rents holds that the equilibrium price for a parcel of land will be the present value of the stream of rents produced by the land. Economic theory has long recognized that the productivity of land differs across sites. These productivity differentials will yield differential rents to land, and therefore differential land values. Where land is a producer's good, competition and free entry are sufficient to ensure that productivity differentials are fully reflected in the land rent structure. For any property where the land rent is less than the productivity, the activity occupying that land must be earning a profit. Some potential entrant will be willing to bid above the going rent in order to occupy that site and reap the rewards of a superior productivity. It is this competition that bids up land rents and eliminates the profit. Rent differentials will be equal to productivity differentials; and since the price at which a unit of land sells in the market is the present value of the stream of future rents, productivity differentials will also be reflected in land prices.

Some environmental characteristics such as air or water quality may affect the productivity of land as either a producer's good or a consumer's good. Where this is so, the structure of land rents and prices will reflect these environmentally determined productivity differentials. These results from classical rent theory aroused considerable interest among economists about the possibility of using data on land rent or land value for residential properties to measure the benefits to households brought about by improvements in environmental characteristics such as air or water quality. Ridker (1967) was the first economist to attempt to use residential property value data as the basis for estimating the benefits of changes in measures of environmental quality such as air pollution. He reasoned as follows:

> If the land market were to work perfectly, the price of a plot of land would equal the sum of the present discounted streams of benefits and costs derivable from it. If some of its costs rise (for example, if additional maintenance and cleaning costs are required) or if some of its benefits fall (for example, if one cannot see the mountains from the terrace), the property will be discounted in the market to reflect people's evaluation of these changes. Since air pollution is specific to locations and the supply of location is fixed, there is less likelihood that the negative effects of pollution can be significantly shifted on to other markets. We should therefore expect to find the majority of effects reflected in this market, and we can measure them by observing associated changes in property values.
>
> (Ridker 1967, 25)

The last sentence of the passage raises three questions. The first is whether environmental variables such as air pollution do systematically affect land prices. Assuming an affirmative answer to this question, the second is whether knowledge of this relationship is sufficient to predict changes in land prices when, say, air pollution levels change. The third question is whether changes in land prices accurately measure the underlying welfare changes.

Ridker (1967) and Ridker and Henning (1967) provided the first empirical evidence that air pollution affects property values by regressing median census tract property values in an urban area on a measure of sulfate air pollution. They then asserted positive answers to the second and third questions. Specifically, they argued that the coefficient on the air pollution variable in the regression equation could be used to predict the change in the price of any residence, conditioned on a change in its air pollution level. The sum of all such changes, they argued, could be taken as a measure of the benefit of improving air quality in an urban area (Ridker 1967, 136–137; Ridker and Henning 1967, 254).

This work stimulated a now large literature on the proper theoretical interpretation of the observed air pollution–property value relationship. Early contributions included those by Freeman (1971, 1974a, 1974b), and Anderson and Crocker (1972). Subsequent efforts to provide a sound theoretical basis for interpreting the air pollution–property value relationship have taken one of two paths. The first has been the development of models of the urban land market to determine whether and under what circumstances changes in aggregate land values accurately measure the benefits associated with environmental improvements. Early efforts in this direction included those by Strotz (1968), Lind (1973), Pines and Weiss (1976), Polinsky and Shavell (1976), and Kanemoto (1988). Although some of these models can be given an interpretation in the context of hedonic price theory, they do not lend themselves to empirical application, so they are not covered in this book. For further discussion of this branch of the literature, see Bartik and Smith (1987) and Palmquist (1991).

The second path drew on hedonic price theory, introduced in Chapter 4, to interpret the derivative of the cross-section regression equation with respect to air pollution as a marginal implicit price, and therefore, a marginal value for the air quality improvement (see Freeman 1974b and Rosen 1974). This section describes how hedonic price theory provides a basis for deriving welfare measures for public goods from observed differences in the prices of houses. The primary emphasis is on model specification and interpretation rather than econometric estimation. The goal of the section is to provide an overview of the methods of welfare measurement based on hedonic price theory and to identify the major conceptual issues. For other current treatments of these matters with greater emphasis on econometric and estimation issues, see Palmquist (2005) and Taylor (2003).

The Basic Hedonic Property Value Model

Assume that each individual's utility is a function of that person's consumption of a composite commodity z and a vector of amenities ($\boldsymbol{Q}$) associated with the house that the person occupies. These amenities, of course, include the structural characteristics of the house (such as size, number of rooms, age, and type of construction). They also include characteristics of the neighborhood in which the house is located (such as quality of local schools, accessibility to parks, stores, and work place, and crime rates), as well as location-specific environmental amenities (such as the local air and water quality).

Any large area has in it a wide variety of sizes and types of housing with different structural, neighborhood, and environmental characteristics. An important assumption of the hedonic technique is that the area as a whole can be treated as a single market for housing services. Individuals must have information on all alternatives and must be free to choose a house anywhere in the market. It is as if the area were one huge supermarket offering a wide selection of varieties. Indeed, Rosen's (1974) model assumed that the household can choose from a continuum of housing attributes, an assumption that is, at best, approximated in actual housing market settings.

Since the focus here is on the values of characteristics to buyers of houses, there is no need to model formally the supply side of this market. Instead, assume that the housing market is in equilibrium—that is, that all individuals have made their utility-maximizing residential choices given the prices of alternative housing locations, and that these prices just clear the market given the existing stock of housing and its characteristics. Under these assumptions, the rental price of the *j*th residential location can be taken to be a function of the structural, neighborhood, and environmental characteristics of that location. In other words,

$$R_j = R\left(\boldsymbol{Q}_j\right). \tag{10.1}$$

As explained in Chapter 4, this relationship can be linear in a characteristic if repackaging of that characteristic is possible. However, in general this need not be the case. Two living rooms with six-foot ceilings are not equal to one living room with a twelve-foot ceiling. Where repackaging is not possible, equation (10.1) will be nonlinear.

To model the problem more formally, consider an individual who occupies house *j*. Her utility is given by

$$u = u\left(z, \boldsymbol{Q}_j\right), \tag{10.2}$$

where z is a Hicksian composite good with a price of 1. This assumption makes the demands for characteristics independent of the prices of other goods, a convenient property for empirical work. The individual maximizes $u(\cdot)$ subject to the budget constraint:

$$M - R(\boldsymbol{Q}) - z = 0. \tag{10.3}$$

A typical first-order condition for the choice of amenity q (an element in $\boldsymbol{Q}$) is

$$\frac{\partial u/\partial q}{\partial u/\partial z} = \frac{\partial R(\boldsymbol{Q})}{\partial q}. \tag{10.4}$$

Assume now that the hedonic price function $R(\boldsymbol{Q})$ has been estimated for an urban area. Its partial derivative with respect to any of its arguments, for example q, gives the implicit marginal price of that characteristic—that is, the additional amount that must be paid by any household to move to a housing bundle with a higher level of that characteristic, other things being equal. If this function is nonlinear, the marginal implicit price of a characteristic is not constant, but depends on its level and perhaps the levels of other characteristics as well. If the individual is assumed to be a price taker in the housing market, that person can be viewed as facing an array of implicit marginal price schedules for various characteristics. An individual maximizes utility by simultaneously moving along each marginal price schedule until she reaches a point where her marginal willingness to pay for an additional unit of that characteristic just equals the marginal implicit price of that characteristic. If an individual is in equilibrium, the marginal implicit prices associated with the housing bundle actually chosen must be equal to the corresponding marginal willingness to pay for those characteristics.

Panel A in Figure 10.1 shows the partial relationship between q and $R(\boldsymbol{Q}) = R\left(q, \boldsymbol{Q}_{-q}\right)$ as estimated from equation (10.1), where $\boldsymbol{Q}_{-q}$ denotes all of

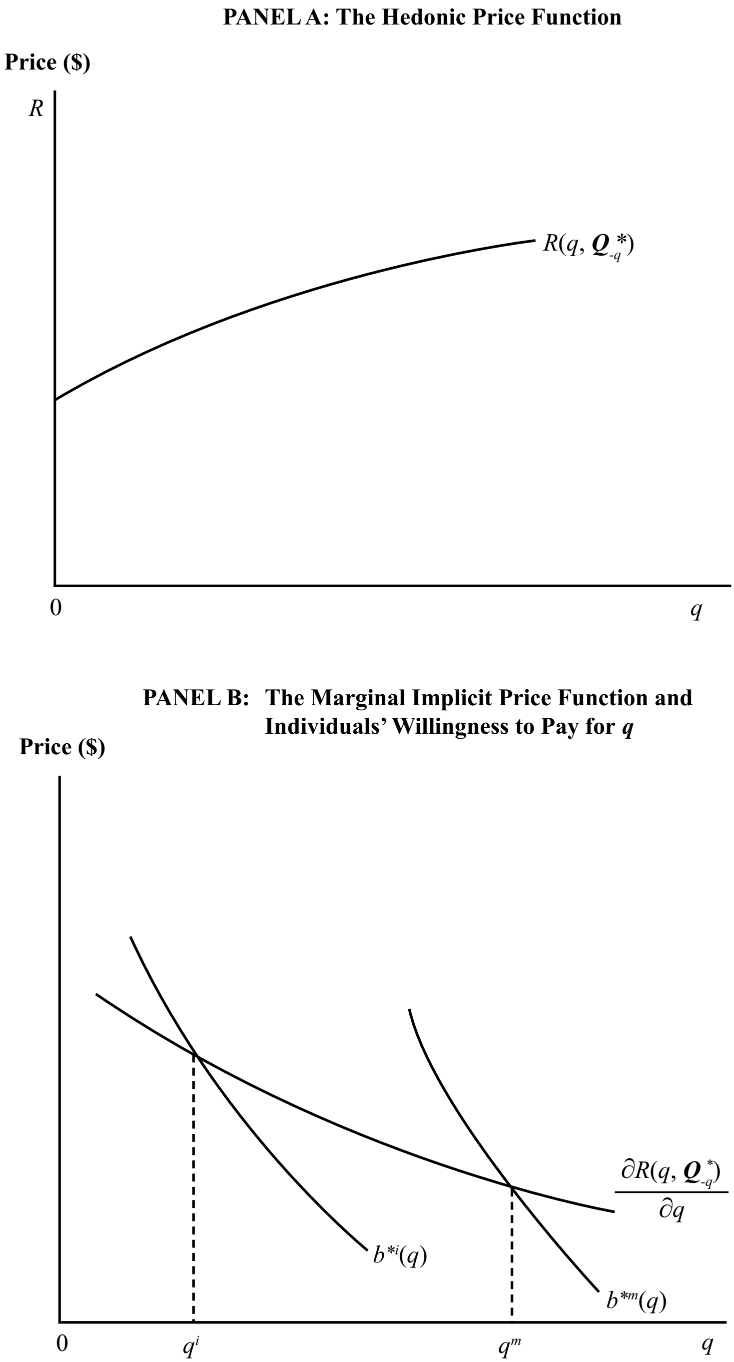

Figure 10.1 Property prices and environmental quality

the attributes except q. In the figure, $\boldsymbol{Q}_{-q}$ is fixed at some level $\boldsymbol{Q}^*_{-q}$. Panel B shows the marginal implicit price of q, $\partial R\left(q,\boldsymbol{Q}^*_{-q}\right)/\partial q$. It also shows the marginal willingness-to-pay curves for two individuals, i and m, who have chosen utility maximizing bundles of housing characteristics, labeled $b^{*i}\left(q\right)$ and $b^{*m}\left(q\right)$. These curves show each individual's marginal willingness to pay for changes in the characteristic, holding utility constant at the level achieved by maximizing equation (10.2) subject to (10.3). Let this level be u^*. Both individuals in Figure 10.1 have chosen locations where their marginal willingness to pay for q is equated with its marginal implicit price.

The analysis described here results in a measure of the price of, and the marginal willingness to pay for, q, but it does not directly reveal the marginal willingness-to-pay function. The fundamental problem is that, for any one individual i, we observe only a single point along their marginal willingness-to-pay function, namely the intersection between $b^{*i}\left(q\right)$ and $\partial R\left(q,\boldsymbol{Q}_{-q}\right)/\partial q$. The second stage of the hedonic technique seeks to combine the quantity and implicit price information, together with restrictions on the structure of preferences, in an effort to identify the marginal willingness-to-pay function for q. The individual's demand price or willingness to pay for q is a function of the level of q. Since there may be substitute and complementary relationships among characteristics, the willingness to pay for q may also depend on the levels (or marginal implicit prices) of other characteristics. It is convenient to assume that the utility function is weakly separable in housing so that prices of other goods can be omitted in the specification of the marginal willingness-to-pay function. Also, it is convenient to assume that each individual purchases only one housing bundle. Purchasing more than one would necessitate that the bundles be identical, or that the hedonic price function be linear in all characteristics. This is because there can be only one marginal implicit price recorded for each individual for each characteristic.

Given these assumptions, for the ith individual we can derive a marginal willingness-to-pay function for q by differentiating the expenditure function, as shown in Chapter 4. The result is

$$b^{*i} = b^{*i}\left(q,\boldsymbol{Q}^*_{-q},u^*\right). \tag{10.5}$$

If equation (10.5) can be estimated, it can be used to estimate the welfare change of an individual associated with changes q, assuming that other things are held equal. Specifically, if the quantities of other characteristics and amenities do not change, the welfare change can be found by integrating b^{*i} over the relevant range of the change in q. However, a change in the quantity of one characteristic can result in changes in the quantities of other characteristics the individual chooses and in changes in the hedonic price function itself. The task of welfare measurement when individuals can fully adjust to the new supply of amenities and characteristics is discussed below in the section Measuring Welfare Changes.

Estimating the Hedonic Price Function

The Dependent Variable

In the discussion so far, the hedonic price has been the annual rental price of the property. What is observed usually is the purchase price of the house, which will be denoted as P_h for the rest of this chapter. P_h can be interpreted as the discounted present value of the stream of expected rental values; but this leads to two complications. First, when housing price differentials are used to estimate welfare changes that are usually expressed as annual flows, care must be taken to convert the house price measures into the appropriate temporal dimension. This topic is discussed in a later section. Second, it might be necessary to take account of expected changes in the characteristics of a house, especially environmental changes when estimating and interpreting the hedonic price function. For example, if air pollution at a given location is expected to improve over time, the present price of the house should be bid up to reflect not only the current conditions but the expected improvement as well.

One question to be asked is whether the dependent variable to be explained should be a pure site or land value or the full price of the house and land together. Since the environmental amenities of interest are location specific but not a part of the structure, the values of the environmental amenities should be reflected in the price of land alone. However, at least in the United States, land is not usually traded separately from the structures placed upon it, so the observed prices reflect the values of both the land and its structural improvements. This causes no problems at the theoretical level, but it does require that the hedonic price equation adequately control for structural characteristics.

Another question concerns the source of data on housing prices. Data on actual market transactions are preferable. For rental housing there is a regular monthly "market transaction," from which fairly accurate data on housing rents could be gathered. However, the majority of residential housing is owner-occupied, and only a small percentage of the total owner-occupied housing stock is exchanged through the market each year. The most preferred source of data is systematically collected information on actual sales prices of individual dwellings, along with relevant characteristics. Fortunately, in many parts of the country these data are collected by multiple-listing services and tax assessing agencies. Increasingly, these data are available electronically, either directly from the primary source (for example, online sites for individual county assessor offices) or from third-party data aggregators (for example, Dataquick). In recent years, virtually all published hedonic property value studies have used micro data on individual transaction prices.

An alternative source of property value data would be professional appraisals of individual properties constructed for taxation or other purposes. The appraisals themselves are typically based on one of two approaches: (a) a sales approach that assesses a home's value based on transaction prices for similar homes in the same area, and (b) a cost approach that assesses the cost of replacing the home. Some jurisdictions have developed computer-based systems of appraisals, made for tax purposes, which include data not only on appraised values but also on a variety of structural and site characteristics. For citations to studies examining the accuracy of appraisals and tax assessments, see Kiel and Zabel (1999, Table 1).

There are two fundamental problems with the use of appraisal data. First, to the extent that a sales approach is used, the data will mask underlying variability in property values, ignoring idiosyncratic factors influencing a home's value. Indeed, as the sales approach often involves a statistical analysis of recent transactions, the resulting assessments can be viewed as providing fitted values from a first stage hedonic regression. Subsequent analysis of these assessed values will be "successful" (that is, in the sense of fitting the data) only to the extent that the researcher happens to choose a functional form that mimics what was used by the assessor. Second, for assessed values based on a cost approach, the reported home prices reflect only the cost side of an equilibrium price function.

There are, of course, concerns with individual transaction data. For example, it is important to ensure that the transactions reflect so-called "arms-length" transactions, eliminating sales between individuals who are related or those in which the sales price has been altered for reasons unrelated to the home's underlying value. It is also good to keep in mind the underlying assumption in hedonic property analysis—that the housing market is in equilibrium (that is, that all opportunities for possible gains from further trade at the revealed set of prices have been exhausted). This is a heroic assumption, because buyers and sellers often operate with substantial ignorance about the true willingness to pay and willingness to accept offers of other potential buyers and sellers. Sellers typically state an asking price that effectively truncates potential offers at that price. Sellers must choose to accept or reject offers more or less at the time they are received, without knowing when (or even if) a higher offer might come along; and buyers lack information on possible prior bids made by others for a given property (Horowitz 1986). Horowitz developed an alternative model of the bidding and acceptance strategies of buyers and sellers, and estimated both a standard hedonic model and his alternative model with the same data set. He found that the statistical performance of his bidding model was substantially superior to the standard model in predicting sales prices. However, since he used principal components in his estimation procedure rather than actual attributes, it is not possible to analyze the impact of his alternative modeling strategy on marginal implicit prices for environmental attributes.

Explanatory Variables

In choosing the appropriate explanatory variables, the first question to be addressed is the way in which environmental amenities and location characteristics enter the hedonic price function. The typical practice has been to enter a simple scalar measure of an amenity—for example, parts per million (ppm) of an air pollutant, or distance to a park. However, Parsons (1990) showed that this practice is not consistent with a restriction imposed on the hedonic price function by profit maximizing behavior on the supply side of the housing market. The implication of profit maximization is that the effect of the environmental amenity can only be captured without bias by weighting the amenity by the area of the lot on which the house sits. The restriction is that if an area of land of given q is developed and sold in two or more different lot sizes, the prices of the lots must be such that the return per acre is independent of the sizes of the lots. For example, if two one-acre lots sell for $\$X$ apiece, one two-acre lot must sell for $\$2X$. This means that the premium on lots with higher levels of q must be twice as high for the two-acre lot compared with the one-acre lots. The higher premium on the larger lot is necessary to compensate the landowner for the forgone opportunity to capture two premiums with the smaller lots.

Although the argument is correct in principle, there is some question about its relevance in practice. As Parsons points out, once lots are developed, the cost of changing the size of lots on which houses sit may be too high to force amenity premiums to take the weighted form in the secondhand market for houses. Although Parsons showed that biased estimates of implicit prices for characteristics are possible, few empirical researchers have used his proposed weighted amenity values.

The levels of some environmental amenities are fixed by location, while the levels of others, especially those related to air quality, vary over time with changes in emissions and meteorological conditions. With time-varying amenities, there is the question of how best to represent the level of the amenity in the regression equation. The typical practice in air pollution–property value studies has been to use the annual mean as a summary statistic. However, Murdoch and Thayer (1988) have shown that in the case of visibility, using more information on the probability distribution of visual range improves the statistical performance of the hedonic price function.

A number of other conceptual and practical issues must be resolved in the course of selecting a set of explanatory variables for a hedonic price function. These include:

- Which measures of environmental quality should be used to characterize environmental amenities?
- Is it possible to separate the effects of different amenities on property values when measures of the amenities are correlated?
- What objective data best capture "neighborhood" characteristics?
- Does the spatial scale of the socioeconomic data often used in these studies correspond to peoples' perceptions of these characteristics? (For example, is there sufficient homogeneity within census tracts so that census tract means or medians adequately measure "neighborhood"?)
- Is there sufficiently close correspondence between peoples' perceptions of amenity levels (which presumably govern the choices reflected in property prices) and the objective measures of amenity levels that are available to the researcher?

Since the objective of the hedonic analysis is to determine the effect of one amenity on property values, other things being equal, a key issue is the control for structural, neighborhood, and other environmental variables. The issue is made more difficult by the likelihood of multicollinearity among housing characteristics. This raises the troublesome question of the tradeoff between increasing bias through the omission of variables that are correlated with the variable of concern and increasing the variance or imprecision of coefficient estimates when collinear variables are included. Theory does not provide any hard-and-fast answers to this question. Work by Atkinson and Crocker (1987) and Graves et al. (1988) suggested the value of approaching this question systematically, using Bayesian principles. These authors have also examined the effects of errors in the measurement of other explanatory variables on the estimates of the coefficients on the environmental variables of concern.

Finally, hedonic property analysis has often been employed to assess the impact of NIMBY ("not in my backyard") sites, such as toxic waste facilities (Kohlhase 1991), incinerators (Kiel and McClain 1995), Superfund sites (Kiel 1995), and animal confinement units (Palmquist, Roka, and Vukina 1997; Herriges, Secchi, and Babcock 2005). A difficulty in this setting is that a property may be exposed to multiple NIMBY sites, raising the question as to which site or sites to include in the analysis, and how to quantify their joint and marginal impacts. Most studies focus on the nearest site and assume that its marginal impact diminishes with distance. However, for some externalities, distance alone is a poor proxy for exposure. In the context of animal confinement units, for example, exposure to the odor generated by the various facilities can depend upon the prevailing winds. Cameron (2006) developed a more general representation of the hedonic price function that allows the marginal impact of a single site to depend upon its directional location relative to the property and to identify the direction from which the property value gradient is highest. The approach, however, still requires that the analyst focus on a single NIMBY site.

Functional Form

Functional forms for the hedonic price function that have been proposed or used in the literature include the linear, the quadratic, the log-log, the semi-log, the inverse semi-log, the exponential, and the Box–Cox transformation. The first step in choosing a functional form is to see what theory can tell us. According to theory, a hedonic price function is an equilibrium relationship derived from the interaction of individuals' preferences and suppliers' cost or profit functions. The only obvious general restriction on the form of the hedonic price function is that its first derivative with respect to an environmental characteristic be positive (negative) if the characteristic is a good (bad).

Rosen (1974) and Epple (1987) showed that it is possible to solve for the hedonic price function analytically after making specific assumptions about the form of individual utility functions and the distribution of suppliers' characteristics (Rosen) or the exogenous supply of housing characteristics (Epple). These analytical solutions are only possible for a very limited set of assumed forms of preferences and supply. For example, Rosen assumed that individuals' utility functions were linear. This is not an attractive assumption, especially if the ultimate objective of the analysis is to measure welfare values for changes in supplies of environmental characteristics.

Early researchers tried alternative functional forms for the hedonic price function (typically the semi-log, inverse semi-log, and log-linear) and selected one on the basis of goodness-of-fit criteria. Goodman (1978) was one of the first to experiment with a flexible functional form. He employed a Box–Cox transformation of the dependent variable:

$$P_h^{[\lambda]} = \frac{P_h^{\lambda} - 1}{\lambda}. \tag{10.6}$$

For $\lambda = 1$, this is a simple linear function. As λ approaches zero, this becomes the semi-log form. Some authors have found estimates of λ that were significantly different from both zero and one, indicating that this more complicated form fits the data better than either the linear or semi-log forms.

Transforming only the dependent variable still produces a very limited range of possibilities. Halvorsen and Pollakowski (1981) proposed estimating what they called a quadratic Box–Cox functional form. It would have the form

$$P_h^{[\lambda]} = \alpha_0 + \sum_{j=1}^{J} \alpha_j \cdot q_j^{[\gamma]} + \tfrac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk} \cdot q_j^{[\gamma]} \cdot q_k^{[\gamma]}, \tag{10.7}$$

where j and k index the characteristics and λ and γ are estimated from the data.

Cassel and Mendelsohn (1985) pointed out that for welfare analysis, it is not the goodness-of-fit of the hedonic price function that matters; rather, it is the estimate of the marginal implicit price of the environmental attribute. In the regression equation for housing price, the environmental variable is likely to have relatively little influence in determining the estimated magnitude of γ. However, the estimate of γ would have a major impact on the estimated marginal implicit price of the environmental characteristic.

A more general flexible form would be an extension of the quadratic Box–Cox of equation (10.7) to

$$P_h^{[\lambda]} = \alpha_0 + \sum_{j=1}^{J} \alpha_j \cdot q_j^{[\gamma_j]} + \tfrac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk} \cdot q_j^{[\gamma_j]} \cdot q_k^{[\gamma_k]}. \tag{10.8}$$

This functional form allows for different transformations for each independent variable.[1] This general form may not be estimable when there are a large number of characteristics. A compromise, proposed by Palmquist (1991), would be to set $\gamma_j = \gamma_k$ for all $j,k = 1, \ldots, n-1$, where n indexes the environmental attribute of

1 For dummy variables (for example, indicating whether or not a housing unit is in a particular neighborhood), γ_j cannot be identified, and is simply set equal to one.

interest. This would be responsive to the point raised by Cassel and Mendelsohn. Allowing for a separate transformation of the environmental amenity should give better results.

One question about functional form is whether the form chosen allows the marginal implicit price of the environmental characteristic to depend on the levels of the other attributes of houses. Of the commonly used functional forms, only the log and the Box–Cox transformation make the implicit prices of characteristics depend on the levels of other characteristics—the other forms impose independence. However, this is a question that should be answered by the data, not by assumption.

One of the early efforts to systematically consider the question of functional form was a study by Cropper, Deck, and McConnell (1988). The authors simulated the performance of a housing market using real data on buyer and housing characteristics drawn from the Baltimore, Maryland, area. After specifying the functional form and parameters of individuals' utility functions and the distribution of characteristics that reflect taste differences across individuals, they solved the assignment problem, producing a housing market equilibrium with each house being sold to the individual with the highest willingness to pay for its bundle of characteristics. With knowledge of the utility function parameters, it was then possible to calculate the true marginal implicit price for each individual and for the mean across all individuals. The authors used the equilibrium prices to estimate six alternative functional forms for the hedonic price function. They were then able to compare the mean true marginal bids with the bids calculated from each hedonic price function. They found that when all of the housing characteristics were included in the hedonic price function, the linear and quadratic versions of the Box–Cox transformation provided the most accurate estimates of marginal implicit prices. However, when they experimented with various forms of incorrectly specified hedonic price functions (by omitting variables or using proxy variables) they found that the linear version of the Box–Cox transformation was consistently superior in generating marginal implicit prices.

For over 20 years, this research has provided a rationale for relying on relatively simple functional forms when estimating hedonic price functions. More recently, however, Kuminoff, Parmeter, and Pope (2010) revisited this issue in light of newer econometric techniques, including spatial fixed effects and quasi-experimental methods, designed to control for omitted variables bias. The authors undertook an extensive Monte Carlo analysis evaluating over 540 different hedonic models, with a range of underlying utility functions and assumed functional forms for the price functions, as well differences in the nature of controls used by analyst to deal with omitted variables. Their overall conclusion was that "the more flexible specifications for the price function, such as the quadratic Box–Cox model, outperform the linear, log-linear, and log-log specifications that have dominated the empirical practice for the past two decades" (Kuminoff, Parmeter, and Pope 2010, 159).

The Hedonic Price Function as a Market Equilibrium

Interpreting the marginal implicit prices as measures of households' marginal willingness to pay requires the assumption that each household is in equilibrium with respect to a given vector of housing prices and that the vector of housing prices is the one that just clears the market for a given stock of housing. These conditions assure that the hedonic price function is the price vector that makes all participants in the market in aggregate just willing to hold the existing stock of housing. For these two aspects of equilibrium to be fully achieved, we require first that households have full information on all housing prices and attributes and that their transactions and moving costs be zero; and second, that the price vector adjust instantaneously to changes in either demand or supply. The market for housing can be viewed as a stock-flow model where the flow (change in stock) is a function of prices, but the prices at any point in time are determined only by the stock at that point in time.

This idealized model is clearly not an accurate representation of real-world housing markets. However, in evaluating the strength of this criticism of the hedonic price model, we must focus on several distinct issues. One issue concerns the speed of adjustment of the market to changed conditions of supply and demand. If adjustment is not complete, then observed marginal implicit prices will not accurately measure household marginal willingness to pay. A major question is whether imperfect adjustment will lead to systematic biases in estimates of willingness to pay.

Consider households' imperfect adjustment to changing prices. First, an increase in housing prices need not affect the marginal implicit prices of attributes, in which case no adjustment of the attribute bundles is necessary. Even if marginal implicit prices change, households will not move unless the potential utility gains to returning to full equilibrium exceed the information costs, transactions costs, and moving costs associated with the change. These costs help to define a band within which observed marginal implicit prices can diverge from household marginal willingness to pay for housing attributes. If housing prices change so that the marginal implicit price schedule for an attribute moves consistently in one direction, households will consistently lag in their adjustment to that change; and the marginal willingness to pay will be overstated or understated according to whether the marginal implicit price is rising or falling.

A second issue concerns expectations about future environmental amenity levels. Market prices for long-lived assets such as housing reflect the discounted present value of the stream of expected future services from that asset. A change in expectations about future environmental amenity levels can affect housing prices and marginal implicit prices independently of the present level of these amenities. For example, if there are widespread expectations of an improvement in air quality and the market adjusts reasonably quickly to these expectations, the price differential between presently dirty houses and clean houses should decrease. Correlating these prices with existing levels of air pollution would lead to an underestimate of the marginal implicit price of air quality.

Divergences from full equilibrium of the housing market in many circumstances will only introduce random errors into the estimates of marginal willingness to pay. However, where market forces are moving continuously in one direction or are expected to move in one direction, incomplete market adjustment, or full adjustment to changing expectations, or both, can introduce biases in both directions. We should be much more cautious about utilizing the cross-section hedonic price model in those cities and at those points in time during which market forces and environmental quality levels are changing rapidly (granted that "rapidly" is an imprecise term). However, it is also possible in these circumstances to determine the direction of bias. Thus, estimates of marginal willingness to pay derived from such studies can be labeled as an upper bound or a lower bound on the basis of that analysis.

Another issue concerns the possibility of corner solutions. If there is not a sufficiently wide variety of housing models available, corner solutions are likely. The hedonic price function defines an opportunity locus across attribute space. A household chooses a housing model such that its indifference surface is tangent to the given opportunity locus, provided that a model with that precise set of attributes is available. If the optimum model is not available, the household must pick the nearby housing model that gives the highest utility level; but then the first-order conditions for utility maximization are not satisfied as equalities (Mäler 1977, 361–362).

The hedonic model is based on an assumption that the implicit price function is differentiable and continuous. However, this is an artifact of the statistical and mathematical technique. If this assumption is not satisfied in practice, two sorts of problems can arise. The first problem is that the statistically fitted hedonic price function is a good approximation only when the number of housing units is large and there is more continuous variation in characteristics among units. A small number of distinctly different types of housing units might be better analyzed with one of the discrete choice models described later in this chapter. The second type of problem arises if there are no units available with particular combinations of attributes. If there are substantial gaps in the opportunity locus, some households will not be able to satisfy the first-order conditions as equalities. This could be a problem for certain subsets of the urban population.

Market Segmentation

Straszheim (1974) was the first to raise the question of market segmentation in the context of estimating hedonic price functions for housing within an urban area. He argued that the urban housing market really consisted of a series of separate, compartmentalized markets with different hedonic price functions in each. As evidence in support of the segmentation hypothesis, Straszheim showed that estimating separate hedonic price functions for different geographic areas of the San Francisco Bay area reduced the sum of squared errors for the sample as a whole.

For different hedonic price functions to exist in an urban area, two conditions must be met. The first is that the structure of demand, the structure of supply, or both must be different across segments—either buyers in separate submarkets must have different structures of demands, or the structure of characteristics of the housing stocks must be different. The second condition is that purchasers in one market segment must not participate significantly in other market segments. In other words, there must be some barrier to mobility of buyers among market segments that prevents arbitrage from occurring in response to differences in marginal implicit prices. Such barriers could be due to geography, discrimination, lack of information, or a desire for ethnically homogeneous neighborhoods. Even with buyer immobility, if demand and supply structures are the same they will produce similar structures of hedonic prices. Perfect mobility and information on the part of buyers will also eliminate differences in the implicit prices for any characteristic across market segments.

If market segmentation does exist, the hedonic price function estimated for the urban area as a whole will provide faulty estimates of the implicit prices facing subsets of buyers in different market segments. Thus, estimates of benefits and estimates of demand functions based on faulty price data will also be faulty. If market segmentation does exist, separate hedonic price functions must be estimated for each segment; and benefit and demand functions must be separately estimated for each segment with a different set of implicit prices.

It is not clear how significant the problem of market segmentation is for air pollution–property value studies within single urban areas—although there are enough positive results in the literature to suggest that it is not a problem that can be dismissed out of hand. Some authors have found evidence of different hedonic price functions for submarkets within larger urban areas, suggesting segmentation; however, this could be due to misspecification of the model, as others have not found evidence of segmentation in their data.

The existence of market segmentation does not render the hedonic price technique invalid; but rather, it makes application of the technique more difficult. If the appropriate basis for segmentation can be identified, it is conceptually possible to estimate separate implicit price functions for each submarket. Although these functions would be different across markets, they each would accurately reflect the outcome of the market processes in each submarket. Thus, the functions could be used to estimate equilibrium marginal willingness to pay.

Econometric Concerns

While the emphasis in this chapter has been on model specification and interpretation, a number of econometric developments in recent years are worth noting. Most of them stem from concerns regarding omitted variables. First, given the spatial nature of hedonic analysis, it is likely that the error terms in a hedonic price model are correlated over space, with the correlation being larger for housing units closer to each other, precisely because these units share common omitted variables. This so-called *spatial autocorrelation* is analogous to serial correlation in a time series setting, where in the latter case the correlation increases for observations closer to each other in time. Ignoring spatial autocorrelation will lead the researcher to overstate the precision with which individual parameters are estimated, potentially resulting in erroneous conclusions regarding statistical significance. A more problematic form of spatial dependence arises when the price of one unit depends upon the sales prices of other houses in the same region (known as spatial autoregression). Ignoring spatial autoregression can result in biased estimates of the price function parameters. Palmquist (2005) provided a helpful overview of the econometric techniques for handling both types of spatial dependence. Specific applications in the hedonic pricing literature include Bockstael (1996), Can (1992), and Bell and Bockstael (2000).

Second, as noted earlier, omitted variables can result in biased parameter estimates to the extent that the omitted factors are correlated with the regressors included in the hedonic price function. One approach to dealing with unobserved housing unit attributes (including locational amenities) is the use of repeat sales data (Palmquist 1982). Consider a housing price function given by

$$P_{hjt} = \alpha_0 + \sum_{k=1}^{K} \beta_k q_{kjt} + \xi_j + \varepsilon_{jt} = \alpha_0 + \sum_{k=1}^{K} \beta_k q_{kjt} + \eta_{jt}, \tag{10.9}$$

where P_{hjt} denotes the sales price of house j in time period t and q_{kjt} denotes the kth attribute for the same unit in time period t. The term ξ_j denotes unobserved attributes of the housing unit, which are subsumed into the error term $\eta_{jt} = \xi_j + \varepsilon_{jt}$. To the extent that ξ_j is correlated with the observable housing unit attributes (that is, the q_{kjt}), least squares regression will yield biased parameter estimates. However, if repeated sales data are available (say in time periods t and t'), then the two sales prices can be differenced, yielding a new regression equation:

$$\Delta P_{hj} \equiv P_{hjt'} - P_{hjt} = \sum_{k=1}^{K} \beta_k \left(q_{kjt'} - q_{kjt}\right) + \left(\varepsilon_{jt'} - \varepsilon_{jt}\right) = \sum_{k=1}^{K} \beta_k \Delta q_{kj} + \Delta\varepsilon_j. \tag{10.10}$$

In this case, by focusing on changes in housing prices over time for the same unit, any unobserved attributes of the house that are constant over time are differenced out.

The use of repeat sales data is not without its drawbacks. First, for any given market and time period, there will be fewer housing units that will have sold more than once, limiting the sample size available for the hedonic regression. Increasing the sample size by considering a larger market or a longer time period risks violating the assumption that a single hedonic price function applies for all the observations. Second, first differencing will also eliminate any observable site attributes that have remained constant over time. Third, if there are unobserved housing attributes that have changed over time, the potential for omitted variable bias remains, though it is likely to have been attenuated.[2]

In recent years, a variety of *quasi-experimental* techniques have been developed to eliminate or reduce the potential bias from omitted variables (see, for example, Angrist and Pischke 2008 and Imbens and Wooldridge 2009 for general treatments). Two prominent approaches are:

- *Difference-in-differences*: A natural measure of the impact that a locational amenity (for example, air quality) has had on property values is the difference (D_0) in housing prices before and after changes have occurred in that amenity. The risk, of course, is that other things have changed as well. In its simplest form, the difference-in-differences (DID) approach attempts to control for these other factors by also measuring the difference (D_0) in housing prices over the same time period for a control group that did not experience changes to the locational amenity of interest. The difference in these differences (that is, D_1–D_0) is assumed to eliminate the nuisance of other factors and isolate the impact of the locational amenity. Chay and Greenstone (2005) provided an example of the use of DID in measuring the impact of air quality on the housing market, while Kuminoff and Pope (2012) illustrated the risks in its use.
- *Spatial discontinuities*: The regression discontinuity approach provides a means of controlling for unobservables by examining outcomes (property values in the hedonic setting) on either side of an arbitrary threshold. For example, in considering the impact of school quality on the housing market, homes immediately on either side of a school district's boundary will share commonly unobserved locational attributes (such as crime rates, access to shopping, community "feel," etc.), with differences in property values attributable to differences in observable structural characteristics of the home and differences in the associated schools. Greenstone and Gallagher (2008) examined the impact of Superfund-sponsored cleanups on housing prices using a regression discontinuity approach, comparing housing prices in areas immediately above and immediately below the threshold to qualify for Superfund cleanup. Lee and Lemieux (2010) provided a general discussion of the technique.

Estimating Characteristics Demands

The attractiveness of the hedonic price model for applied welfare analysis lies in the potential for using estimates of individuals' marginal implicit prices for a characteristic to recover the uncompensated inverse demand function for q or information on the underlying structure of preferences. Rosen (1974) had argued

2 For a more extensive discussion of the repeat sales model, see Freeman (2003, 388–390).

that the inverse demand and marginal supply price functions could be estimated from the information contained in the hedonic price function in the following manner:

> … compute a set of implicit marginal prices … for each buyer and seller evaluated at the amounts of characteristics … actually bought or sold, as the case may be. Finally, use estimated marginal prices … as endogenous variables in the second-stage simultaneous estimation of [the inverse demand and supply price functions]. Estimation of marginal prices plays the same role here as do direct observations on prices in the standard theory and converts the second-stage estimation into a garden variety identification problem.
>
> (Rosen 1974, 50)

This suggestion has been the source of a large literature for more than three decades. Since the emphasis in this book is on models and basic economic method rather than on econometric issues, only an overview is provided here of the sources of problems in estimating and identifying demand functions for characteristics and alternative approaches to solving them.[3]

The difficulties in estimating and identifying the inverse demand functions from hedonic price data come in two forms. The first arises from the fact that the source of data for the dependent variable in the marginal willingness-to-pay function is not direct observation of the inverse demand prices; rather, it is the calculation of the marginal implicit price $\partial P_h / \partial q$ from the estimated hedonic price function. However, this variable is itself computed as a function of the same characteristics that are explanatory variables in the marginal willingness-to-pay function. Brown and Rosen (1982) and Mendelsohn (1987) showed that at least in some cases this procedure leads to parameter estimates for the marginal willingness-to-pay function that are identical to the estimated coefficients in the hedonic price function. As Brown and Rosen put it,

> Contrary to Rosen's original statement, we claim that marginal attribute prices constructed as above will not necessarily play the same role in estimation that direct observation on prices would play if they were available. Because such constructed prices are created only from observed sample quantities, any new information that they may provide (that is, any information beyond that already provided directly by observed sample quantities) can only come from a priori restrictions placed on the functional form of the price function $P_h(\cdot)$. In the absence of such additional restrictions, second-stage structural

3 Readers interested in more technical discussion, especially from an econometric perspective, should consult Brown and Rosen (1982), Epple (1987), Bartik (1987), and McConnell and Phipps (1987). Bartik and Smith (1987) and Palmquist (1991) also provided useful reviews.

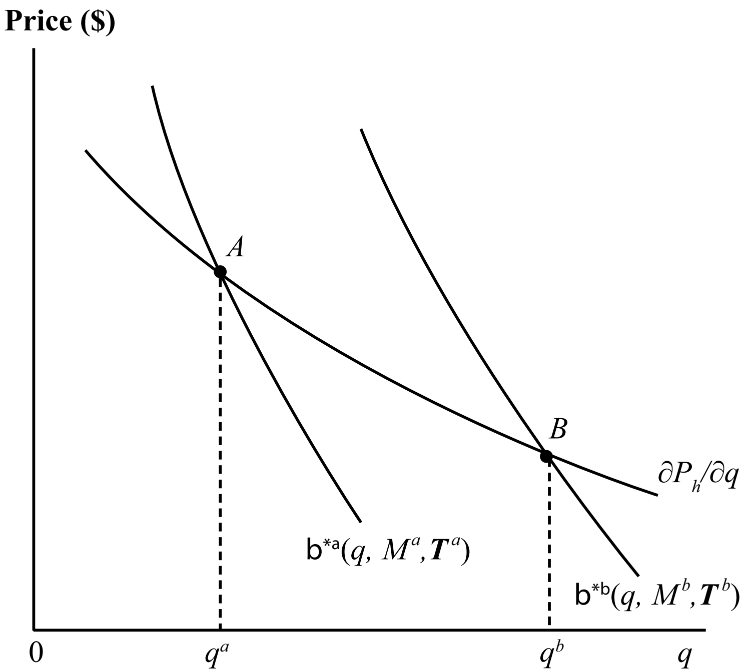

Figure 10.2 The identification problem when income and the quantity of the characteristic are correlated

> estimation of the sort suggested by Rosen may only reproduce the information already provided by the first-stage estimation of the $P_h(\cdot)$ function.
>
> (Brown and Rosen 1982, 176: notation changed by the authors)

In other words, since the second-stage estimation procedure utilizes no additional data beyond that already contained in the hedonic price function, it can do no more than reproduce the coefficients estimated from the hedonic price function.

The nature of the problem can be illustrated with Figure 10.2. This example is due to Bartik (1987, 84–85). Consider two individuals with the same income and uncompensated inverse demand functions of the form

$$b^* = b^*\left(q, M, \boldsymbol{T}\right), \tag{10.11}$$

where M is income, and $\boldsymbol{T}$ is a vector of unobserved determinants of tastes. For the first individual, we observe point A on her inverse demand function; but we have no information on the demand price for other levels of q. If the two individuals choose different levels of q, for example, q^a and q^b, it must be because of differences in tastes. This means that the demand-shifter, which is unobserved in the data of Figure 10.2, is correlated with the observed choices of q. As Palmquist put it, "the other marginal prices [on the individual's marginal willingness-to-pay function] are only observed for other individuals with other socioeconomic characteristics and provide no information on the original consumer's bid for different quantities of the characteristic" (Palmquist 1991, 96). This makes it very difficult to separate out the effects of demand-shifters from the price–quantity relationship itself.

One approach to solving the identification problem is to impose sufficient structure on the problem by assumption to assure that the conditions for identification of the inverse demand function are met. Quigley (1982), who assumed a functional form for preferences that included homotheticity as a property, provided an early example. See Chattopadhyay (1999) for another example of this approach. By specifying the relationship between income and demand, this assumption made it possible to separate the effects of income and quantity change on the marginal willingness to pay for characteristics. Note, however, that the assumptions about functional form are not testable.

Recently, Ekeland, Heckman, and Nesheim (2004) argued that the criticisms of Rosen's two-stage estimation procedure are misleading, based on "arbitrary linearizations that do not use all of the information in the model … Nonlinearities are generic features of equilibrium in hedonic models and a fundamental and economically motivated source of identification" (2004, S60). They went on to examine identification in the context of a normal-linear-quadratic version of a single market hedonic price model, and demonstrated that "[w]ith mild functional form assumptions, the model is completely identified" (2004, S96). While the approaches developed in Ekeland, Heckman, and Nesheim (2004) still rely upon structural assumptions, the line of research would seem to be promising in that the specifications considered are relatively general and authors draw instead on implications of underlying hedonic equilibrium to bolster identification.

An alternative approach to solving the identification problem is to find cases where the marginal implicit prices of characteristics vary independently of the other demand-shift variables. Specifically, this means finding cases where individuals with the same preferences, income, and other traits face different marginal implicit prices. This can only occur if similar individuals must choose in markets with different hedonic price functions, which in turn implies either segmented markets within a city or observations taken from several different housing markets (as, for example, in different cities).

The first step in implementing this approach is to estimate separate hedonic price functions for each housing market, using the same specification. The second step is to compute the marginal implicit price faced by each individual from the hedonic price function in that market. Then the computed marginal implicit prices can be regressed on the observed quantities of the characteristics and the exogenous demand-shifters to obtain the uncompensated bid function. Assuming sufficient independent variation across markets, and assuming that there are no unobserved differences in preferences across individuals, this approach will lead to reliable and properly identified bid functions. For examples of this approach, see Palmquist (1984), Bartik (1987), and Zabel and Kiel (2000).

The second source of difficulty in estimating inverse demands for attributes lies in the fact that both the quantity of the characteristic and its marginal implicit price are endogenous in the hedonic price model. Unlike the standard market model in which an individual faces an exogenously determined price and chooses a quantity, and unlike a quantity-rationed market in which an individual faces an exogenously determined quantity and reveals a marginal willingness to pay, the individual chooses both a point on the hedonic price schedule and its associated quantity. The choice of that point simultaneously determines the marginal willingness to pay and the quantity of the characteristic.

One approach to solving this problem is to find truly exogenous variables to be used as instruments. This appears to be a difficult task, however. For some of the suggested possibilities and their problems, see Mendelsohn (1984, 1985), Bartik (1987), Bartik and Smith (1987), and Palmquist (1991). Another possibility is to assume that there is a characteristic in the marginal implicit price function that is not an argument in the marginal willingness-to-pay function for another characteristic. This makes it possible to use the omitted characteristic as an instrument. Recently many authors have been critical of this approach, since the results are only as good as the assumptions imposed to obtain them—see Mendelsohn (1987), Bartik and Smith (1987), Horowitz (1987), and Palmquist (1984). Unfortunately, such assumptions are not testable. More recently, Ekeland, Heckman, and Nesheim (2004) suggested the use of nonlinear instrumental variables as an alternative approach to addressing the endogeneity problem.

Measuring Welfare Changes

It has been established that in a housing market in equilibrium, utility-maximizing individuals equate their marginal willingness to pay for housing characteristics with the marginal implicit prices of these characteristics, and that in some circumstances it may be possible to estimate inverse demand functions based on this information. The question then is how this information on prices and preferences extracted from the hedonic housing market can be used to calculate measures of welfare change for changes in environmental amenities. The basic concepts of welfare measurement at the level of the individual are straightforward and were introduced in Chapter 4. However, measurement of aggregate welfare changes based on hedonic prices is made difficult by the adjustments that people are likely to make in response to changes in environmental attributes and the possibility that the hedonic price function will change. Also, in principle, it is necessary to consider possible changes in the supply side of the hedonic property market.

In this subsection, the basic welfare measure is defined for marginal changes in a characteristic or environmental amenity, holding other things constant, in particular, individuals' choices of housing bundles. The benefits of a nonmarginal change in an amenity are then considered, assuming that individuals cannot adjust their housing bundles by moving. This measure looks only at benefits to purchasers of housing bundles. A fully general measure of welfare change is next examined, one that includes possible changes in profits on the supply side of the hedonic market as well as the consequences of individuals' adjustments on the demand side of the market. Since this conceptually correct measure is not implementable in practice, lower or upper bound approximations of the correct measure are considered. Finally, the subsection concludes with the consideration of localized changes in environmental quality that lead to benefits to some people without changing the hedonic price function.

Since a change in an environmental amenity in an urban area is nonexcludable and nondepletable, it is, in effect, a public good. The marginal value of the change, then, is simply the sum of the marginal willingness to pay of each of the N affected individuals evaluated at the existing housing equilibrium. In other words, for the amenity q:

$$w_q = \sum_{i=1}^{N} b^{*i} = \sum_{i=1}^{N} \left(\frac{\partial P_h}{\partial q_i} \right), \tag{10.12}$$

where w_q is the aggregate marginal welfare change and b^{*i} is the ith individual's marginal willingness to pay. Although most proposed environmental policy changes are nonmarginal in magnitude, the ease of calculating equation (10.12) may make it useful for indicating whether some improvement is desirable, by comparing this measure with an estimate of the marginal cost of the improvement.

Welfare Changes without Adjustments

Recall from Chapter 4 that the equilibrium hedonic price function is given by the double envelope of the bid and offer curves for all of the characteristics. A change in the level of q will place at least some individuals out of equilibrium, given the existing set of marginal implicit prices. Their efforts to restore their equilibria will result in changes in the hedonic price function and marginal implicit prices. Also, in principle, changes in the hedonic price function could trigger changes in the supplies of houses with different bundles of characteristics. In the next section, welfare measurement in the context of such changes will be discussed. But for now, assume that all individuals are constrained to stay at their original location, as might be the case with high transactions and moving costs or if a very short run perspective is taken; and also, assume that there is no supply response to the change in q.

Given these assumptions, the welfare value of the change in q from q^0 to q^1 is given by the sum of the areas under each individual's marginal willingness-to-pay curve over the change in q, or

$$W_q = \sum_{i=1}^{N} \int_{q^0}^{q^1} b^{*i} \left(q, \boldsymbol{Q}^*_{-q}, u_i^* \right) dq \,, \tag{10.13}$$

where W_q is the aggregate benefit. Notice that this measure requires knowledge of the marginal willingness-to-pay functions of individuals. If the uncompensated bid functions from the second stage of the hedonic price estimation are used, the welfare gain from an increase in q will be overestimated.

There is a method for calculating exact welfare measures for nonmarginal changes in a characteristic, holding all other things constant. This method is based on an adaptation by Horowitz (1984) of Hausman's technique for exact welfare measurement for price changes (see Chapter 3). Suppose that the ith individual's uncompensated inverse demand function for q,

$$b^i = b^i\left(q, \boldsymbol{Q}^*_{-q}, M - P_h\right), \tag{10.14}$$

has been identified. Using the indirect utility function, in equilibrium,

$$\frac{\partial v/\partial q}{\partial v/\partial M} = b^i(\cdot) \tag{10.15}$$

and for individual i

$$b^i = \left(\partial P_h/\partial q\right). \tag{10.16}$$

The left-hand side of equation (10.15) is the slope of the indifference curve between the numeraire, M, and q. So, in equilibrium

$$\frac{dM}{dq} = b^i(\cdot) = \left(\partial P_h/\partial q\right)_i. \tag{10.17}$$

This expression can be solved for

$$M = f\left(q, \boldsymbol{Q}^*_{-q}, C\right), \tag{10.18}$$

where C is a constant of integration. The benefit of an increase in q is

$$W_q = f\left(q^0, \boldsymbol{Q}^*_{-q}, C\right) - f\left(q^1, \boldsymbol{Q}^*_{-q}, C\right). \tag{10.19}$$

See Horowitz (1984) and Palmquist (2005) for examples of this approach with various forms of the utility function.

In those cases where neither the uncompensated or compensated inverse demand functions are available, welfare changes could be estimated by making some assumption as to the shape of the marginal willingness-to-pay function through the original equilibrium point. Three alternative assumptions can be used to establish bounds on the true measure. One is to assume that the marginal willingness to pay for each individual is constant—that is, that the marginal willingness-to-pay function for each individual is a horizontal line through the known point. In this case, each individual's benefit for the postulated improvement in the amenity is approximated by the product of the (assumed) constant marginal willingness to pay and the change in the amenity. The aggregate benefit is obtained by summing over all individuals. This assumption leads to an estimate of aggregate benefits that is biased upward.

A second convenient assumption would be that each individual's marginal willingness-to-pay curve decreases linearly from its observed point to the point of the highest attainable level of the amenity. Marginal willingness to pay would be zero at this point. It is not clear whether this approximation would lead to an overestimation or an underestimation of true benefits. The third assumption would be that all individuals' marginal willingness-to-pay functions are identical. Then, as discussed in Chapter 4, the marginal implicit price curve is identified as the marginal willingness-to-pay curve for the representative individual.

Welfare Changes with Full Adjustment

Bartik and Smith (1987, 1223) presented a welfare measure that takes account of all of the adjustments that individuals make in response to the nonmarginal change in q. At any location, the value of a nonmarginal change can be taken to be the integral of the values of a series of infinitesimal changes in the amenity. The value of each small change is taken to be the willingness to pay of the occupant of that site at that point in the sequence of changes. The measure for all sites together is the sum of the values for each site. It is given by

$$W_q = \sum_{j=1}^{J} \int_{q_j^0}^{q_j^1} \frac{\partial P_h\left(q, \boldsymbol{Q}_{-q}^*\right)}{\partial q} dq \, , \tag{10.20}$$

where $\boldsymbol{Q}_{-q}^*$ is the vector of all other site characteristics that are held constant by assumption, j indexes locations, and where the hedonic price function itself is changing in response to the adjustments that people are making.

In principle this measure allows individuals to relocate in response to changes in the quantity and price of the amenity, since, in effect, it sums individuals' marginal values as the amenity changes at each site. This is important, because a major limitation of some of the measures to be described below is their inability to account for individual relocation decisions. Furthermore, this measure does not require knowledge of either the marginal willingness to pay or the bid function. It relies on the fact that at each point in the sequence of changes, each individual's marginal bid is revealed by the marginal implicit price of the characteristic. However, since the hedonic price function is shifting as a consequence of the change in the amenity level, it is necessary to know how the hedonic price function and the marginal implicit prices at each location change as the levels of the amenities at each location change along the path of integration. As a practical matter, this is a major limitation of the measure.

This limitation has forced researchers to look for practical measures that can be interpreted as approximations or upper or lower bounds on the true welfare change. Following the analysis by Bartik (1988), suppose there are increases in several environmental amenities in an urban area. These increases need not be uniform across the area. Specifically, consider the case where the vector $\boldsymbol{Q}$ increases from $\boldsymbol{Q}^0$ to $\boldsymbol{Q}^1$.

First, assuming that individuals cannot move to new locations and that the hedonic price function does not change, the benefit to individuals is given by an expanded version of equation (10.11):

$$W_q = \sum_{i=1}^{N} \int_{Q_i^0}^{Q_i^1} \boldsymbol{B}^i \left(\boldsymbol{Q}, u_i^* \right) d\boldsymbol{Q}_i , \tag{10.21}$$

where each individual's welfare gain is computed from a path-independent line integral over the changes in the individual elements in $\boldsymbol{Q}$, and $\boldsymbol{B}^i(\cdot)$ is the vector of individual marginal willingness-to-pay functions for the characteristics.

Now, at the existing hedonic price function, some people may wish to choose different bundles of characteristics. If they do change, it must be because they perceive themselves to be better off after the adjustment. This welfare gain is in addition to that given by equation (10.21). Thus, (10.21) can be interpreted as a lower bound on the true measure; and it requires knowledge of only the bid or compensated inverse demand functions for the characteristics that change. Furthermore, the effort to adjust to different characteristics bundles is likely to affect the hedonic price function, unless the number of people wishing to do so is quite small relative to the market. It is also possible that the suppliers of housing will respond to changes in the hedonic price function by offering different bundles of housing characteristics. This could have further repercussions on the hedonic price function, and it will increase the profits of housing suppliers.

When all of these repercussions have worked themselves out, the aggregate benefit to individuals can be defined in terms of each individual's total willingness to pay for a housing unit with given characteristics, holding utility constant. Let this total willingness to pay be given by

$$WH_i \left(\boldsymbol{Q}_i^{*j}, u_i^* \right), \tag{10.22}$$

where $\boldsymbol{Q}_i^{*j}$ ($j = 0, 1$) indicates the vectors of environmental and other characteristics actually chosen by the individual in the original and new equilibrium. Each individual's total benefit is the increase in total willingness to pay for the characteristics actually chosen, holding utility constant, minus any increase in actual expenditure on housing. Summing across all individuals, we obtain

$$W_q = \sum_{i=1}^{N} \left[WH_i \left(\boldsymbol{Q}_i^{*1}, u_i^* \right) - WH_i \left(\boldsymbol{Q}_i^{*0}, u_i^* \right) \right] - \sum_{i=1}^{N} \left[P_h^1 \left(\boldsymbol{Q}_i^{*1} \right) - P_h^0 \left(\boldsymbol{Q}_i^{*0} \right) \right]. \tag{10.23}$$

Turning to the supply side of the market, producers, in aggregate, realize a change in aggregate profits given by the increase in expenditures on housing net of any change in their costs. This is given by

$$\Delta\text{Profit} = \sum_{i=1}^{N} \left[P_h^1 \left(\boldsymbol{Q}_i^{*1} \right) - P_h^1 \left(\boldsymbol{Q}_i^{*0} \right) \right] - \sum_{i=1}^{N} \left[C_h^1 \left(\boldsymbol{Q}_i^{*1} \right) - C_h^1 \left(\boldsymbol{Q}_i^{*0} \right) \right], \tag{10.24}$$

where $C_h^j(\cdot)$ is the cost function for producers.

The welfare change for society as a whole is the sum of equations (10.23) and (10.24). Notice that one component of this sum is simply a transfer of revenue from buyers to sellers, so it nets out. The total welfare change is the sum of the increase in total willingness to pay of individuals minus any cost increase on the part of producers of housing. Full implementation of this welfare measure would require enormous amounts of information. However, note that this measure reduces to equation (10.21) if the hedonic price function does not change and if the change in environmental amenities does not affect the costs of supplying housing for producers.

Even if this set of conditions is not satisfied, equation (10.21) can be interpreted as a lower bound on the true measure of benefits. This can be seen by decomposing the true benefit measure into a three-step sequence of changes and adjustments. Consider first the change in amenity levels without any adjustment on the part of individuals or suppliers. The welfare change associated with this step is given by equation (10.21) plus any reduction in the costs of supplying existing houses at the affected locations. Second, suppose hypothetically that the hedonic price function is shifted to its new equilibrium position but that no adjustments to the new price function by individuals or suppliers are permitted. At this stage, although some individuals and suppliers may gain while others lose, on net all of the price changes sum to zero. At this stage, there is no net change in welfare.

Finally, allow individuals and suppliers to respond to the new hedonic price function. Any adjustments that take place at this stage must represent welfare improvements for those responding. The total welfare change is the sum of equation (10.21), any costs reduction to suppliers, and the benefits of adjusting to the price change. The latter two components are either zero or positive. Thus, equation (10.21) represents a lower bound on the true measure of benefits; and the smaller the adjustment to the changes in the hedonic price function, the smaller is the error involved in using (10.21).

Localized Amenity Changes: A Special Case

If the hedonic price function does not shift, then exact welfare measurement may be a relatively easy task. Palmquist (1992a) identified one situation in which the hedonic price function could be assumed to be constant. That is when the number of sites at which there is a change in the amenity level is small relative to the total urban market. If this is the case, and if individuals can move without cost from one site to another in response to the change in environmental amenity levels, then exact welfare measurement is straightforward. The hedonic price function can be used to predict the changes in the prices of affected properties. Benefits are exactly measured by the increase in the values of the affected properties; and knowledge of the marginal bid functions is not required.

Consider the case of an improvement from q^0 to q^1 at just one site, as shown in Figure 10.3. Assume that moving costs for occupants who choose to relocate are zero (the impact of positive moving costs is discussed below). The change in the amenity level results in an increase in the price of this house from P_h^0 to P_h^1. The owner of the property is made better off by this increase in wealth. Even though the occupant of the property experiences the increase in amenity level, he or she is worse off because of the increase in the cost of occupying the property. The occupant is shifted from point A on the curve B^0 to point B on the curve B^1. However, with costless moving, the occupant can relocate to his or her

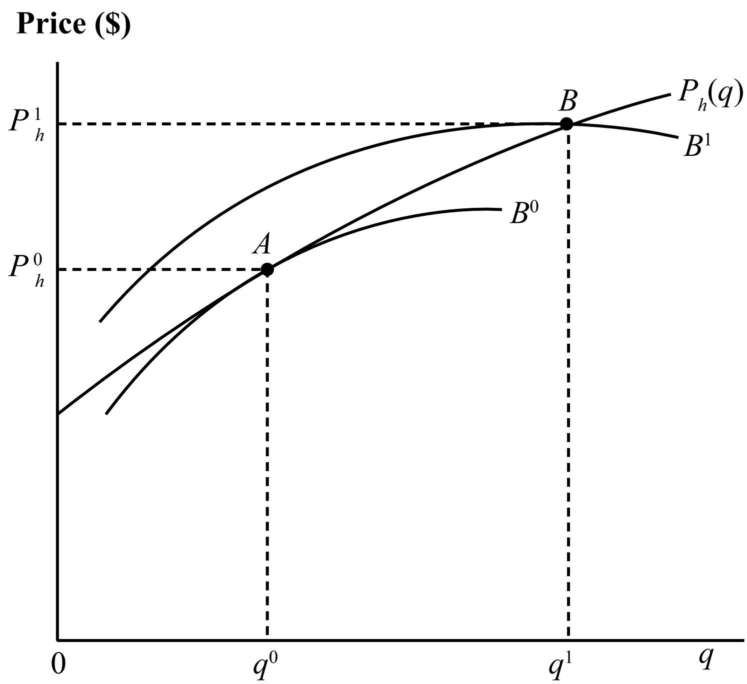

Figure 10.3 The benefit of an amenity improvement at one site when the hedonic price function is unchanged

original equilibrium position. So, the net welfare change is the increase in wealth to the owner. If the owner and occupant are the same person, the result is still the same. This individual might choose to move to a property with an amenity level somewhat greater than q^0 because of a wealth effect. However, the increase in wealth fully captures the benefit of the amenity improvement to this individual.

Now consider the case where the number of affected sites is small so that the hedonic price function does not change, but where moving costs are positive. The renter either loses the moving costs involved in adjusting his/her housing bundle or bears a loss of utility associated with staying at a less preferred location after the amenity change. In either case, the increase in property prices is an upper bound on the total benefit (Palmquist 1992b).

Rents, Taxes, and Property Prices

In the development and exposition of the theoretical model, the discussion has ignored the temporal dimension of housing prices and how welfare measures based on property prices might be converted to the annualized form usually used in welfare evaluation. It is typically the market price of a property that is observed. Inferences about the streams of rents and of benefits are drawn by converting observed present values to annual streams. The institutions of income and real property taxation affect the way in which the market capitalizes rents (and changes and differentials in rents) into market prices for properties. These effects must be properly understood if the process of retracing these steps to infer rents from property value observations is to be successful. This subsection develops and expands on some ideas first presented by Niskanen and Hanke (1977).

In the simplest case of a stream in perpetuity and with no taxes, the conversion of property value to rent is given by

$$R = P_h \cdot r \,, \tag{10.25}$$

where r is the appropriate discount rate. The discussion proceeds by examining first the effects of the two forms of taxation separately, then their combined effects. It is shown that the effects of these two kinds of taxation on the relationship between observed differences in property values and welfare measures depends of the specific features of the tax system and parameter values such as interest rates and tax rates.

Ad valorem taxation of property can be viewed as a device for capturing some of the rent of land for the government. Since taxation affects the net return to the property owner, it should affect the market value of property as an asset. An individual would purchase a property as an asset only if its market price, P_h, is equal to or less than the discounted present value of the rental stream net of property taxation. Market forces would establish the following relationship between property values and rents:

$$P_h = \frac{R - t \cdot P_h}{r} \,, \tag{10.26}$$

where t is the ad valorem tax rate.

If property values are known, the rental stream they represent can be computed by rearranging equation (10.24):

$$R = \left(r + t\right) P_h \,. \tag{10.27}$$

Assume that the property value–amenity relationship, $P_h\left(\cdot\right)$, has been estimated. The marginal benefit of a change in q at a site is

$$w_q = \frac{\partial R}{\partial q} = \left(r + t\right) \frac{\partial P_h}{\partial q} \,. \tag{10.28}$$

The present value of this stream of benefits is

$$\frac{w_q}{r} = \left(1 + \frac{t}{r}\right)\left(\frac{\partial P_h}{\partial q}\right). \tag{10.29}$$

In other words, when the hedonic price function is defined in terms of property value, ignoring the effect of property taxation on the capitalization of rents can lead to the underestimation of benefits. The term (t/r) is a measure of the percentage error resulting from omitting the tax term in the calculation of benefits. For an interest rate of 10 percent and an effective tax rate of 10–20 mils per dollar of market value (1–2 percent), the error is between 10 and 20 percent.

However, this is not the whole story. The income tax code treats the imputed rental income of homeowners differently than it does the rental income of landlords. The absence of a tax liability for imputed rent further complicates the task of inferring annual rents and benefits from observations of (capitalized) market prices for housing assets. This is because the market will place different values on two assets with the same rental stream if one is subject to income taxation while the other is not.

Assume two perpetual assets indexed as *a* and *b* with equal annual returns of R per year. The return to the first asset is taxable at the rate g percent, while the return to the second asset incurs no tax liability. If r represents the market rate of return on assets with taxable returns, the two assets will be priced so as to equalize the after-tax rate of return:

$$P_a = R_a / r \tag{10.30}$$

and

$$P_b = \frac{1}{(1-g)} \frac{R_a}{r} > P_a\,. \tag{10.31}$$

If P_b is observed, the tax-free rental stream can be computed as

$$R_b = r(1-g)P_b\,. \tag{10.32}$$

Taking account of this adjustment and using equations (10.26) and (10.27), the marginal annual benefits of amenity changes to homeowners and their present value are

$$w_q = r(1-g)\partial P_b / \partial q \tag{10.33}$$

$$\frac{w_q}{r} = (1-g)\partial P_b / \partial q\,. \tag{10.34}$$

Ignoring the effects of income taxation leads to an overestimation of benefits. The discount factor $r(1-g)$ is analogous to the municipal bond rate, and it arises for the same reason. However, where the marginal tax rate is itself a function of income, g varies across individuals, and equations (10.33) and (10.34) must be computed separately for each individual.

The tax code confers additional benefits on homeowners by exempting them from taxation on capital gains realized on the sale of a primary residence and by allowing them to deduct property tax payments in calculating taxable income. This latter provision lowers the real cost of the property tax by g percent. Combining these effects (ad valorem taxation, deductibility, and exemption of imputed rental income), we have

$$P_b = \left[R_b - (1-g)tP_b\right] / r(1-g)\,. \tag{10.35}$$

Solving for R_b gives

$$R_h = (r+t)(1-g)P_h. \tag{10.36}$$

Marginal benefits are calculated by

$$w_q = (r+t)(1-g)\left(\frac{\partial P_h}{\partial q}\right) \tag{10.37}$$

and

$$\frac{w_q}{r} = \left(1+\frac{t}{r}\right)(1-g)\left(\frac{\partial P_h}{\partial q}\right). \tag{10.38}$$

The effects of ignoring taxation in calculating benefits depend on the magnitudes of g and t/r. The higher the marginal income tax rate, the more likely benefits would be overstated if taxes were ignored. For an example, suppose the marginal income tax rate is 30 percent, the opportunity cost of capital is 10 percent, and the property tax rate is 2 percent. Then the terms in parentheses come to 0.84. Ignoring tax effects would lead to an overstatement of benefits by almost 20 percent. However, lower income tax and discount rates can reverse this conclusion. An alternative approach to dealing with taxation and discounting is to base the hedonic equation on measures of user cost (called gross rent by Sonstelie and Portney 1980). This variable captures the full cost of owning (and using) an asset such as a house. User cost would include property taxes and the opportunity cost of capital plus any change in the market price of the asset over the interval, say a year. It would be calculated as follows:

$$u = (r+t+m)P_h, \tag{10.39}$$

where m is the percentage rate of change in market value.

The user cost approach differs from that outlined above in two respects. The first is the inclusion of the change in market value over time. P_h could be changing because of physical depreciation of the house, general price inflation, changes in the price of housing relative to other goods, and changes in the variables determining P_h. Only the latter changes have relevance for benefit estimation, and they would be captured by modified versions of equations (10.37) and (10.38), which are generalized from the assumption of constant streams in perpetuity. However, the depreciation term might be useful in empirical work as an approximation of expected changes in these variables, provided that it were adjusted to net out general price level effects.

The second difference arises in considering the effects of some provisions of the income tax code on user cost. For one thing, the tax exemption for imputed rent does not affect the user cost of holding a house, since user cost is an opportunity cost. For another, the tax deductibility of mortgage interest does affect user cost, but it does not affect the market capitalization of streams of benefits. If user cost is used to compute benefits, the net result of these two effects is to overstate benefits in comparison with equations (10.37) and (10.38).

Discrete Choice Models

The hedonic price model is based on the assumption that each attribute of the housing bundle is a continuous variable and that an individual can choose any point on the continuous and differentiable hedonic price function in the n-dimensional attribute space. As noted above, this is clearly not a completely realistic assumption, and in some respects it may seriously misrepresent the problem of choosing a bundle of housing attributes. For example, the number of bedrooms in a house is not a continuous variable. There may be no one-bedroom houses on one-acre lots, or four-bedroom houses with swimming pool and attached garage on one-quarter-acre lots. Discrete choice models provide an alternative way of looking at housing choice and inferring values for housing attributes.

Some of the discrete choice models that have been used in the literature focus on the individual's bid function for housing bundles. Such models are based on the probability that an individual will be the highest bidder for a specified bundle of housing attributes. These are known as bid rent models or random bidding models. An alternative approach is to focus on the individual's utility function defined on housing attributes. These models investigate the probability that a specified bundle of housing attributes (including the price of the bundle) will be chosen by the individual, drawing on the Random Utility Maximization (RUM) models described in Chapter 3, and used extensively in the recreation demand literature.

Both types of models can be used to derive the marginal bid or marginal willingness-to-pay function for individual attributes from an estimate of the bid function or indirect utility function. Thus, in principle, both types of models allow for the calculation of the benefits of changes in an environmental attribute, at least assuming no relocation and no changes in the hedonic price function. Both types of models start with the assumption of utility maximization.

In the bid rent model developed by Ellickson (1981), the utility maximization problem is solved, subject to the standard budget constraint defined by income, prices of market goods, and the hedonic price function, to obtain the individual's bid function—that is, the bid as a function of the housing attributes and income, holding utility constant. As a practical matter, individuals are then grouped into broad "type" categories (based on, for example, income, household size, and other socio-demographic attributes), assuming that bid functions are homogenous within type. The bid function for individuals of type t for a particular housing unit j can be written as a function of the observable housing attributes ($\boldsymbol{Q}_j$) plus a random error term reflecting unobserved attributes of either the individual or the housing bundle:

$$B_{tj} = B_t\left(\boldsymbol{Q}_j\right) + \varepsilon_{tj}. \tag{10.40}$$

For examples of this type of model, see Ellickson (1981), Lerman and Kern (1983), and Gross (1988).

The probability that a household of type t will occupy housing unit j will be determined by whether or not they outbid any other household type; that is,

$$\begin{aligned}\Pr(t \mid j) &= \Pr\left(B_{tj} > B_{t'j}\ \forall t' \neq t\right) \\ &= \Pr\left[B_t\left(\boldsymbol{Q}_j\right) - B_{t'}\left(\boldsymbol{Q}_j\right) + \varepsilon_{tj} - \varepsilon_{t'j}\ \forall t' \neq t\right].\end{aligned} \tag{10.41}$$

If the random error terms are independently and identically distributed with a Type I Extreme Value distribution, this probability can be written in the logit form:

$$P\left[t \mid \boldsymbol{Q}_j\right] = \frac{\exp\left[B_t\left(\boldsymbol{Q}_j\right)\right]}{\sum_{s=1}^{T} \exp\left[B_s\left(\boldsymbol{Q}_j\right)\right]}\ , \tag{10.42}$$

where T denotes the number of household types. As Lerman and Kern (1983) and Gross (1988) pointed out, estimation of equation (10.42) fixes only the slope of the bid function, not its level. However, information on the bids actually paid can be used in the estimation process to fix these values and to make it possible to calculate the marginal bid functions for individual attributes.

Note that the focus in the bid-rent model is on which type of individual occupies a given housing unit. In contrast, in the RUM model, the emphasis is on modeling which housing unit is chosen by a given individual. Specifically, suppose that the conditional utility that individual i receives from choosing housing unit, j, is given by

$$V_{ij} = V\left(R_j, \boldsymbol{Q}_j, M_i, \boldsymbol{S}_i; \beta\right) + \varepsilon_{ij}\ , \tag{10.43}$$

where R_j denotes the rental cost of housing unit j, $\boldsymbol{S}_i$ is a vector of socio-demographic characteristics for individual i, β is a vector of parameters for the conditional utility function, and ε_{ij} is a random error term assumed to capture unobservable attributes of the individual/housing unit, known to the individual but unobserved by the analyst. The individual is presumed to choose the housing unit that maximizes their utility, so that

$$\begin{aligned}\Pr(j \mid i) &= \Pr\left(V_{ij} > V_{ij'}\ \forall j' \neq j\right) \\ &= \Pr\left[V\left(R_j, \boldsymbol{Q}_j M_i, \boldsymbol{S}_i; \beta\right) - V\left(R_{j'}, \boldsymbol{Q}_{j'} M_i, \boldsymbol{S}_i; \beta\right) + \varepsilon_{ij} - \varepsilon_{ij'}\ \forall j' \neq j\right].\end{aligned} \tag{10.44}$$

If, as in Cropper et al. (1993), the error terms are assumed to be independent and identically distributed with a Type I extreme value distribution, then the probability that individual i will choose housing unit j can be written in the standard logit form, with

$$\Pr(j \mid i) = \frac{\exp\left[V\left(R_j, \boldsymbol{Q}_j M_i, \boldsymbol{S}_i; \beta\right)\right]}{\sum_{j'=1}^{J} \exp\left[V\left(R_{j'}, \boldsymbol{Q}_{j'} M_i, \boldsymbol{S}_i; \beta\right)\right]}\ , \tag{10.45}$$

where J denotes the total number of housing units in the individual's choice set. Knowledge of the parameters of the indirect utility function makes it possible to compute welfare measures for changes in any of the housing attributes, including an environmental amenity. One practical issue with the RUM model in this setting is defining the choice set. An individual moving to a large metropolitan region may have literally thousands of housing units to choose from, but may actually only consider a small fraction of these options. In practice, analysts often focus on modeling the choice of a community or neighborhood, assuming that the housing units are largely homogeneous within a given neighborhood.

In an interesting simulation study, Cropper et al. (1993) compared estimates of welfare measures derived from a hedonic price model with those from the random utility model for given known household preferences. After simulating an equilibrium in an urban housing market, they used the resulting hedonic price and individual choice data to estimate both a hedonic price model with its marginal bid functions and a random utility model. They then calculated welfare measures for 25 percent and 100 percent changes in each of ten attributes, including both neighborhood attributes and attributes of individual houses (such as number of bathrooms, lot size, and age). They found that the random utility model provided more accurate estimates of the known welfare measure than the hedonic price model, and this was true even when they assumed that the researcher did not know the true form of individuals' utility functions. They suggest that the reason for this is the difficulty in identifying and obtaining accurate estimates of the marginal bid functions with the hedonic price model when data are generated by only a single market.

Other researchers (see for example, Bartik and Smith 1987, 1224–1225, and Palmquist 1991, 119) have suggested that while one of the strengths of the discrete choice model is its ability to generate welfare measures for nonmarginal changes relatively easily, the model only does so because it forces the researcher to make strong assumptions about the functional form of the utility function or bid function. If similar strong assumptions are made about the functional form of the inverse demand functions for attributes in the hedonic model, these functions can be identified too. Cropper et al. (1993) suggested that even when the functional form of preferences is known, the discrete choice model outperforms the hedonic model as a way of measuring welfare change. For other examples of empirically based efforts to compare welfare measures from standard hedonic models with random utility and random bidding models, see Chattopadhyay (1998, 2000) and Palmquist and Israngkura (1999).

Equilibrium Sorting Models

During the past decade, an alternative paradigm has emerged for modeling the supply and demand for differentiated commodities such as housing. Whereas the hedonic pricing literature focuses attention on the equilibrium outcome of the market, the equilibrium sorting literature seeks to characterize the sorting process itself (that is, how the distributions of individual preferences, production costs, and locational amenities interact to yield a market equilibrium in which the supplies and demands for housing amenities are equated, and both the distributions of housing types and their prices are determined).[4] With a model of the sorting process in hand, it is conceptually straightforward to evaluate both marginal and nonmarginal policy scenarios by simply simulating the new sorting equilibrium that would emerge. These capabilities, of course, come at a cost, as they invariably rely upon assumptions regarding the structure and/or distribution of preferences.

This section provides an overview of equilibrium sorting models. Much of the discussion draws on a recent and more comprehensive review of this literature by Kuminoff, Smith, and Timmins (2013). Attention here is focused on two particular sorting model frameworks: (a) the *pure characteristics sorting model* developed by Epple and Sieg (1999) and (b) the *random utility sorting model* developed by Bayer, McMillan, and Reuben (2004).[5] These are sometimes referred to in the literature as the *vertical* and *horizontal* sorting models, respectively. At the heart of both models is Tiebout's (1956) notion that households "vote with their feet," choosing among available communities both in terms of the private amenities that a specific house in the community would provide and the accompanying exposure to both positive and negative public goods. Specifically, let $\boldsymbol{G}_k$ denote the vector of public amenities provided by community $k = 1,\ldots, K$ and $\boldsymbol{H}_j$ denote the vector of private amenities provided by housing unit $j \in k$ in community k. In the notation of the previous section, the complete vector of amenities for housing unit j is then given by $\boldsymbol{Q}_j = \left(\boldsymbol{G}_k, \boldsymbol{H}_j\right)$. The individual is assumed to choose a community k, a housing unit $j \in k$ in that community, and a level of expenditures on a numeraire good z so as to maximize her utility subject to a budget constraint; that is,

$$\max_{k, j\in k, z} u\left(z, \boldsymbol{G}_k, \boldsymbol{H}_j; \alpha_i, \boldsymbol{S}_i\right) \qquad \text{s.t.} \quad z + R_{j\in k} = M_i, \tag{10.46}$$

where $R_{j\in k}$ denotes the annual rental expenditure for housing unit j in community k, α_i denotes unobserved individual characteristics influencing the individual's preferences and $\boldsymbol{S}_i$ denotes observable individual characteristics. In general, the individual's income might also depend upon the community chosen, but for ease of notation, this generalization is ignored here. The two modeling frameworks differ in the additional structure imposed on the consumer's optimization problem.

4 To be fair, of course, the hedonic pricing literature incorporates in theory many of these factors as well, as in Ekeland, Heckman, and Nesheim (2004), but has focused on the equilibrium price outcome as a practical matter.

5 Kuminoff, Smith, and Timmins (2013) also consider a third framework, which they refer to as the *Nechyba–Ferreyra model*, developed by Ferreyra (2007) and based on theoretical models of Nechyba (1997, 1999, 2000). This approach, however, has largely been applied to modeling housing and school choices, and the interaction of these choices with the production of education.

The Pure Characteristics Sorting Model

The vertical sorting model begins by assuming that the annual rental expenditures associated with living in housing unit *j* can be decomposed into a housing quantity index $h_j = h(\boldsymbol{H}_j)$, reflecting the private attributes of the unit, and a price index $p_k = p(\boldsymbol{G}_k)$, reflecting the cost of purchasing a home in community *k* (per unit of the housing index); that is, $R_{j\in k} = p_k h_j$. Thus, $h(\boldsymbol{H}_j)$ reflects the *amount* of housing one receives from a specific unit as it varies in physical attributes, such as square footage, number of bathrooms, etc., whereas $p(\boldsymbol{G}_k)$ captures how the price of the housing changes across communities with different levels of public amenities. In addition, it is traditionally assumed that the public good attributes of the community can be summarized by a one-dimensional index of the attribute vector $\boldsymbol{G}_k$ (i.e., $\tilde{g}_k = g(\boldsymbol{G}_k)$). Conditional on choosing community *k*, the individual agent's maximum utility becomes[6]

$$v_k = v(\tilde{g}_k, p_k, \alpha_i, M_i) \equiv \max_{h,z} u(z, \tilde{g}_k, h; \alpha_i) \qquad \text{s.t.} \quad z + p_k h = M_i . \tag{10.47}$$

In this context, α_i is treated as an indicator of the individual's preference for public goods; that is, all else equal, an individual with a high value of α_i will prefer to live in a community with more public goods.

The household is assumed to choose the community that maximizes its utility. Prices adjust to reflect the demand for housing in the community and the available stock of housing and public amenities. An equilibrium emerges when individuals are in their desired community and no longer wish to move. The problem is that, without further structure, it is not clear that such an equilibrium exists, or that it is unique.

The solution employed in the vertical sorting literature is the so-called *single crossing condition*, introduced by Ellickson (1971) and generalized by Epple and Platt (1998). The single crossing condition requires that indifference curves in $(\tilde{g}, p)$ space, implied by $v(\tilde{g}, p, \alpha, M)$ in equation (10.47), cross only once for different individuals. Formally, as Kuminoff, Smith, and Timmins (2013) note, the slope of these indifference curves is given by

$$F(\tilde{g}, p, \alpha, M) = -\frac{\partial v(\tilde{g}, p, \alpha, M)/\partial \tilde{g}}{\partial v(\tilde{g}, p, \alpha, M)/\partial p} . \tag{10.48}$$

The single crossing condition is satisfied if F is increasing in $(\alpha | M)$ and is increasing in $(M | \alpha)$. Figure 10.4 illustrates this condition for three individuals with the same preference parameter α, but different income levels. For any one individual, with their given level of income (*M*) and attitude towards public goods (α), utility is increasing as one moves down (due to lower housing prices) and to the right (due to higher levels of public goods). Individual *i* in the figure is

6 For simplicity, the observed individual attributes are dropped, subsumed for now into the preference parameter *i*.

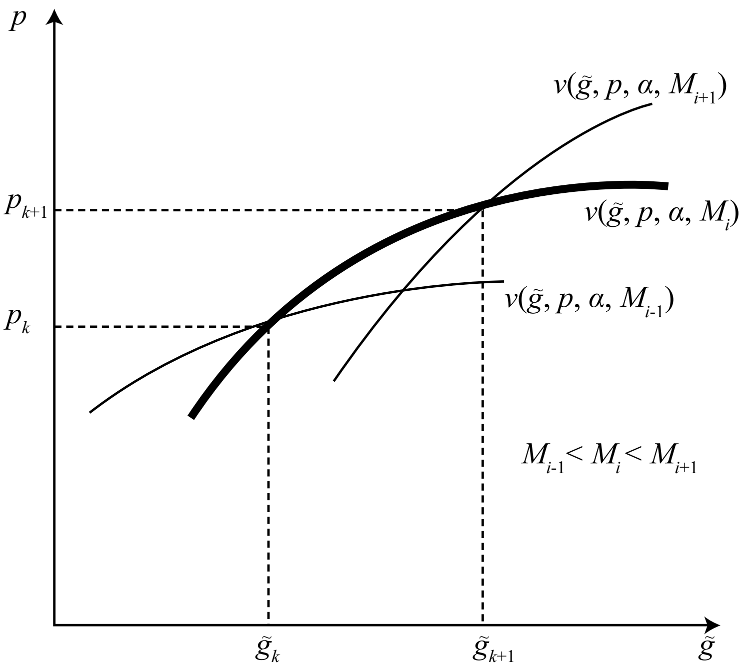

Figure 10.4 The single crossing condition

indifferent between communities k and $k + 1$, whereas individual $i - 1$ would prefer community k and individual $i + 1$ prefers community $k + 1$.

The single crossing condition yields three properties of the sorting equilibrium that are key to estimation: *boundary indifference*, *increasing bundles*, and *stratification* (Epple and Platt 1998). To define these properties, suppose that the K communities are ordered by the public good index $\tilde{g}_k$, with $\tilde{g}_1 < \tilde{g}_2 < \ldots < \tilde{g}_K$. Then the three properties are:

- Boundary indifference: There exists a "boundary" B_k between any two adjacent communities (k and $k + 1$) defined in (α, M)-space such that

$$B_k = \left\{ (\alpha, M) \middle| v\left(\tilde{g}_k, p_k, \alpha, M\right) = v\left(\tilde{g}_{k+1}, p_{k+1}, \alpha, M\right) \right\}. \tag{10.49}$$

  The term "adjacent" in this context refers to the communities' proximity to each other in the ordering from 1 to K, not necessarily their proximity geographically.
- Increasing bundles: The communities' rankings by the public good index match the rankings by price (i.e., $p_1 < p_2 < \cdots < p_K$).
- Stratification: For a given level of income, individuals sort themselves by preference parameter α; and individuals in communities $k - 1$, k, and $k + 1$ with the same income sort such that

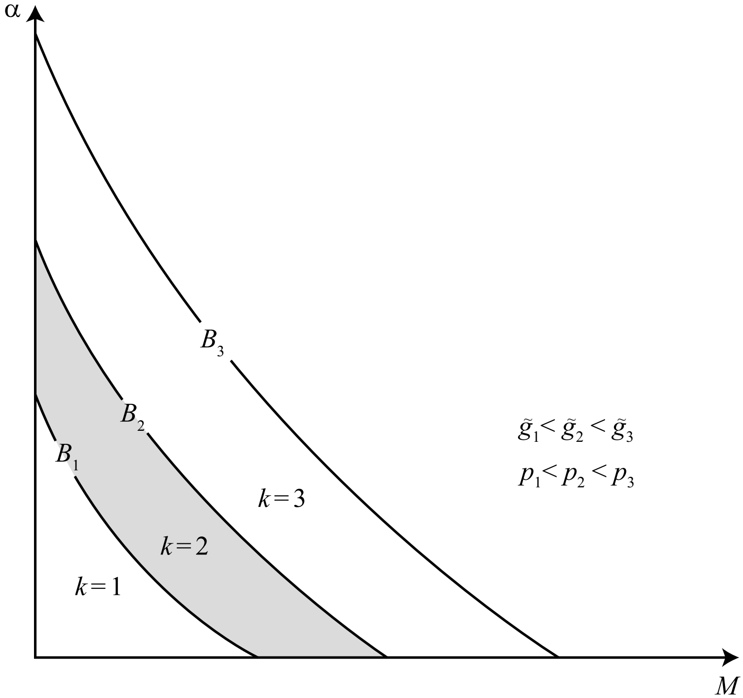

Figure 10.5 Market segmentation

$$\left(\alpha_{k-1}\middle|M\right)<\left(\alpha_k\middle|M\right)<\left(\alpha_{k+1}\middle|M\right). \tag{10.50}$$

Likewise, for a given level of the preference parameters, individuals sort themselves by income. In other words,

$$\left(M_{k-1}\middle|\alpha\right)<\left(M_k\middle|\alpha\right)<\left(M_{k+1}\middle|\alpha\right). \tag{10.51}$$

The implication of these three properties in terms of sorting is illustrated in Figure 10.5. The solid lines indicate the boundaries (B_k) between communities. The shaded region indicates the set of individuals (defined in terms of α and M) choosing to live in community k = 2. Note that the make-up of individuals in a given community is not defined solely in terms of either income (M) or solely in terms of preferences (i.e., α), but by their interaction. An individual with a relatively low income will choose to live in community 2 rather than the less expensive community 1 if they place a high enough value in the additional public goods that community 2 offers (that is, they have a high enough α). Conversely, an individual with a relatively high income may choose to live in community 2 rather than community 3 if they care relatively little for the available public goods (that is, they have a low enough α).

In order to close the model, additional structure is still needed. In particular, one needs a functional form for the indirect utility function in (10.47), an assumption regarding the joint distribution of income and preferences, and a functional form for the public goods index $\tilde{g}_k = g(\boldsymbol{G}_k)$. A commonly used specification for the indirect utility function is a constant elasticity of substitution (CES) form, with

$$v(\tilde{g}, p, \alpha, M) = \left\{ \alpha \cdot \tilde{g}^{\rho} + \left[\exp\left(\frac{M^{1-\upsilon} - 1}{1-\upsilon} \right) \exp\left(\frac{\beta \cdot p^{\eta+1}}{1+\eta} \right) \right]^{\rho} \right\}^{\frac{1}{\rho}}. \tag{10.52}$$

As Kuminoff, Smith, and Timmins (2013) noted, this structure implies that housing demand in the chosen community is given by

$$h = \beta p_k^{\eta} M^{\upsilon}, \tag{10.53}$$

so that η represents the price elasticity of housing demand and υ denotes the income elasticity of housing demand. The parameter ρ determines the substitutability between private and public goods.

The public good index is typically assumed to be a linear function of both the observed and unobserved public good attributes of a community; that is,

$$\tilde{g}_k = \gamma_1 g_{1,k} + \gamma_2 g_{2,k} + \cdots + \gamma_L g_{L,k} + \xi_k, \tag{10.54}$$

where $g_{\ell,k}$ denotes the ℓth observed public good for community k and ξ_k is a composite of all the unobserved attributes of community k. An important characteristic of this index is that it does not vary by individual. This implicitly assumes that all individuals weigh the component elements of the public good index the same. This is likely to be a strong assumption in some settings. For example, if one of the public goods is elementary education and another is public health care facilities, individuals with children may weigh these attributes differently than those who are single or elderly. In other settings, the assumption may be more innocuous, but it represents an important limitation of the vertical sorting model as it has been implemented to date and an area where additional research is needed.

The last element in specifying the model is to choose a joint distribution for the preference parameter and income pair (α, M). These are typically assumed to be from a joint lognormal distribution. Parameters of this distribution are inferred, in part, by comparing the observed income distributions in individual communities with those implied by the stratification boundaries (such as for community 2 in Figure 10.5).

Details of the econometric procedures required to estimate vertical sorting models are not described here, as they have varied somewhat across individual applications. Kuminoff, Smith, and Timmins (2013) provide a general discussion of some of the approaches used. To get a sense of at least part of the process, consider a stylized and simple version of Figure 10.5 depicted in Figure 10.6, in which it is assumed that there are only three communities, $k = 1, 2, 3$. Furthermore, suppose that *both* income (M) and preferences for public goods (α) are observable and known to be independently and uniformly distributed in the population; that is, $M \sim U[0, \bar{M}]$ and $\alpha \sim [0, \bar{\alpha}]$. Finally, suppose that the shaded region in Figure 10.6 depicts those individuals actually living in community 2. According to the

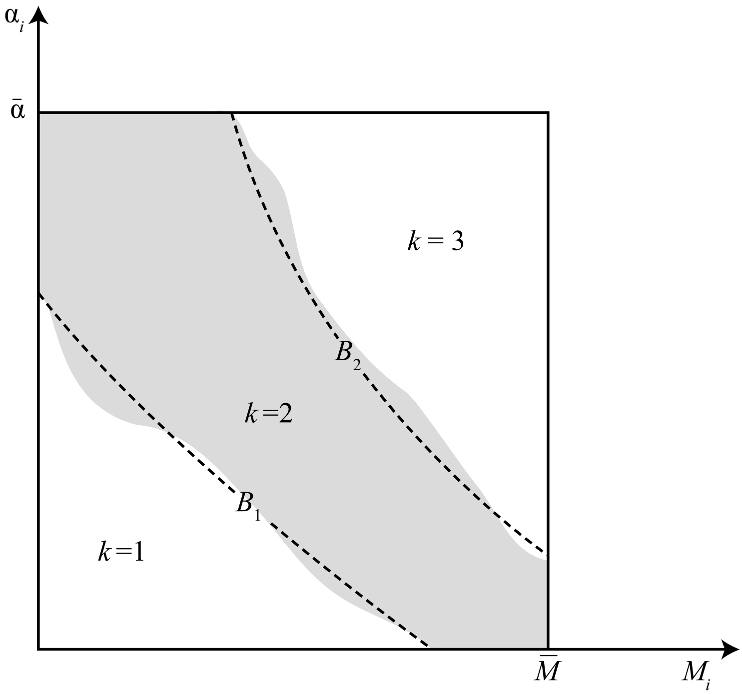

Figure 10.6 Estimating the parameters of a vertical sorting model

vertical sorting model outlined above, the proportion of the population living in community 2 is dictated by the region between the boundaries B_1 and B_2 defined in equation (10.10). Given observed housing prices in the three communities, these boundaries are in turn determined by the preference parameters in (10.13); that is, (β, η, ν, ρ). Estimation of these parameters involves adjusting them so that the resulting estimated boundaries (the dashed lines in Figure 10.6) imply as closely as possible the observed distribution of individuals in the three communities.

With the parameter estimates in hand, the general equilibrium implications of a policy changing the levels of public goods can be evaluated. Changes in the public goods will cause a shift in the community boundaries, which will in turn lead to a mismatch between existing housing supply and the new levels of demand in each community. The new equilibrium is constructed by finding the levels of housing prices that restore supply and demand in each community. Evaluation of the impact of the proposed policy can then include the overall impact of the policy and distribution of its impact on various subpopulations.

Individual applications include Epple and Sieg's (1999) study of school quality and public safety in the Boston metropolitan area, Smith et al.'s (2004) examination of general versus partial equilibrium implications of air quality improvements in the Los Angeles basin, and Walsh's (2007) analysis of open space as an endogenous public amenity.

The Random Utility Sorting Model

The second locational sorting model, initially developed by Bayer, McMillan, and Reuben (2004), builds on the discrete choice literature and McFadden's (1974) RUM model. Indeed, the starting point for this approach is essentially the same as the discrete choice RUM model discussed above at the end of the hedonic pricing model section. Specifically, an individual is assumed to face a discrete set of alternatives in their choice set, defined in terms of communities ($k = 1,\ldots,K$) and housing units (or housing types) available within each community (that is, $j \in k, j = 1,\ldots,J_k$). The conditional utility that individual i receives from choosing housing unit j in community k is assumed to take the form[7]

$$\begin{aligned} u^i_{j\in k} &= \boldsymbol{\alpha}^i_H \boldsymbol{H}_{j\in k} + \boldsymbol{\alpha}^i_G \boldsymbol{G}_k + \alpha_R R_{j\in k} + \xi_{j\in k} + \varepsilon^i_{j\in k} \\ &= v^i_{j\in k} + \varepsilon^i_{j\in k}\ , \end{aligned} \tag{10.55}$$

where $\boldsymbol{\alpha}^i_H$ and $\boldsymbol{\alpha}^i_G$ are vectors of individual-specific parameters associated with the respective vectors of housing and community characteristics, $\xi_{j\in k}$ represents unobserved attributes of the housing unit j (including potentially unobserved community characteristics), and $\varepsilon^i_{j\in k}$ denotes idiosyncratic unobservable factors influencing $u^i_{j\in k}$. The individual-specific parameters are typically modeled as functions of observable individual attributes (i.e., such as age, gender, etc. denoted by $\boldsymbol{S}_i$ in equation (10.46) above), with

$$\boldsymbol{\alpha}^i_a = \boldsymbol{\alpha}_{a0} + \sum_{\ell=1}^{L} \boldsymbol{\alpha}_{a\ell} s^i_\ell , \tag{10.56}$$

where s^i_ℓ denotes the ℓ th attribute in $\boldsymbol{S}_i$.

Substituting (10.56) into (10.55) and collecting terms, we can rewrite (10.55) as

$$u^i_{j\in k} = \alpha_{j\in k} + \tilde{\boldsymbol{\alpha}}^i_H \boldsymbol{H}_{j\in k} + \tilde{\boldsymbol{\alpha}}^i_G \boldsymbol{G}_k + \varepsilon^i_{j\in k} , \tag{10.57}$$

where

$$\alpha_{j\in k} = \boldsymbol{\alpha}_{H0} \boldsymbol{H}_{j\in k} + \boldsymbol{\alpha}_{G0} \boldsymbol{G}_k + \alpha_{R0} R_{j\in k} + \xi_{j\in k} \tag{10.58}$$

denotes a housing-type specific constant and

$$\tilde{\boldsymbol{\alpha}}^i_a = \sum_{\ell=1}^{L} \boldsymbol{\alpha}_{a\ell} s^i_\ell . \tag{10.59}$$

Individuals are assumed to choose a community (k) and housing unit ($j \in k$) that maximizes their utility.

Before getting into either the equilibrium sorting of households or the econometric issues associated with estimating the discrete choice model in (10.57) and (10.58), there are several of its features that are worth highlighting. First, unlike in the vertical sorting model, individuals are able to have differing preferences

7 One can also make the coefficient on rental costs (that is, αR) a function of individual observable attributes (that is, $\boldsymbol{S}i$), but this complication is not considered here in order to simplify the exposition.

regarding both the observable housing characteristics ($\boldsymbol{H}_{j\in k}$) and the community level public goods ($\boldsymbol{G}_k$). The baseline (or common) impact of each attribute is captured by either $\boldsymbol{\alpha}_{H0}$ or $\boldsymbol{\alpha}_{G0}$, but $\tilde{\boldsymbol{\alpha}}_H^i$ and $\tilde{\boldsymbol{\alpha}}_G^i$ allow for differences in the marginal utility of an attribute depending on socio-demographic characteristics. Thus, families with children can value local public schools and health care facilities differently than those who are single or elderly. Recall that in the vertical sorting model, everyone is assumed to weigh the community's public goods in the same manner (see equation 10.54). Second, the horizontal sorting model can readily handle discrete quality attributes for both communities and housing units. This is in contrast to the standard hedonic pricing model, wherein individuals are assumed to face a continuum of housing units in all quality dimensions. Third, the structure can also allow for friction in the sorting process in the form of moving costs, for example by incorporating a fixed cost associated with alternatives in either different communities or different cities (see, for example, Bayer, Koehane, and Timmins 2009). Finally, notice that the conditional indirect utilities in (10.57) include an alternative specific constant (ASC) $\alpha_{j\in k}$. These ASCs reflect the overall appeal of the alternative, which, as indicated by equation (10.58), depends upon the observed features of the housing unit ($\boldsymbol{H}_{j\in k}$), the public goods available in the community ($\boldsymbol{G}_k$), the cost of the unit ($R_{j\in k}$), and features of the unit not observed by the analyst ($\xi_{j\in k}$). The terms $\tilde{\boldsymbol{\alpha}}_H^i$ and $\tilde{\boldsymbol{\alpha}}_G^i$ capture how the appeal of the unit varies by observable individual attributes. The advantage of this structure is that it controls for a myriad of possible unobservable attributes for the housing unit. Unfortunately, it also means that there are as many ASCs to estimate as there are alternatives in the choice set. This creates a practical problem in terms of estimation, as the choice set can become large in individual applications. This has led to the use of a two-stage estimation procedure outlined below.

There are also econometric issues associated with insuring the consistency and asymptotic normality of the estimators used (see Berry, Linton, and Pakes 2004). This has led researchers to aggregate housing units into types or classes within each community. For example, Klaiber and Phaneuf (2010) organized housing units within communities by size (so that j references small, medium, and large housing types), whereas Tra (2010) organized housing units by ownership status, number of bedrooms, dwelling type, and when the housing unit was built.

Estimation of the model typically proceeds in two steps. If the $\varepsilon_{j\in k}^i$ are assumed to be independently and identically distributed according to a Type I extreme value distribution, then the probability that an individual i chooses housing unit (or housing type) j in community k is given by the usual logit structure, with

$$P_{j\in k}^i = \frac{\exp\left(v_{j\in k}^i\right)}{\sum_{m=1}^{K}\sum_{\ell=1}^{J_m}\exp\left(v_{\ell\in m}^i\right)}. \tag{10.60}$$

The alternative specific constants (i.e., the $\alpha_{j\in k}$) and the socio-demographic parameters $\tilde{\boldsymbol{\alpha}}_H^i$ and $\tilde{\boldsymbol{\alpha}}_G^i$ are recovered. One of the challenges at this stage of estimation is the potentially large number of alternative specific constants to be estimated. Researchers have typically drawn on a number of convenient features of the logit model to simplify estimation (see, for example, Murdock 2006 and Klaiber and Phaneuf 2010).

The second stage in the estimation process involves using the fitted alternative specific constants in order to estimate the parameters in (10.59). The key issue here is that the unobserved housing and community characteristics, represented by $\xi_{j\in k}$, are likely correlated with the observed housing and community characteristics, as well as the housing price itself (i.e., $R_{j\in k}$). In order to resolve this potential for omitted variables bias, the horizontal sorting literature has drawn on instrumental variables procedures developed in the industrial organization literature (for example, Berry, Levinsohn, and Pakes 1995).

Up to this point, the model is essentially no different from the discrete choice housing models discussed in the previous section. The difference arises when one recognizes that the choice probabilities themselves provide a measure of aggregate demand ($d_{j\in k}$) for housing by community and housing type. Specifically,

$$\begin{aligned} d_{j\in k} &= d_{j\in k}\left[\boldsymbol{H}_{j\in k}, \boldsymbol{G}_k, \alpha_{j\in k}\left(\boldsymbol{H}_{j\in k}, \boldsymbol{G}_k, R_{j\in k}\right)\right] \\ &= \sum_{i=1}^{N} P_{j\in k}^{i}\left[\boldsymbol{H}_{j\in k}, \boldsymbol{G}_k, \alpha_{j\in k}\left(\boldsymbol{H}_{j\in k}, \boldsymbol{G}_k, R_{j\in k}\right), \boldsymbol{S}_i\right]. \end{aligned} \tag{10.61}$$

Note that the price of housing ($R_{j\in k}$) has its impact in this specification entirely through the alternative specific constant. If $t_{j\in k}$ denotes the total available housing of type *j* in community *k*, then the equilibrium is characterized by

$$t_{j\in k} = d_{j\in k}\left[\boldsymbol{H}_{j\in k}, \boldsymbol{G}_k, \alpha_{j\in k}\left(\boldsymbol{H}_{j\in k}, \boldsymbol{G}_k, R_{j\in k}\right)\right]. \tag{10.62}$$

The general equilibrium impact of a change in the community public goods (for example, an improvement in air quality or public education, changing $\boldsymbol{G}_k$ to $\tilde{\boldsymbol{G}}_k$) is found by solving for the housing prices (say $\tilde{R}_{j\in k}$) that re-establish the equilibrium:

$$t_{j\in k} = d_{j\in k}\left[\boldsymbol{H}_{j\in k}, \tilde{\boldsymbol{G}}_k, \alpha_{j\in k}\left(\boldsymbol{H}_{j\in k}, \tilde{\boldsymbol{G}}_k, \tilde{R}_{j\in k}\right)\right]. \tag{10.63}$$

This involves solving for the new set of alternative specific constants

$$\tilde{\alpha}_{j\in k} = \alpha_{j\in k}\left(\boldsymbol{H}_{j\in k}, \tilde{\boldsymbol{G}}_k, \tilde{R}_{j\in k}\right), \tag{10.64}$$

that re-establishes the equilibrium in (10.63) and backs out the implied housing prices. Changes in welfare can then be inferred using standard log-sum formulas for evaluating changes in an RUM model. Changes in the composition of individuals living in a given community can be inferred by decomposing the housing demand in (10.61) by groups of interest (for example, the elderly, households with children, etc.).

There have been a number of applications of the horizontal sorting model to date, including Klaiber and Phaneuf's (2010) analysis valuing open space in the Twin Cities, Tra's (2010) examination of the impact in the Los Angeles basin of the 1990 Clean Air Act Amendments, and Bayer, Ferreira, and McMillan's (2007) consideration of how school quality impacts housing and community selection. As noted above, there are a number of appealing features of the horizontal sorting approach to valuing quality-differentiated commodities. There are, of course, limitations as well. First, much of this literature has relied on the standard logit structure in modeling choice probabilities. This specification makes the econometrics substantially easier, but requires the rather strong Independence of Irrelevant Alternatives (IIA) assumption (discussed in Chapter 3). In particular, it requires that the idiosyncratic error terms in (10.57) (that is, $\varepsilon^{i}_{j\in k}$) are uncorrelated across the choice options for a given individual. Since these terms capture unobserved individual specific factors influencing housing choice, this can be a strong assumption, depending upon how much information is observed about individual (and housing) characteristics. Relaxing this structure and using a nested logit or mixed logit model would seem worth consideration. Second, defining the choice set may be difficult, much like it is in the context of recreation demand. The housing units considered by individuals may span differing time periods (i.e., they may be looking for a house over a short or long time period) and communities. Banzhaf and Smith (2007) provided an excellent discussion of this issue, examining the impact of a wide range of possible choice sets. Their analysis, however, focused on the use of a single choice set across individuals, whereas the choice set itself may be individual specific. More research is needed into defining the choice set.

Summary

Hedonic price theory provides a coherent basis for explaining housing prices as a function of the levels of characteristics embedded in each house, including the environmental amenities and disamenities determined by the housing unit's location. Measures of value for marginal and nonmarginal changes in local public goods can be derived from a properly specified hedonic price model. Values for marginal changes in amenity levels are found simply by adding up the observed or computed marginal willingness to pay for all affected individuals. However, for nonmarginal amenity changes, when the hedonic price function itself might shift, welfare measurement requires knowledge of the inverse demand function or the income-compensated bid function for the amenity. These, in turn, require a solution to the daunting identification problem. As a result, much of the empirical literature has limited its attention to the so-called first stage of hedonic analysis—that is, estimating the hedonic price equation.

Beyond the difficulty in estimating both stages of the hedonic price model, and hence using it to assess nonmarginal shifts in amenities, there are some additional limitations to the hedonic property value model for use in estimating welfare effects.

First, the hedonic model assumes that consumers of housing can select their most preferred bundle of characteristics from a complete range of levels of all characteristics. In practice, the available housing units are finite both in numbers and in individual housing characteristics (for example, number of bathrooms and bedrooms, styles, etc.). This weakens the theoretical connection, depicted in equation (10.4), between the marginal cost of a housing attribute and the individual's marginal willingness to pay for the attribute that lies at the heart of using hedonic price analysis for welfare evaluation.

Second, since the property value models are based on the consequences of individuals' choices of residence, they do not capture willingness to pay for improvements in environmental amenities at other points in the urban area—for example, in the work place, shopping areas, or parks and recreational areas, or at second homes (see for example, Smith 2007).

Third, because the property value models are based on observing behavioral responses to differences in amenity levels across houses, they only capture willingness to pay for perceived differences in amenities and their consequences. For example, if there are subtle, long-term health effects associated with reduced environmental quality at some housing sites, but people are unaware of the causal link between these effects and the housing site, their willingness to pay to avoid the effects will not be reflected in housing price differences. Despite these limitations, the traditional hedonic price model provides a valuable tool in assessing or bounding the welfare implications of changing environmental amenities. It should also be noted that, while most applications of the hedonic model have used residential property values, the technique is also applicable to commercial and agricultural properties. For example, Mendelsohn, Nordhaus, and Shaw (1994) provided an analysis of the impact of climate on agricultural land values with a goal of assessing the effects of global warming on agriculture.

The equilibrium sorting literature has emerged in recent years as a means of addressing some of the issues associated with the hedonic pricing model. Both the vertical and horizontal sorting models allow for evaluation of nonmarginal shifts in public goods, including environmental amenities, by modeling the equilibrium process itself. This is particularly important in that many policy scenarios, such as those resulting from the 1990 Clean Air Act Amendments, involve significant changes in the environment that are likely to alter the housing market's equilibrium. Many of the applications to date suggest that ignoring general equilibrium effects of a policy (changing the hedonic pricing function and the resorting of households) can significantly bias the estimated welfare impacts of that policy. The horizontal sorting model also provides resolutions to two other issues in the hedonic price model by allowing for (a) a discrete, rather than continuous, stock of available housing units and (b) frictions in the housing market stemming from moving costs. The gains from the equilibrium sorting models, of course, come at a cost, including the need for additional assumptions regarding the structure of consumer preferences and the choice set available to consumers. At the same time, they represent a promising avenue for future research in efforts to understand how environmental amenities are capitalized into the housing market.

References

Anderson, Robert J., and Thomas D. Crocker. 1972. "Air Pollution and Property Values: A Reply." *Review of Economics and Statistics* 54(4): 470–473.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Atkinson, Scott E., and Thomas D. Crocker. 1987. "A Bayesian Approach to Assessing the Robustness of Hedonic Property Value Studies." *Journal of Applied Econometrics* 2(1): 27–45.

Banzhaf, H. Spencer, and V. Kerry Smith. 2007. "Meta-analysis in Model Implementation: Choice Sets and the Valuation of Air Quality Improvements." *Journal of Applied Econometrics* 22(6): 1013–1031.

Bartik, Timothy J. 1987. "The Estimation of Demand Parameters in Hedonic Price Models." *Journal of Political Economy* 95(1): 81–88.

Bartik, Timothy J. 1988. "Measuring the Benefits of Amenity Improvements in Hedonic Price Models." *Land Economics* 64(2): 172–183.

Bartik, Timothy J., and V. Kerry Smith. 1987. "Urban Amenities and Public Policy." In *Handbook of Regional and Urban Economics*, edited by Edwin S. Mills. Amsterdam: Elsevier.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115(4): 588–638.

Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins. 2009. "Migration and Hedonic Valuation: The Case of Air Quality." *Journal of Environmental Economics and Management* 58(1): 1–14.

Bayer, Patrick, Robert McMillan, and Kim Reuben. 2004. "An Equilibrium Model of Sorting in an Urban Housing Market." NBER Working Paper No. 10865. Available at: www.nber.org/papers/w10865.

Bell, Kathleen P., and Nancy E. Bockstael. 2000. "Applying the Generalized Method of Moments Approach to Spatial Problems Involving Micro-Level Data." *Review of Economics and Statistics*. 82(1): 72–82.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63(4): 841–890.

Berry, Steven, Oliver B. Linton, and Ariel Pakes. 2004. "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems." *Review of Economic Studies* 71(3): 613–654.

Bockstael, Nancy E. 1996 "Modeling Economics and Ecology: The Importance of a Spatial Perspective." *American Journal of Agricultural Economics* 78(5): 1168–1180.

Brown, James N., and Harvey S. Rosen. 1982. "On the Estimation of Structural Hedonic Price Models." *Econometrica* 50(3): 765–768.

Cameron, Trudy A. 2006 "Directional Heterogeneity in Distance Profiles in Hedonic Property Value Models." *Journal of Environmental Economics and Management* 51(1): 26–45.

Can, Ayle. 1992. "Specification and Estimation of Hedonic Price Models." *Regional Science and Urban Economics* 22(3): 453–474.

Cassel, Eric, and Robert Mendelsohn. 1985. "The Choice of Functional Forms for Hedonic Price Equations: Comment." *Journal of Urban Economics* 18(2): 135–142.

Chattopadhyay, Sudip. 1998. "An Empirical Investigation into the Performance of Ellickson's Random Bidding Model, with an Application to Air Quality Valuation." *Journal of Urban Economics* 43(2): 292–314.

Chattopadhyay, Sudip. 1999. "Estimating the Demand for Air Quality: New Evidence Based on the Chicago Housing Market." *Land Economics* 75(1): 22–38.

Chattopadhyay, Sudip. 2000. "The Effectiveness of McFadden's Nested Logit Model in Valuing Amenity Improvement." *Regional Science and Urban Economics* 30(1): 23–43.

Chay, Kenneth Y., and Michael Greenstone. 2005. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy* 113(2): 376–424.

Colwell, Peter F., and Gene Dilmore. 1999. "Who Was First? An Examination of an Early Hedonic Study." *Land Economics* 75(4): 620–626.

Cropper, Maureen L., Leland B. Deck, Nalin Kishor, and Kenneth E. McConnell. 1993. "Valuing Product Attributes Using Single Market Data: A Comparison of Hedonic and Discrete Choice Approaches." *Review of Economics and Statistics* 75(2): 225–232.

Cropper, Maureen L., Leland B. Deck, and Kenneth E. McConnell. 1988. "On the Choice of Functional Form for Hedonic Price Functions." *Review of Economics and Statistics* 70(4): 668–675.

Ekeland, Ivar, James J. Heckman, and Lars Nesheim. 2004. "Identification and Estimation of Hedonic Models." *Journal of Political Economy* 112(S1): S60–S109.

Ellickson, Bryan. 1971. "Jurisdictional Fragmentation and Residential Choice." *American Economic Review* 61(2): 334–339.

Ellickson, Bryan. 1981. "An Alternative Test of the Hedonic Theory of Housing Markets." *Journal of Urban Economics* 9(1): 56–79.

Epple, Dennis. 1987. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." *Journal of Political Economy* 87(1): 59–80.

Epple, Dennis, and Glenn J. Platt. 1998. "Equilibrium and Local Redistribution in an Urban Economy when Households Differ in both Preferences and Incomes." *Journal of Urban Economics* 43(1): 23–51.

Epple, Dennis, and Holger Sieg. 1999. "Estimating Equilibrium Models of Local Jurisdiction." *Journal of Political Economy* 107(4): 645–681.

Ferreyra, Maria Marta. 2007. "Estimating the Effects of Private School Vouchers in Multi-District Economies." *American Economic Review* 97(3): 789–817.

Freeman, A. Myrick, III. 1971. "Air Pollution and Property Values: A Methodological Comment." *Review of Economics and Statistics* 53(4): 415–416.

Freeman, A. Myrick, III. 1974a. "Air Pollution and Property Values: A Further Comment." *Review of Economics and Statistics* 56(4): 454–456.

Freeman, A. Myrick, III. 1974b. "On Estimating Air Pollution Control Benefits from Land Value Studies." *Journal of Environmental Economics and Management* 1(1): 74–83.

Freeman, A. Myrick, III. 2003. *The Measurement of Environmental and Resource Values: Theory and Methods*. Washington, D.C.: Resources for the Future Press.

Goodman, Allen C. 1978. "Hedonic Prices, Price Indices and Housing Markets." *Journal of Urban Economics* 5(4): 471–484.

Graves, Phil, James C. Murdoch, Mark A. Thayer, and Don Waldman. 1988. "The Robustness of Hedonic Price Estimation: Urban Air Quality." *Land Economics* 64(3): 220–233.

Greenstone, Michael, and Justin Gallagher. 2008. "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program." *Quarterly Journal of Economics* 123(3): 951–1003.

Gross, David J. 1988. "Estimating Willingness to Pay for Housing Characteristics: An Application of the Ellickson Bid-Rent Model." *Journal of Urban Economics* 24(1): 95–112.

Halvorsen, Robert, and Henry O. Pollakowski. 1981. "Choice of Functional Form for Hedonic Price Equations." *Journal of Urban Economics* 10(1): 37–49.

Herriges, Joseph A., Silvia Secchi, and Bruce A. Babcock. 2005. "Living with Hogs in Iowa: The Impact of Livestock Facilities on Rural Residential Property Values." *Land Economics* 81(4): 530–545.

Horowitz, Joel L. 1984. "Estimating Compensating and Equivalent Income Variations from Hedonic Price Models." *Economics Letters* 14(4): 303–308.

Horowitz, Joel L. 1986. "Bidding Models of Housing Markets." *Journal of Urban Economics* 20(2): 168–190.

Horowitz, Joel L. 1987. "Identification and Stochastic Specification in Rosen's Hedonic Price Model." *Journal of Urban Economics* 22(2): 165–173.

Imbens, Guido M., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 4**7**(1): 5–86.

Kanemoto, Yoshitsugu. 1988. "Hedonic Prices and the Benefits of Public Projects." *Econometrica* 56(4): 981–989.

Kiel, Katherine A., 1995. "Measuring the Impact of the Discovery and Cleaning of Identified Hazardous Waste Sites on House Values." *Land Economics* 71(4): 428–435.

Kiel, Katherine A., and Katherine T. McClain. 1995. "House Prices During Siting Decision Stages: The Case of an Incinerator from Rumor Through Operation." *Journal of Environmental Economics and Management* 28(2): 241–255.

Kiel, Katherine A., and Jeffrey E. Zabel. 1999. "The Accuracy of Owner-Provided House Values: The 1978–1991 American Housing Survey." *Real Estate Economics* 27(2): 263–298.

Klaiber, H. Allen, and Daniel J. Phaneuf. 2010. "Valuing Open Space in a Residential Sorting Model of the Twin Cities." *Journal of Environmental Economics and Management* 60(2): 57–77.

Kohlhase, Janet E., 1991. "The Impact of Toxic Waste Sites on Housing Values." *Journal of Urban Economics* 30(1): 1–26.

Kuminoff, Nicholai V., and Jaren C. Pope. 2012. "Do 'Capitalization Effects' for Public Goods Reveal the Public's Willingness to Pay." Working paper, Arizona State University. Available at: www.public.asu.edu/~nkuminof/KP12.pdf.

Kuminoff, Nicholai V., Christopher F. Parmeter, and Jaren C. Pope. 2010. "Which Hedonic Models Can We Trust to Recover the Marginal Willingness to Pay for Environmental Amenities." *Journal of Environmental Economics and Management* 60(3): 145–160.

Kuminoff, Nicholai V., V. Kerry Smith, and Christopher Timmins. 2013. "The New Economics of Equilibrium Sorting and Policy Evaluation Using Hedonic Markets." *Journal of Economic Literature* 51(4): 1007–1062.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Economic Literature* 48(2): 281–355.

Lerman, Steven R., and Clifford R. Kern. 1983. "Hedonic Theory, Bid Rents, and Willingness-to-pay: Some Extensions of Ellickson's Results." *Journal of Urban Economics* 13(3): 358–363.

Lind, Robert C. 1973. "Spatial Equilibrium, the Theory of Rents, and the Measurement of Benefits from Public Programs." *Quarterly Journal of Economics* 87(2): 188–207.

Mäler, Kärl-Göran. 1977. "A Note on the Use of Property Values in Estimating Marginal Willingness to Pay for Environmental Quality." *Journal of Environmental Economics and Management* 4(4): 355–369.

McConnell, K.E., and T.T. Phipps. 1987. "Identification of Preference Parameters in Hedonic Models: Consumer Demands with Nonlinear Budgets." *Journal of Urban Economics* 22(1): 35–52.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by Paul Zarembka. New York: Academic Press.

Mendelsohn, Robert. 1984. "Estimating the Structural Equations of Implicit Markets and Household Production Functions." *Review of Economics and Statistics* 66(4): 673–677.

Mendelsohn, Robert. 1985. "Identifying Structural Equations with Single Market Data." *Review of Economics and Statistics* 67(3): 525–529.

Mendelsohn, Robert. 1987. "A Review of Identification of Hedonic Supply and Demand Functions." *Growth and Change* 18(1): 82–92.

Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw. 1994. "The Impact of Global Warming on Agriculture: A Ricardian Analysis." *American Economic Review* 84(4): 753–771.

Murdoch, James C., and Mark A. Thayer. 1988. "Hedonic Price Estimation of Variable Urban Air Quality." *Journal of Environmental Economics and Management* 15(2): 143–146.

Murdock, Jennifer. 2006. "Handling Unobserved Site Characteristics in Random Utility Models of Recreation Demand." *Journal of Environmental Economics and Management* 51(1): 1–25.

Nechyba, Thomas, J. 1997. "Existence of Equilibrium and Stratification in Local and Hierarchical Tiebout Economies with Property Taxes and Voting." *Economic Theory* 10(2): 277–304.

Nechyba, Thomas, J. 1999. "School Finance Induced Migration and Stratification Patterns: The Impact of Private School Vouchers." *Journal of Public Economic Theory* 1(1): 5–50.

Nechyba, Thomas, J. 2000. "Mobility, Targeting, and Private-School Vouchers." *American Economic Review* 90(1): 130–146.

Niskanen, William A., and Steve H. Hanke. 1977. "Land Prices Substantially Underestimate the Value of Environmental Quality." *Review of Economics and Statistics* 59(3): 375–377.

Palmquist, Raymond B. 1982. "Measuring Environmental Effects on Property Values Without Hedonic Regressions." *Journal of Urban Economics* 11(3): 333–347.

Palmquist, Raymond B. 1984. "Estimating the Demand for Characteristics of Housing." *Review of Economics and Statistics* 64(3): 394–404.

Palmquist, Raymond B. 1991. "Hedonic Methods." In *Measuring the Demand for Environmental Improvement*, edited by John B. Braden and Charles D. Kolstad. Amsterdam: Elsevier.

Palmquist, Raymond B. 1992a. "Valuing Localized Externalities." *Journal of Urban Economics* 31(1): 59–68.

Palmquist, Raymond B. 1992b. "A Note on Transactions Costs, Moving Costs, and Benefit Measurement." *Journal of Urban Economics* 32(1): 40–44.

Palmquist, Raymond B. 2005. "Property Value Models." In *Handbook of Environmental Economics*, edited by Karl-Göran Mäler and Jeffery R. Vincent. Amsterdam: Elsevier.

Palmquist, Raymond B., and Adis Israngkura. 1999. "Valuing Air Quality with Hedonic and Discrete Choice Models." *American Journal of Agricultural Economics* 81(5): 1128–1133.

Palmquist, Raymond B., Fritz M. Roka, and Tomislav Vukina. 1997. "Hog Operations, Environmental Effects, and Residential Property Values." *Land Economics* 73(1): 114–124.

Parsons, George R. 1990. "Hedonic Prices and Public Goods: An Argument for Weighting Locational Attributes in Hedonic Regressions by Lot Size." *Journal of Urban Economics* 27(3): 308–321.

Pines, David, and Yoram Weiss. 1976. "Land Improvement Projects and Land Values." *Journal of Urban Economics* 3(1): 1–13.

Polinsky, A. Mitchell, and Steven Shavell. 1976. "Amenities and Property Values in a Model of an Urban Area." *Journal of Public Economics* 5(1–2): 119–129.

Quigley, John M. 1982. "Nonlinear Budget Constraints and Consumer Demand: An Application to Public Programs for Residential Housing." *Journal of Urban Economics* 12(2): 177–201.

Ridker, Ronald G. 1967. *Economic Costs of Air Pollution: Studies in Measurement*. New York: Praeger.

Ridker, Ronald G., and John A. Henning. 1967. "The Determinants of Residential Property Values with Special Reference to Air Pollution." *Review of Economics and Statistics* 49(2): 246–257.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Perfect Competition." *Journal of Political Economy* 82(1): 34–55.

Smith, V. Kerry. 2007. "Reflections on the Literature." *Review of Environmental Economics and Policy* 1(1): 152–165.

Smith, V. Kerry, and Ju Chin Huang. 1995. "Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models." *Journal of Political Economy* 103(1): 209–227.

Smith, V. Kerry, Holger Sieg, H. Spencer Banzhaf, and Randall P. Walsh. 2004. "General Equilibrium Benefits for Environmental Improvements: Projected Ozone Reductions under EPA's Prospective Analysis for the Los Angeles Air Basin." *Journal of Environmental Economics and Management* 47(3): 559–584.

Sonstelie, Jon C., and Paul R. Portney. 1980. "Gross Rents and Market Values: Testing the Implications of the Tiebout Hypothesis." *Journal of Urban Economics* 7(1): 102–118.

Straszheim, Mahlon. 1974. "Hedonic Estimation of Housing Market Prices: A Further Comment." *Review of Economics and Statistics* 56(3): 404–406.

Strotz, Robert H. 1968. "The Use of Land Value Changes to Measure the Welfare Benefits of Land Improvements." In *The New Economics of Regulated Industries*, edited by Joseph E. Haring. Los Angeles: Occidental College.

Taylor, Laura O. 2003. "The Hedonic Method." In *A Primer on Nonmarket Valuation*, edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Dordrecht: Kluwer Academic Press.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64(5): 416–424.

Tra, Constant I., 2010. "A Discrete Choice Equilibrium Approach to Valuing Large Environmental Changes." *Journal of Public Economics* 94(1–2): 183–196.

Walsh, Randall L. 2007. "Endogenous Open Space Amenities in a Locational Equilibrium." *Journal of Urban Economics* 61(2): 319–344.

Zabel, Jeffrey E., and Katherine A. Kiel. 2000. "Estimating the Demand for Air Quality in Four U.S. Cities." *Land Economics* 76(2): 174–194.

Chapter 11

Hedonic Wage Models

The hedonic wage model is a formalization of the concept of compensating wage differentials, which can be traced back to Adam Smith. The basic idea is that, other things being equal, workers will prefer jobs with more pleasant working conditions as opposed to those that are less pleasant. The greater supply of workers for pleasant jobs will depress the wage levels of such jobs. In equilibrium, the difference in wages between two jobs with different working conditions will reflect the workers' monetary valuations of these differences.

The basic hedonic wage model has been refined and has been applied empirically to two important questions of particular interest to environmental and resource economists and policymakers. One question concerns the value of reducing the risk of death, injury, or illness. The hedonic wage model has been used to estimate the wage–risk tradeoff as a revealed preference measure of this value. The other question concerns the values of the environmental and social amenities that vary across regions. Wage differences across regions have been used as indicators of the values of region-specific environmental, cultural, and social amenities.

From a worker's perspective, a job can be viewed as a differentiated product; that is, a good with a bundle of characteristics such as working conditions, prestige, training and enhancement of skills, and levels of risk of accidental injury and exposure to toxic substances. If workers are free to move from one urban area to another, then jobs are also differentiated, in part, by the environmental and other characteristics of the urban areas in which the jobs are located. If workers are free to choose from a menu of differentiated jobs, the hedonic price technique can be applied to the data on wages, job characteristics (including their locations), and worker characteristics to estimate the marginal implicit prices of these job characteristics.

Employers, from their perspective, can be viewed as choosing from among a set of workers of different characteristics. This is a distinguishing feature of labor markets. In the typical application of the hedonic theory to differentiated goods, producers are viewed as selling a good embodying a package of characteristics and as being indifferent to the characteristics of the purchaser of the good. In hedonic wage studies, the employer is viewed as selling a package of job characteristics (including the quality of the work environment); but at the same time the employer is purchasing work effort and cannot be indifferent to the productive characteristics of the firm's employees. Thus, the hedonic wage equation must be interpreted as an equilibrium relationship that reflects not only the interaction of supply and demand for job characteristics, but also the interaction of supply and demand of worker characteristics (see Lucas 1977 and Rosen 1979). This means that both worker and job characteristics must be included as arguments in the estimated hedonic wage equation.

As in the case of hedonic property values, the derivative of the hedonic wage function with respect to any job characteristic can be interpreted as the marginal implicit price of that characteristic. If the worker is maximizing utility, the marginal implicit price can be taken as an estimate of the worker's marginal willingness to pay for the characteristic. It gives the change in income necessary to just compensate for a small change in the characteristic. Since in general the hedonic wage function need not be linear, these marginal values may be different for different workers. Similarly, the derivative of the hedonic wage function with respect to any worker characteristic gives its marginal implicit price and, assuming profit maximization, the marginal value of that characteristic to the employer.

In order to estimate the value of a nonmarginal change in a characteristic, it is necessary to know the compensated inverse demand function for it. As in the case of hedonic property values, the inverse demand function cannot be estimated from data from a single labor market unless additional restrictions are imposed. Some examples of efforts to identify these functions are described in the next section.

The interpretation of the hedonic wage function as revealing marginal implicit prices and marginal values requires that all of the transactions that make up the data be undertaken in the same market. In other words, each buyer (seller) in the market must have had the opportunity to match up with any of the other sellers (buyers) and to choose the most preferred given prices, and so forth. In the terminology of Chapter 10, the market must be in equilibrium and must not be segmented into submarkets with incomplete mobility among segments. When hedonic wage equations are estimated using data from several urban areas, it is necessary to assume that these areas are part of a single market. In practice, labor markets can be segmented on the basis of geography, with moving costs and lack of information on job alternatives imposing barriers between labor markets in different parts of the country. Markets can also be segmented on the basis of education and skill requirements, and between blue-collar and professional–managerial workers. Geographic segmentation can lead to different marginal implicit price schedules in different regions. Segmentation on the basis of occupation or education level can lead to different marginal implicit price functions across occupational categories. One approach to the problem of geographic segmentation is to estimate the hedonic wage function only for occupational groups that are believed, on a priori grounds, to be part of a national labor market. In general, the extent of market segmentation and its significance for empirical estimation of hedonic wage functions are not known.

The next section of this chapter discusses the application of the hedonic wage model to measure the value of risk reduction, and reviews the evolution of empirical applications of this model. Much of the recent literature calls into question earlier estimates of the marginal value of risk reduction, as well as their applicability to valuing the impact of environmental programs (see, for example, Cropper, Hammitt, and Robinson 2011, Black and Kniesner 2003, and Black, Galdo, and Liu 2003). At the same time, new data sources and econometric procedures hold out hope for resolving a number of these issues. This chapter then goes on in the second section to outline and describe some of the models that have been developed to explain and interpret interregional wage differences as reflections of the values people place on regional amenities. Included in this discussion is the recent introduction of equilibrium sorting models, described in Chapter 10, to the process of modeling the impact of local amenities on both labor and housing markets.

Wage Differences and the Value of Reducing Risks

Chapter 7 introduced the concept of the value of statistical life (VSL) and its role in estimating the value of policies that reduce the risk of death. It was also noted that inferences about the value of reductions in risk and the VSL could be extracted from revealed preferences (through the wage–risk tradeoff, for example). This section discusses the use of the hedonic wage model as one approach to estimating the individuals' willingness to pay for reductions in risk. Most of the applications of the hedonic model to risk valuation have dealt with risks of death due to accidents on the job.

Suppose that each individual (with attributes $\boldsymbol{S}$) chooses a job so as to maximize expected utility from consumption of the numeraire, z, and from the vector of job characteristics, $\boldsymbol{Q}$. In addition to $\boldsymbol{Q}$, each job is characterized by its risk of accidental death, δ. Individuals face a hedonic wage function that is the locus of points at which firms' marginal wage offers (as functions of job and worker characteristics) equal workers' marginal acceptance wages (see Chapter 4 for a description of equilibrium in hedonic markets). This function is

$$p_w = p_w(\delta, \boldsymbol{Q}, \boldsymbol{S}) \tag{11.1}$$

where p_w is the weekly or monthly wage and where hours of work per period could be one of the characteristics in $\boldsymbol{Q}$. Individual characteristics enter the hedonic wage function because they may impact both an individual's preferences and their productivity as a worker. The individual chooses a job to maximize expected utility subject to the wage constraint—that is

$$\max E[u] = \pi \cdot u(z, \boldsymbol{Q}, \boldsymbol{S}) + \lambda[p_w(\delta, \boldsymbol{Q}, \boldsymbol{S}) - z], \tag{11.2}$$

where π is the probability of surviving the period and being able to consume z. In wage–risk studies, it is risk of death, δ, rather than survival probability, π, which is observed. The relationship between the two is given by $\pi = (1-\delta)(1-\varphi)$, where φ is the probability of dying from non-work related causes. Since φ is usually small for the working-age population, π is approximately equal to $(1-\delta)$.

The first-order conditions governing the choices of z, job risk (δ), and job characteristics ($\boldsymbol{Q}$) are

$$\pi \cdot \left(\frac{\partial u}{\partial z}\right) = \lambda, \tag{11.3}$$

$$\frac{u(\cdot)}{\lambda} = \frac{\partial p_w}{\partial \delta}, \tag{11.4}$$

and

$$\frac{\pi \partial u / \partial Q}{\lambda} = -\frac{\partial p_w}{\partial Q}, \tag{11.5}$$

for all job characteristics, $\boldsymbol{Q}$.

From equation (11.3), λ is the expected marginal utility of consumption, which, by assumption, is positive. According to equation (11.4), the marginal willingness to pay for an increase in the probability of surviving the job risk must equal its marginal implicit price. Equation (11.4) also implies that wages must be lower for jobs that are safer; that is, the marginal implicit price of an increase in π is a decrease in the wage rate. Equation (11.5) requires that the marginal willingness to pay for each job characteristic equal its marginal implicit price.

If workers know the relationship between market wages and job attributes and risks, then each worker selects the collection of job attributes and risks that equates the marginal benefit of each attribute to its marginal cost. In the case of risk of death, the marginal cost of working in a less risky job is the lower wage received, $\partial p_w / \partial \delta$, and this must be equated with the marginal willingness to pay for lower job risk. In other words, the risk premium associated with a higher-risk job must be equal to the individual's marginal willingness to accept compensation for risk.

Estimating Marginal Values for Risk

Data on wages, job attributes, and worker attributes are used to estimate the hedonic wage function—an equilibrium relationship between wages on the one hand, and job characteristics and variables affecting worker productivity on the other. If the hedonic wage function can be estimated satisfactorily, the risk premium for a marginal change in risk can be calculated by evaluating the partial derivative of the function at a given risk level and set of individual and job attributes. However, if the wage–risk tradeoff locus is nonlinear, this marginal willingness to pay will vary with the baseline risk of each worker. For nonmarginal changes in risk, the value of the change to the individual cannot be calculated from the wage–risk tradeoff curve alone because of the convexity of the individual's indifference curve. For example, see Figure 4.5.

The task of estimating the hedonic price function is, of course, not without its challenges. Indeed, a number of recent reviews call into question a large portion of the existing literature. Cropper, Hammitt, and Robinson (2011, 331), for example, argue that "many of the older studies suffer from both data and econometric problems" and "it is time to replace the older set of studies with newer results." Fortunately, improvements in both the available data and econometric techniques have emerged in recent years, holding out hope for improved estimates of the marginal value of risk. This subsection provides an overview of the key issues in estimating the hedonic price function.

Measurement Error

Perhaps the biggest challenge in hedonic wage studies of risk is constructing the appropriate measure(s) of risk. Most studies rely upon objective measures of fatality risk, matched to individuals based on their occupation and/or the industry in which they work. These objective risk measures, however, are subject to a variety of potential measurement errors and may bear little resemblance to the risk perceptions of either the firms or the individuals that underlie the hedonic wage equilibrium.

One of the major questions in interpreting estimates of willingness to pay for risk reduction is whether individuals perceive differences in risks across jobs and, if so, whether these perceptions are accurate. Compensating wage differentials for risk can exist in the labor market only if workers perceive differences in risks across jobs. The absence of compensating differentials need not mean that workers do not value reducing the risk of accidental death, only that they are unaware of the differences in risks; and if individuals have inaccurate estimates of job risks, then risk premia can exist. However, these risk premia will yield biased estimates of individuals' marginal willingness to pay for risk reduction unless the researcher can identify what it was that individuals thought they were buying when they accepted a particular job with its bundle of characteristics. Since most hedonic wage studies use objective measures of job risk, it is important to find out if individuals' perceptions correspond well with these objective measures.

The only evidence pertaining directly to individuals' perceptions of job-related risks is found in research that compares workers' risk perceptions with data on frequency of job-related death and injury. This evidence suggests that workers' perceptions are positively correlated with objective risks, but may overstate them. Viscusi (1979) reported a positive correlation between a dichotomous risk variable (posed as a question: "Is your job dangerous?") and accident rate data from the Bureau of Labor Statistics (BLS). In a subsequent study, Viscusi and O'Connor (1984) reported that workers in the chemical industry perceived risk of injury on their job to be about 50 percent higher than BLS estimates. Gerking, de Haan, and Schulze (1988) also found that workers' perceptions of risk of death on the job overstated BLS accidental death rates. Liu and Hammitt (1999) found evidence that workers' risk perceptions evolve over time with their experience on the job, which would in turn influence their propensity to stay with a firm and the resulting hedonic price equilibrium.

Other studies providing evidence on the accuracy of risk perceptions compare relative frequency of deaths, by cause, with individuals' perceptions of these frequencies. Slovic, Fischhoff, and Lichtenstein (1979) found that, on average, people overestimate the likelihood of infrequent causes of death (for example, deaths due to botulism, floods, tornadoes) but underestimate the probability of deaths with higher frequencies (for example, deaths due to heart disease, cancer). However, Fischhoff et al. (1981) noted that one must distinguish between an individual's perception of the relative frequency of death in some population and the individual's estimate of his or her *own* risk of death. There is evidence that the latter is often underestimated (see Hamermesh 1985 and Fischhoff et al. 1981).

There is, of course, an analogous issue for firms, but there appears to be no research regarding the relationship between objective risk measures and firms' subjective perceptions of them. To the extent that the perceptions of firms and workers differ, it is not enough to include only one side's perceptions in the analysis, as it is the interaction of these two sets of agents that determines the resulting hedonic wage function. Ideally, one would want to understand the mechanisms by which objective risk measures are converted to subjective perceptions of these risks by *both* sides of the market.

Even ignoring the potential disconnect between subjective and objective risk measures, there are significant concerns regarding the objective risk measures underlying most of the hedonic wage models prior to the year 2000. Until relatively recently, the bulk of hedonic wage studies have relied on fatality data from two sources: the Bureau of Labor Statistic's Survey of Worker Conditions (BLS-SWC) and the National Institute of Occupational Safety and Health's National Traumatic Occupation Fatality Survey (NIOSH-NTOFS). In both instances, the available sources provide only fatality measures, requiring independent information on total employment in an industry and/or occupation in order to convert the fatality rates to risk measures. In the case of the BLS-SWC, the data were available only by two-digit and three-digit Standard Industrial Classification (SIC) code. The problem here is that risks can vary significantly within an industry depending on the individual's specific occupation (e.g., the steel mill worker versus clerical workers in the steel industry). The NIOSH data, on the other hand, were aggregated along different lines, providing fatality measures at the one-digit occupation or industry levels and by state. This form of aggregation, of course, creates its own problems. For both data sources, concerns have been raised regarding the accuracy of the fatality counts themselves (see, for example, Viscusi and Aldy 2003 and Dorman and Hagstrom 1998).

Black, Galdo, and Liu (2003) provided an extensive analysis of these two data sources and the implications for their use in valuing risk reductions. They argued that the data suffer from significant measurement error and that studies that fail to account for this measurement are likely to significantly understate the value of risk reduction. They went on to conclude that their findings lead them "to have severe doubts about the usefulness of existing estimates to guide public policy. These estimates are so highly sensitive to the risk measure used and the specification of the wage equation that the selection of any particular value of the price of risk seems arbitrary" (2003, 3). In another study, Black and Kniesner (2003) found that comparable risk measures from the two data sources "are not highly correlated, with a maximum correlation being 0.53," again drawing into question hedonic wage studies of risk that are based upon them.

Fortunately, several databases have emerged in recent years that seek to address the shortcomings in the BLS-SWC and NIOSH data. Of particular importance is the Census of Fatal Occupational Injuries (CFOI), which the BLS began collecting in 1992. One advantage of the CFOI is that it is available at the two- or three-digit SIC industry and occupation level. Moreover, fatalities reported in the CFOI are based on at least two independent sources, avoiding some of the reporting errors thought to plague both the BLS-SWC and NIOSH databases (see, for example, Mellow and Sider 1983 and Bound, Brown, and Mathiowetz 2001). A number of hedonic wage studies have employed this relatively new source of risk measures, including Kniesner et al. (2012), Kniesner, Viscusi, and Ziliak (2010), and Viscusi (2003, 2004). In a novel study, Lee and Taylor (2011) took advantage of plant-level injury and fatality data gathered by the Occupational Safety and Health Administration (OSHA) to construct plant-level risk measures, linking these to plant-level wages and worksite characteristics and avoiding some of the concerns with employing national measures of risk. Research based upon these alternative data sources is relatively new, but seems promising in terms of ameliorating concerns regarding measurement errors in the earlier BLS and NIOSH data.

Finally, even with these newer databases, an ongoing concern will be the inevitable imprecision with which fatality risk can be measured given the infrequent occurrence of on-the-job deaths in many industries and occupations. This problem is likely to only get worse over time with improvements in workplace safety and as researchers seek to provide a better (i.e., finely-tuned) match between the individual and the available fatality risk data. In order to alleviate this issue, it is not uncommon for researchers to average fatality risks over several years.

Omitted Variables Bias

The hedonic wage function represents the equilibrium outcome from the interaction between workers and firms. As such, it depends on the characteristics of both workers (influencing their productivity and preferences) and jobs (influencing worker utility and firm costs). Failure to control for characteristics in either category can lead to significant omitted variables bias. The classic example in the hedonic wage literature is the importance of controlling for both the fatality *and* injury risks associated with a job. Industries with high fatality rates are also likely to have high injury rates. Since workers presumably value reductions in both risks, and the two risks are positively correlated, including only fatality risk in the estimated hedonic wage function will tend to overstate its value to workers.

Equally important is controlling for worker characteristics. For example, it is not uncommon to find that the highest paying jobs also have some of the lowest levels of risk. This (unconditionally) negative correlation between risk and wage rates should not be construed as an indication that workers are willing to pay (in the form of lower wages) for increased risk. Rather, the high salary positions are usually associated with higher levels of education and/or experience. Firms are willing to pay more (both in wages and in the costs associated with providing lower levels of risk) in order to attract such workers.

A related concern is that of *endogenous risk*, with workers sorting themselves into jobs both within and across sectors based on unobserved skills in managing risk (Shogren and Stamland 2002). Individuals with "cool hands" may choose high-risk jobs, not because they place little value in risk reduction, but because their perceived (and perhaps actual) risk is lower than the objective risk measures in the industry. Failure to control for this sorting process will tend to understate the value of risk reductions.

Finally, an important development in recent years is the availability and use of panel data techniques to control for possibly omitted variables, including worker and industry/occupation characteristics. The use of individual fixed effects, for example, will control for any unobserved worker characteristics that are constant over time, mitigating the problem of omitted variables bias. Kniesner et al. (2012), for example, found that controlling for unobserved individual characteristics reduced their estimated wage–risk premium by up to 60 percent. Other examples include Hinterman, Alberini, and Markandya (2010), Kniesner, Viscusi, and Ziliak (2010), and Scotton and Taylor (2011). A potential downside of using panel data in hedonic wage analysis is that it implicitly assumes that the market equilibrium does not shift over the course of the panel, or that it shifts in ways that can be readily accounted for (using, for example, time-fixed effects). Changes in workplace regulations and the costs to firms of reducing risk, as well as shifts in the overall market, will endanger this assumption.

Inter-Industry and Inter-Personal Wage Differentials

In their recent review of the hedonic wage literature, Bockstael and McConnell (2007) highlighted the importance of controlling for inter-industry wage differentials. They noted that there is substantial evidence that "wage premia exist for some industries, irrespective of type of job or job attributes" (2007, 201). If labor markets are segmented across different industries, and wages are not allowed to equilibrate, then potential problems will arise with inferring risk values from differences in wages and risks across industries. This will be particularly the case if the factors impeding the equilibration (say differences in market power) also influence the levels of risk in the industry. This suggests, at a minimum, including industry dummy variables in the hedonic wage model, though it may also call into question treating such disparate industries as part of a single labor market.

A related issue can arise if there are persistent wage differentials due to discrimination. If an ethnic or gender class is effectively barred from a segment of the job market, their choices can no longer fully reflect their willingness to pay for risk reduction. Their choice of a low paying job with high levels of risk may simply reflect the more narrowly defined segment of the labor market that they have access to. Viscusi (2003) examined differences in labor market by race and the implications for these differences in terms of measuring wage–risk tradeoffs. This highlights, again, the importance of controlling for worker characteristics in estimating a hedonic price function.

Functional Form

Relatively little consideration has been given to the choice of functional form in hedonic wage literature. The vast majority of applications model log-wages, with risk typically entering linearly into the hedonic wage function. However, there have been a number of studies allowing for more flexible functional forms. Just over 20 percent of the cases included in Mrozek and Taylor's (2002) meta-analysis of the VSL literature allow risk to enter quadratically into the hedonic wage function, a term that is often found to be statistically significant. Moore and Viscusi (1988) and Shanmugam (1997) considered Box–Cox representations of the hedonic wage function that nest the semi-log and linear representations. In both cases, both the linear and semi-log representations are rejected as restrictions on the more general functional form. Given these results, and the related findings in the hedonic property literature by Kuminoff, Parmeter, and Pope (2010), it would seem prudent for researchers to consider more flexible functional forms moving forward.

Identifying Marginal Willingness-to-Pay Functions

We know that the partial derivative of the hedonic wage function with respect to each characteristic is its marginal implicit price, and that in equilibrium we can take the marginal implicit price to be a point estimate of each worker's marginal willingness to pay for that characteristic. However, as explained in Chapter 10, a second stage of analysis is required in order to identify the marginal willingness-to-pay function for each characteristic. As in the property value literature, there have been very few efforts to identify the willingness-to-pay function for job characteristics.

Two studies have dealt with the identification problem in somewhat different ways. In both cases, the authors imposed structure on the problem by making explicit assumptions about the form of the underlying utility functions. In one study (Biddle and Zarkin 1988), identification of the marginal willingness-to-pay function for reducing risks of an accidental injury on the job was achieved by an instrumental variables approach. The second study (Viscusi and Moore 1989) is noteworthy for two reasons. The first is that the authors identified the marginal willingness-to-pay function for reduced risks of death by making use of interregional variation in the hedonic wage function. To identify the parameters of the marginal bid function, the marginal price of job risk was estimated for different regions of the United States, thus assuring variation in marginal price that is independent of the variables entering the marginal bid function. The second reason is that the authors explicitly took into account the age of each worker and the number of expected life-years at risk.

Viscusi and Moore specified a lifetime utility maximization problem, which was a simplification of equation (7.17) in Chapter 7 of this book, in that the exogenous probability of death was constant across all time periods. This assumption made it possible to derive a simplified expression for the first-order conditions and the marginal willingness to pay for any given explicit utility function. The model also made it possible to derive estimates of the implied discount rate on life-years. Viscusi and Moore used data from the 1982 wave of the University of Michigan Panel Study of Income Dynamics, supplemented by National Traumatic Occupational Fatality data published by NIOSH. Depending upon the specification of the utility function in the estimation technique, the implied discount rate ranged from 11 to 17 percent for the sample as a whole (see also Moore and Viscusi 1990).

Conclusions

There is an extensive empirical literature suggesting the existence of a positive wage–risk premium (Kniesner and Leeth 1991; Liu, Hammitt, and Liu 1997; Shanmugam 2000), though the size of the premium varies substantially depending upon the data set being used, the time period analyzed, and modeling assumptions employed (Mrozek and Taylor 2002; Viscusi and Aldy 2003). It appears that (a) workers perceive differences in risks across jobs; (b) these perceptions are correlated with objective measures of on-the-job risks; and (c) workers prefer jobs with lower risks, other things being equal, and are willing to pay for safety in the form of reduced wages. Yet, serious concerns have been raised about the accuracy of the available empirical estimates, particularly those prior to the year 2000. While drawing on the best data available at the time (such as the BLS-SCW and NIOSH data), these studies are plagued by measurement and omitted variables bias stemming from the disconnect between both firms' and workers' perceptions of risks and available objective measures, the accuracy of the available risk data that are aggregated over occupational groupings and/or industry categories, and the ability to control for other individual and job-related determinants of wage differentials. Recent advances in econometric techniques, as well as the advent of new data sources (such as the BLS-CFOI), provide promise in terms of mitigating these concerns, though more work clearly remains.

It has also become clear that there is no one single value for risk reduction; rather, individuals' values depend upon such things as age and income, and even the type of fatality risk the person is exposed to. Studies that have examined the role of age in the wage–risk premium include Aldy and Viscusi (2007, 2008) and Evans and Smith (2006, 2008). Scotton and Taylor (2011) provided one of the few studies to differentiate the value of risk reduction according to type of risk, finding that the risk premium associated with homicide risk is substantially higher than that associated with traditional workplace risk. Understanding these individual determinants of behavior toward risk and the preferences and values that lie behind them is critical to their continued use in the policy arena. Indeed, this issue represents a potentially fundamental concern with the use of the hedonic wage model to value environmental risk. As Scotton and Taylor noted, "the application of VSL estimates from labor market studies, which our results clearly indicate are driven by traditional sources of workplace risks (e.g., electrocutions, falls, traffic accidents), to policy contexts involving reducing latent cancer risks, or premature mortality from acute asthmatic events, for example, is very likely inappropriate" (2010, 394). Similarly, if the value of risk reduction is age dependent, then estimated values for risk reduction based solely on hedonic wage models will be suspect, relying as they do on attitudes of working-age individuals and not on the preferences of young children (or their parents) and the elderly that typically face the highest risks from environmental pollutants.

Interurban Wage Differences and the Value of Amenities

All of the studies discussed in the previous section share the common characteristic of assuming that the location of the job is unimportant; that is, they do not control for differences in urban amenities and the potential effects of amenities on wage levels across cities. However, in a series of papers Smith and Gilbert have included location-specific characteristics such as air pollution levels in their hedonic wage equations (see Smith 1983, and Smith and Gilbert 1984, 1985). More recently, Rehdanz and Maddison (2004) examined the tradeoff between wages and climate amenities. The modeling of determinants of interurban wage differentials is discussed in this section.

Those cities that are more desirable places to live and work in will attract workers from less desirable cities and regions. The in-migration of labor will exert downward pressure on wage rates in the desirable city. An equilibrium occurs when wages have fallen to the point where the marginal worker is indifferent between moving to this city and staying in his or her next-best alternative location. The difference in wages between this city and the next-best alternative is a compensating wage differential.

The possibility that such compensating differentials could be used as estimates of the monetary value of amenity differences spawned a series of empirical studies during the 1970s (Nordhaus and Tobin 1972; Hoch and Drake 1974; Hoch 1977; Meyer and Leone 1977). In these studies, measures of average wage rates, earnings, or income were regressed on variables reflecting such things as climate (temperature, humidity, frequency of rain, for example), environmental quality (say, measures of air pollution, water pollution, and access to recreational resources such as beaches), cultural amenities (for example, number of museums, newspapers, and universities), the disamenities of urban life (for example, crime rates, population density), and city size itself. Some of these studies controlled for differences in the occupational structure of urban labor forces by estimating separate equations for individual occupations. Others used more aggregated measures such as average earnings. In none of these early studies was the estimating equation derived from a formal model of individual choice, interurban migration, or supplies and demands of labor in a system of interconnected urban labor markets.

Rosen (1979) was apparently the first to attempt to provide a formal model for deriving the structural equations relating wages to urban amenities and disamenities. Such a formal model is necessary to provide a welfare-theoretic interpretation of regressions that explain wage differences across cities. Rosen pointed out that there are really two hedonic markets in which individuals are making choices—one for labor, and one for land or housing. A decision to work in one city is also a decision to purchase housing services in that city. As individuals are drawn toward the more desirable cities and push wages down in those cities, they are also pushing out the demand for land and housing and increasing their prices. Not only are there compensating wage differentials, then, but there are also compensating land rent and housing price differentials across cities. The labor market model must also provide a coherent explanation for why firms in some cities are able to pay higher wages and still compete in markets for goods traded among cities. A formal model of both sets of markets is required in order to draw inferences about amenity values and willingness to pay from data on wage differentials and housing prices.

A number of such formal models have now been presented in the literature. See, for example, Rosen (1979), Cropper and Arriaga-Salinas (1980), Cropper (1981), Henderson (1982), Hoehn, Berger, and Blomquist (1987), Blomquist, Berger, and Hoehn (1988), and Roback (1982, 1988). Bartik and Smith (1987) reviewed most of these models, discussed similarities and differences in their theoretical structures, and described some of the major empirical results (also see Palmquist 1991). These models share the common feature of explicitly dealing with the interaction between markets for labor and for land across urban areas. However, they differ in the way in which they model certain specific features of the determination of the interurban equilibrium. For example, some models treat the size of each city as fixed, while others allow the city boundary to expand to accommodate higher populations. Some models treat the costs of firms as exogenous, while other models allow costs to be affected by environmental amenities and the population of the city itself. Also, some models treat amenity levels as uniform within each city, while others allow for variation in amenity levels within, as well as among, urban areas.

Apparently models that allow all of the relevant variables (such as city size, firm location, and cost) to be determined endogenously are analytically intractable. As a consequence, the results of these models depend upon the specific features of the model. None of these models fully captures all of the complexity of the general interurban equilibrium. The models produce what Bartik and Smith have termed "partial descriptions" of the more complex reality (1987, 1232).

This section does not try to explain all of the models and their different results. Instead, a simple model of individual choice is first introduced to show how the interaction between wages and rent or housing price affects our ability to interpret regression coefficients as welfare measures. One version is then sketched of a simple model of the equilibrium of the land and labor markets that permits the estimation of a marginal amenity value. The discussion ends by examining in a more qualitative manner the implications of constructing richer models with more endogenous variables and more interactions.

The Welfare Measure with Two Hedonic Markets

Most of the early studies of interurban wage differences were based either implicitly or explicitly on the assumption that wage differences measure the values of amenity differences. For example, Nordhaus and Tobin (1972) and Meyer and Leone (1977) used their results to make adjustments to the national income accounts for nonpriced positive and negative welfare effects of urbanization. Clark and Kahn (1989) interpreted their coefficient on an interurban wage equation as a marginal implicit price and used it in a second-stage estimation of a marginal bid function for an amenity. However, a very simple model of individual choice of a city to live and work in can be used to show that the assumption that wage differences measure amenity values is not valid in general. This is because the wage differential is also affected by an interaction between the markets for labor and land.

To see this, consider the simplest case of an individual who derives utility from the consumption of a numeraire good, z, the quantity of housing consumed, h, and the level of an urban-specific amenity, q. By selecting a city to live in, the individual determines her annual wage, p_w, and the level of the urban amenity. Assume that q_j is the same at all locations within city j. Let z be normalized with a price of one that is constant across cities, since z is a nationally traded good. The individual selects q_j along with z and h so as to maximize

$$u = u\left(z, h, q_j\right) + \lambda\left[p_w\left(q_j\right) - p_h\left(q_j\right) \cdot h - z\right] \ . \tag{11.6}$$

Note that both wages and the price of housing vary across cities according to the level of the amenity in each city. If one city is relatively more desirable as a place to live and work, other things being equal, workers will move to that city;

the increase in the supply of labor will push down wages in that city; and, since people must live in the same city in which they work, the increased demand for housing will bid up its price. Of course, a complete model must aggregate across individuals as well as specify the supply sides of these markets in order to solve for the wage and housing price equations (more on this below).

The first-order conditions for the individual choice problem are:

$$\frac{\partial u}{\partial z} = \lambda \tag{11.7}$$

$$\frac{\partial u}{\partial h} = \lambda \cdot p_h \tag{11.8}$$

$$\frac{\partial u}{\partial q} = \lambda \left(h \cdot \frac{\partial p_h}{\partial q} - \frac{\partial p_w}{\partial q} \right) . \tag{11.9}$$

Using the first-order conditions, the individual's optimum choice of z and q can be found. Graphically, the optimum combination is at the tangency between the individual's indifference curve for z and q and the budget line giving the terms at which z and q can be exchanged, both holding the quantity of housing h constant (see Figure 11.1). The marginal rate of substitution (MRS) between z and q is

$$MRS_{zq} = \frac{\partial u / \partial q}{\partial u / \partial z} = - \left. \frac{dz}{dq} \right|_{u=u^0} . \tag{11.10}$$

The slope of the budget line is found by taking the total differential of the budget:

$$\frac{dz}{dq} = \frac{\partial p_w}{\partial q} - h \cdot \frac{\partial p_h}{\partial q} . \tag{11.11}$$

The marginal rate of substitution can also be interpreted as the marginal willingness to pay for q or w_q . In equilibrium, w_q can be inferred from knowledge of the slope of the budget line:

$$w_q = h \cdot \frac{\partial p_h}{\partial q} - \frac{\partial p_w}{\partial q} . \tag{11.12}$$

This marginal willingness to pay has two components. The first component is the willingness to spend more on housing in a higher quality city with a higher price for housing; and the second is the willingness to accept a lower wage in the higher quality city.

This analysis makes it clear that knowledge of the hedonic wage gradient across cities is not sufficient for inferring marginal amenity values. The hedonic wage gradient is also shown in Figure 11.1. The budget line is more steeply sloped than the wage gradient because it includes the negative term for the effect of q on the price of housing. This means that the marginal implicit price of q in the labor

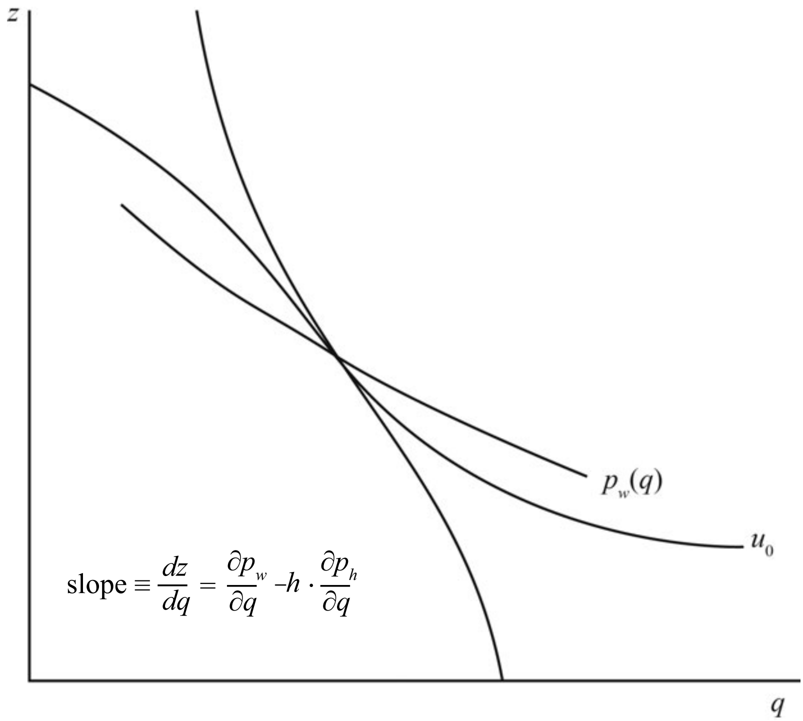

Figure 11.1 The individual's choice of an urban amenity

market is an underestimate of the individual's marginal valuation of q when q is in fact purchased simultaneously through two hedonic markets.

A General Model of Interurban Equilibrium

In order to estimate a welfare measure such as equation (11.12), it is necessary to develop a formal model of the interurban equilibrium so that the hedonic wage and housing price equations can be properly specified. The model presented here does not capture all of the complexities of the interurban location equilibrium problem, but it does show the general features of most of the models that have been presented in the literature. This model is patterned most closely after those of Hoehn, Berger, and Blomquist (1987) and Roback (1982), and incorporates the following features:

- It takes account of the variation in housing prices within cities caused by the spatial character of a city. Specifically, since land rents vary inversely with commuting costs, distance from the city center is an argument in the rent function.
- It treats the size of a city as endogenous. As cities that are more desirable draw workers to them and the price of housing is bid up, it becomes

profitable to convert some of the surrounding agricultural land to housing. This helps to dampen the upward pressure on housing prices.
- It treats the number of firms in each city as endogenous, and firms' costs are made to depend on the level of one or more amenities and disamenities and, perhaps, on city size, in order to reflect agglomeration effects.

This model assumes that all individuals have identical preferences and identical endowments of wealth, which in turn are assumed to be zero for simplicity. From the initial assumptions, it follows that an equilibrium must be characterized by equal levels of utility for all individuals. Each individual's preferences are represented by the utility function

$$u = u(z, a, q), \tag{11.13}$$

where z is the numeraire good traded in a "world" market at a price of one, a is the quantity of land (acres) occupied by the individual for housing, and q is the amenity that varies across cities but is uniform within each city. Let there be J cities in this economy. Each city offers a wage of p_{w_j} and an amenity level q_j, $j = 1, \ldots, J$.

Each city also has the standard circular form in which all jobs are located at the city center. Individuals choose a residential location at distance d, and incur commuting costs of $t \cdot d$ per period, where t is the unit per period transportation cost. In order to compensate for higher commuting costs, the price of land p_a will decline with increasing distance from the city center. The boundary of the city will be at distance d^* where in equilibrium $p_a\left(d^*\right)$ is equal to p_{a^*}, the rental price of undeveloped or agricultural land.

Each individual chooses a wage and amenity package by selecting a city in which to live and work, a location for his or her residence, a quantity of residential land, and a quantity of the numeraire. Formally, each individual's choice problem is to maximize

$$u = u(z, a, q) + \lambda\left[p_w\left(q_j\right) - t \cdot d - p_a(d) \cdot a - z\right]. \tag{11.14}$$

The solution to this problem gives the indirect utility function:

$$v = v\left[p_w\left(q_j\right) - t \cdot d, p_{a_j}(d), q_j\right]. \tag{11.15}$$

An equilibrium is achieved when wages and land prices have adjusted so that all individuals are indifferent as to the choice of a city and their location within the city chosen. From equation (11.15), this equilibrium must satisfy

$$v\left[p_w\left(q_j\right) - t \cdot d, p_{a_j}(d), q_j\right] = v^* \equiv v\left[p_w\left(q_j\right) - t \cdot d_j^*, p_{a^*}, q_j\right]. \tag{11.16}$$

The left-hand side of equation (11.16) can be solved for each city's land price schedule, $p_{a_j}\left(d, q_j, t, v^*\right)$. The right-hand side can be solved for the city's physical size, $d_j^*\left(t, p_{a^*}, q_j, v^*\right)$. The application of Roy's identity to equation (11.16) gives the individual's demand function for land:

$$a = -\frac{(\partial v/\partial p_a)}{(\partial v/\partial p_w)} = a\left(p_w - t\cdot d, p_{a_j}, q_j\right) , \tag{11.17}$$

and since the supply of land at any distance is given by $2\pi d$, the total number of people that any city can accommodate is

$$N = \int_0^{d^*} \frac{2\pi d}{a(\cdot)} dd \ . \tag{11.18}$$

The model must be closed by specifying the production sector of the economy, which also determines the demand for labor. Assume that the industry producing good z is competitive and subject to constant returns to scale. Then price, which has been normalized at one, equals average cost as well as marginal cost. If firms purchase materials and capital at fixed world prices, then some economic characteristic of each city must be an argument in the firms' cost functions; otherwise, all firms would be forced by competition to pay the same wage. Hoehn, Berger, and Blomquist (1987) assumed that q and total population are arguments in the cost function, the latter through an agglomeration effect. Alternatively, Roback (1982) assumed that firms use land in production, so that p_a is an argument in the cost function, along with q.

This system of equations can be solved to determine p_w and N for each city, and p_a and d for each individual within the city. These variables depend upon the exogenous q_j , t, and p_{a^*} , which are the variables to be included in hedonic wage and land price equations. Because of all of the interdependencies, the comparative statics of this type of model can be fairly complex. Hoehn, Berger, and Blomquist (1987) analyzed the comparative statics of their model with respect to changes in q. The results depend on how q and N affect the unit cost function.

To derive the marginal willingness to pay for q, return to equation (11.15) , take its total differential, and set it equal to zero, obtaining

$$dv = \frac{\partial v}{\partial p_w} dp_w + \frac{\partial v}{\partial p_a} dp_a + \frac{\partial v}{\partial q} dq = 0 \ , \tag{11.19}$$

and

$$w_q \equiv \frac{(\partial v/\partial q)}{(\partial v/\partial p_w)} = -\frac{(\partial v/\partial p_a)}{(\partial v/\partial p_w)} \cdot \frac{dp_a}{dq} - \frac{dp_w}{dq} \ . \tag{11.20}$$

Again, employing Roy's identity, this becomes

$$w_q = a \cdot \frac{dp_a}{dq} - \frac{dp_w}{dq} \ . \tag{11.21}$$

This expression is similar to the one derived above from the model of individual choice; see equation (11.12). Here, each individual's marginal willingness to pay for q consists of two components, the change in the expenditure on land associated with an increase in q, and the willingness to accept a lower wage rate for an improvement in q. The comparative static analysis of Hoehn, Berger, and Blomquist (1987) showed that it is possible for dp_a/dq to be negative or for dp_w/dq to be positive. It should be noticed that although the level of q is uniform within each city, individuals can have different marginal willingness to pay for q, depending upon the other elements in their consumption bundle, in particular, the quantity of land.

Other models of the interurban equilibrium have been developed to examine other forms of interaction. The additional features that can be captured in these models include the following:

- Intra-urban variation in amenity levels. Some amenities and disamenities that vary on average across cities also vary systematically within each city. For example, crime rates and air pollution tend to be higher in the center of each city than at the city boundary. This variation in amenity levels within cities will affect the spatial pattern of land rents and housing prices within each city (see Cropper 1981).
- The existence of both traded and nontraded goods. With zero transportation costs, traded goods sell at the same prices in all cities. However, the prices of nontraded goods can vary across cities, leading to differences in the cost of living. The explicit treatment of nontraded goods is necessary to determine how cost of living differences should be treated in the specification of the hedonic wage function. Cropper (1981) has shown that housing prices should not be included in the index of prices used as an argument in hedonic wage functions.
- Variation in the cost of supplying housing across cities. Although the simpler urban models are formulated in terms of land rent, housing prices are easier to observe for purposes of hedonic price estimation. Since the concept of interurban equilibrium is inherently long run, it is necessary to model the supply side of the housing market. The cost of producing housing will depend on, among other things, the price of land (which varies within and across cities) and the wage rate (which also varies across cities). For examples, see Roback (1982) and Hoehn, Berger, and Blomquist (1987).

Alternative Sorting Models

As noted in the previous chapter, a variety of sorting models have emerged in recent years as alternatives to traditional hedonic pricing models of quality-differentiated products. Not surprisingly, this line of research has made its way into models of migration. This section closes with a brief look at several examples of these efforts.

The first example is Bayer, Keohane, and Timmins (2009). The authors developed a discrete choice model of migration, with the utility from choosing a specific location depending upon local wage scales, housing conditions, and an environmental amenity of interest (in their application it is air quality). A central contribution of the paper is that the authors relaxed the assumption, underlying conventional hedonic techniques, that moving is costless. Instead, households are assumed to incur a psychological cost from moving outside the area in which they are born, a cost that is estimated as a part of the discrete choice model. The argument is that these costs create a barrier to migration and that ignoring them will yield biased estimates of the marginal willingness to pay (MWTP) for changes to the environmental amenity. Indeed, in their application they found that the MWTP for air quality improvements (measured in terms of ambient concentration of particulate matter) is three times larger when migration costs are accounted for than when they are ignored.

Timmins (2007) allowed for migration costs as well, but the focus of the paper is somewhat different. Whereas Bayer, Keohane, and Timmins (2009) were interested in estimating the MWTP for air quality improvements, Timmins' (2007) goal was to model the impact of nonmarginal environmental changes. To do so, he drew on the horizontal sorting models of Bayer, McMillan, and Reuben (2004) and Bayer and Timmins (2007). In his application, measuring the cost of climate change in Brazil, Timmins found that migration costs play an important part in determining both the overall impact of climate change and how the impacts are distributed in the population.

Finally, Kuminoff (2012) extended the vertical sorting model of Epple and Sieg (1999) to a dual-market setting, and developed a structural model of a household's interrelated choices of housing location and job. The framework recognized that the individual's ultimate set of choices may involve compromises between the two decisions—for example, accepting a lower-paying job (or one with few workplace amenities) in exchange for a location with higher local amenities (such as cleaner air or more recreational opportunities). In his application, modeling the MWTP for improved air quality in Northern California, Kuminoff found that accounting for the interaction between job and housing choices, rather than focusing on housing choice alone, led to an increase in the MWTP by up to 110 percent.

Both the vertical and horizontal sorting models that have emerged in recent years are promising tools for understanding the partial and general equilibrium impacts of changing environmental amenities and integrating the interrelated housing and labor market choices. The inclusion of moving costs allowed for in the horizontal sorting models seems particularly important. At the same time, these sorting models are relatively new and often quite complex, both in terms of their structure and in terms of the econometrics required for estimation. Additional research remains to be done in order to understand the implications of the assumptions required to close these models, particularly in terms of the specific functional forms used to represent utility and in the assumed distributions of both underlying preferences and the unobserved factors thought to influence them.

Summary

This chapter has examined two different types of application of hedonic price theory to labor markets. In one application, the focus is on the intrinsic characteristics of jobs. These applications seek to infer willingness to pay for changes in characteristics such as risk of accidental death or injury on the job. In these applications, it is usually assumed implicitly that the location of the job is unimportant. An exception is the work of Smith and Gilbert (1984, 1985) described above. In some wage–risk studies, regional dummy variables may be included to control for differences in wage levels across broad regions of the country, but the location of the job and the amenities or disamenities that go with living near the job are not explicitly considered. Recent advances in both the available data and in econometric techniques have helped to alleviate concerns regarding measurement errors and omitted variables bias in estimating the wage–risk premium. However, these advances do not eliminate these concerns, nor do they resolve the problem that these are best suited for working-age individuals, and not the elderly and infants most substantively impacted by environmental disamenities.

In the other application, the focus is on the choice of a city in which to live and work and, in some cases, where in the city to live. Jobs are typically treated as homogeneous, except for their location. These applications seek to infer willingness to pay for urban amenities. In some of these applications, job characteristics are controlled to some extent by examining interurban differences within broad occupational categories. A key feature of applications of this kind is the need to model explicitly the interaction between the two hedonic markets— those for labor and for land or housing. Introducing heterogeneous jobs into this interurban labor market does not require adding another hedonic market into the analysis. However, it does require a more detailed treatment of the determinants of wage differences within a multi-city, multi-characteristic market for differentiated labor. This is an important subject for further theoretical and empirical research.

References

Aldy, Joseph E., and W. Kip Viscusi. 2007. "Age Differences in the Value of a Statistical Life: Revealed Preference Evidence." *Review of Environmental Economics and Policy* 1(2): 241–260.

Aldy, Joseph E., and W. Kip Viscusi. 2008. "Adjusting the Value of a Statistical Life for Age and Cohort Effects." *Review of Economics and Statistics* 90(3): 573–581.

Bartik, Timothy J., and V. Kerry Smith. 1987. "Urban Amenities and Public Policy." In *Handbook of Regional and Urban Economics*, edited by Edwin S. Mills. Amsterdam: Elsevier.

Bayer, Patrick, and Christopher Timmins. 2007. "Estimating Equilibrium Models of Sorting Across Locations." *Economic Journa*l 117(518): 353–374.

Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins. 2009. "Migration and Hedonic Valuation: The Case of Air Quality." *Journal of Environmental Economics and Management* 58(1): 1–14.

Bayer, Patrick, Robert McMillan, and Kim Reuben. 2004. "An Equilibrium Model of Sorting in an Urban Housing Market." NBER Working Paper No. 10865. Available at: www.nber.org/papers/w10865.pdf.

Biddle, Jeff E., and Gary A. Zarkin. 1988. "Worker Preferences and Market Compensation for Job Risk." *Review of Economics and Statistics* 70(4): 660–667.

Black, Dan A., and Thomas J. Kniesner. 2003. "On the Measurement of Job Risk in Hedonic Wage Models." *Journal of Risk and Uncertainty* 27(3): 205–220.

Black, Dan A., Jose Galdo, and Liqun Liu. 2003. "How Robust are Hedonic Wage Estimates of the Price of Risk?" Center for Policy Research, Syracuse University, Syracuse, New York. Report presented to National Center for Environmental Economics, Office of Policy Economics and Innovation, USEPA. Available at: http://yosemite.epa.gov/ee/epa/eerm.nsf/vwAN/EE-0483-04.pdf/$file/EE-0483-04.pdf

Blomquist, Glenn C., Mark C. Berger, and John P. Hoehn. 1988. "New Estimates of Quality of Life in Urban Areas." *American Economic Review* 78(1): 89–107.

Bockstael, Nancy E., and Kenneth E. McConnell. 2007. *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models*. Dordrecht: Springer.

Bound, J., C. Brown, and N. Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics, vol. 5*, edited by James J. Heckman and Edward E. Leamer. Amsterdam: Elsevier.

Clark, David E., and James R. Kahn. 1989. "The Two-Stage Hedonic Wage Approach: A Methodology for the Evaluation of Environmental Amenities." *Journal of Environmental Economics and Management* 16(2): 106–120.

Cropper, Maureen L. 1981. "The Value of Urban Amenities." *Journal of Regional Science* 21(3): 359–373.

Cropper, Maureen L., and A.S. Arriaga-Salinas. 1980. "Inter-City Wage Differentials and the Value of Air Quality." *Journal of Urban Economics* 8(3): 236–254.

Cropper, Maureen L., James K. Hammitt, and Lisa A. Robinson. 2011. "Valuing Mortality Risk Reductions: Progress and Challenges." *Annual Review of Resource Economics* 3: 313–336.

Dorman, P., and P. Hagstrom. 1998. "Wage Compensation for Dangerous Work Revisited." *Industrial Labor Relations Review* 52(1): 116–135.

Epple, Denis, and Holger Sieg. 1999. "Estimating Equilibrium Models of Local Jurisdiction." *Journal of Political Economy* 107(4): 645–681.

Evans, Mary F., and V. Kerry Smith. 2006. "Do We Really Understand the Age-VSL Relationship?" *Resource and Energy Economics* 28(3): 242–261.

Evans, Mary F., and V. Kerry Smith. 2008. "Complementarity and the Measurement of Individual Risk Tradeoffs: Accounting for Quality and Quantity of Life Effects." *Environmental & Resource Economics* 41(3): 381–400.

Fischhoff, Baruch, Sarah Lichtenstein, Paul Slovic, Stephen L. Derby, and Ralph L. Keeney. 1981. *Acceptable Risk*. New York: Cambridge University Press.

Gerking, Shelby, Menno de Haan, and William Schulze. 1988. "The Marginal Value of Job Safety: A Contingent Valuation Study." *Journal of Risk and Uncertainty* 1(2): 185–199.

Hamermesh, Daniel S. 1985. "Expectations, Life Expectancy, and Economic Behavior." *Quarterly Journal of Economics* 100(2): 389–408.

Henderson, J. Vernon. 1982. "Evaluating Consumer Welfare Amenities and Interregional Welfare Differences." *Journal of Urban Economics*. 11(1): 32–59.

Hintermann, B., Anna Alberini, and Anil Markandya. 2010. "Estimating the Value of Safety with Labor Market Data: Are the Results Trustworthy?" *Applied Economics* 42(7–9): 1085–1100.

Hoch, Irving. 1977. "Variations in the Quality of Urban Life Among Cities and Regions." In *Public Economics and the Quality of Life*, edited by Lowdon Wingo and Alan Evans. Baltimore: Johns Hopkins University Press for Resources for the Future and the Centre for Environmental Studies.

Hoch, Irving, and Judith Drake. 1974. "Wages, Climate, and the Quality of Life." *Journal of Environmental Economics and Management* 1(3): 268–295.

Hoehn, John P., Mark C. Berger, and Glenn C. Blomquist. 1987. "A Hedonic Model of Interregional Wages, Rents, and Amenity Values." *Journal of Regional Science* 27(4): 605–620.

Kniesner, Thomas J., and John D. Leeth. 1991. "Compensating Wage Differentials for Fatal Injury Risk in Australia, Japan, and the United States." *Journal of Risk and Uncertainty* 4(1): 75–90.

Kniesner, Thomas J., W. Kip Viscusi, Christopher Woock, and James P. Ziliak. 2012. "The Value of a Statistical Life: Evidence from Panel Data." *Review of Economics and Statistics* 94(1): 74–87.

Kniesner, Thomas J., W. Kip Viscusi, and James P. Ziliak. 2010. "Policy Relevant Heterogeneity in the Value of Statistical Life: New Evidence from Panel Data Quantile Regressions." *Journal of Risk and Uncertainty* 40(1): 15–31.

Kuminoff, Nicholai. 2012. "Partial Identification of Preferences in a Dual-Market Sorting Equilibrium." Working paper, Arizona State University.

Kuminoff, Nicholai, Christopher Parmeter, and Jaren Pope. 2010. "Which Hedonic Models Can We Trust to Recover the Marginal Willingness to Pay for Environmental Amenities." *Journal of Environmental Economics and Management* 60(3): 145–160.

Lee, Jonathan M., and Laura O. Taylor. 2011. "A Quasi-Experimental Approach to Identifying Compensating Wage Differentials for Occupational Risks." Center for Environmental and Resource Economic Policy. Working paper, North Carolina State University. Available at: www4.ncsu.edu/~jmlee7/research/1job_market.pdf

Liu, Jin-Tan, and James K. Hammitt. 1999. "Perceived Risk and the Value of Workplace Safety in a Developing Country." *Journal of Risk Research* 2(3): 263–275.

Liu, Jin-Tan, James K. Hammitt, and Jin-Long Liu. 1997. "Estimated Hedonic Wage Function and the Value of Life in a Developing Country." *Economics Letters* 57(3): 353–358.

Lucas, Robert E.B. 1977. "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling." *American Economic Review* 67(4): 549–558.

Mellow, Wesley, and Hal Sider. 1983. "Accuracy of Response in Labor Market Surveys: Evidence and Implications." *Journal of Labor Economics* 1(4): 331–344.

Meyer, John R., and Robert A. Leone. 1977. "The Urban Disamenity Revisited." In *Public Economics and the Quality of Life*, edited by Lowdon Wingo and Alan Evans. Baltimore: Johns Hopkins University Press for Resources for the Future and the Centre for Environmental Studies.

Moore, Michael J., and W. Kip Viscusi. 1988. "Doubling the Estimated Value of Life: Results Using New Occupational Fatality Data." *Journal of Policy Analysis and Management* 7(3): 476–490.

Moore, Michael J., and W. Kip Viscusi. 1990. "Models for Estimating Discount Rates for Long-Term Health Risks Using Labor Market Data." *Journal of Risk and Uncertainty* 3(4): 381–401.

Mrozek, Janusz R., and Laura O. Taylor. 2002. "What Determines the Value of Life? A Meta-Analysis." *Journal of Policy Analysis Management.* 21(2): 253–270.

Nordhaus, William D., and James Tobin. 1972. "Is Growth Obsolete?" In *Economic Research: Retrospect and Prospect, Economic Growth,* Fiftieth Anniversary Colloquium, vol. 5. New York: National Bureau of Economic Research.

Palmquist, Raymond B. 1991. "Hedonic Methods." In *Measuring the Demand for Environmental Quality*, edited by John B. Braden and Charles D. Kolstad. Amsterdam: North-Holland.

Rehdanz, Katrin, and David Maddison. 2004. "The Amenity Value of Climate to German Households." *Oxford Economic Papers* 61(1): 150–167.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(6): 1257–1278.

Roback, Jennifer. 1988. "Wages, Rents, and Amenities: Differences Among Workers and Regions." *Economic Inquiry* 26(1): 23–41.

Rosen, Sherwin. 1979. "Wage-Based Indices of Urban Quality of Life." In *Current Issues in Urban Economics*, edited by Peter Mieszkowski and Mahlon Straszheim. Baltimore: Johns Hopkins University Press.

Scotton, Carol R., and Laura O. Taylor. 2011. "Valuing Risk Reductions: Incorporating Risk Heterogeneity into a Revealed Preference Framework." *Resource and Energy Economics* 33(2): 381–397.

Shanmugam, K.R. 1997. "The Value of Life: Estimates from Indian Labour Market." *Indian Economic Journal* 44(4): 105–114.

Shanmugam, K.R. 2000. "Valuation of Life and Injury Risks." *Environmental and Resource Economics* 16(4): 379–389.

Shogren, Jason F., and Tommy Stamland. 2002. "Skill and the Value of Life." *Journal of Political Economy* 110(5): 1168–1173.

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein. 1979. "Rating the Risks." *Environment* 21(3): 14–20, 36–39.

Smith, V. Kerry. 1983. "The Role of Site and Job Characteristics in Hedonic Wage Models." *Journal of Urban Economics* 13(4): 296–321.

Smith, V. Kerry, and Carol C.S. Gilbert. 1984. "The Implicit Valuation of Risks to Life: A Comparative Analysis." *Economics Letters* 16(3–4): 393–399.

Smith, V. Kerry, and Carol C.S. Gilbert. 1985. "The Valuation of Environmental Risks Using Hedonic Wage Models." In *Horizontal Equity, Uncertainty, and Economic Well-Being*, edited by Martin David and Timothy Smeeding. Chicago: University of Chicago Press.

Timmins, Christopher. 2007. "If You Cannot Take the Heat, Get of the Cerrado… Recovering the Equilibrium Amenity Cost of Nonmarginal Climate Change in Brazil." *Journal of Regional Science* 47(1): 1–25.

Viscusi, W. Kip. 1979. *Employment Hazards: An Investigation of Market Performance*. Cambridge: Harvard University Press.

Viscusi, W. Kip. 2003. "Racial Differences in Labor Market Values of a Statistical Life." *Journal of Risk and Uncertainty* 27(3): 239–256.

Viscusi, W. Kip. 2004. "The Value of Life: Estimates with Risks by Occupation and Industry." *Economic Inquiry* 42(1): 29–48.

Viscusi, W. Kip, and Joseph E. Aldy. 2003. "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World." *Journal of Risk and Uncertainty* 27(1): 5–76.

Viscusi, W. Kip, and Michael J. Moore. 1989. "Rates of Time Preference and Valuation of the Duration of Life." *Journal of Public Economics* 38(3): 297–317.

Viscusi, W. Kip, and Charles J. O'Connor. 1984. "Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?" *American Economic Review* 74(5): 942–956.

Chapter 12

Stated Preference Methods for Valuation

Chapters 9 through 11 described methods for inferring individuals' values for environmental amenities from their observed choices of related market goods. These approaches are broadly categorized as "revealed preference" (RP) techniques because, it is argued, a person's actions in the marketplace reveal information about his core preferences, including his preferences for public goods.[1] However, discussions in Chapter 4 also showed that there are many circumstances under which value measures cannot be derived from market transactions. This is the case, for example, when individuals are thought to place value in an environmental amenity that they do not directly use (that is, so-called "passive-use" values). Because such values are not tied to the use of a related market commodity, they leave no footprint in the marketplace from which their magnitude can be inferred. Even in the case of use values, revealed preference techniques may be hampered by a lack of sufficient independent variation in the amenity of interest from which to infer its impact on behavior and preferences. In the context of recreation demand models, for example, it may be difficult to discern the marginal value associated with a specific site amenity (e.g., a reduction in lake algae) when there is limited variation in the amenity across existing sites, or a high degree of correlation with other site attributes (e.g., fish catch rates).

This chapter examines a broad class of alternative techniques that can be used to assess the welfare impact of changing environmental conditions, either in lieu of the RP approaches described thus far, or as a complement to them. These methods have in common their source of data for analysis: individuals' responses to questions about hypothetical situations such as "Would you pay $\$X$ for …?", "What is the most that you would be willing to pay for …?", "What would you do if …?", or "Which of the following alternatives do you prefer …?" Since values are inferred from stated responses to such questions, these methods are now commonly referred to as stated preference (SP) methods.

Although not all authors use the same terminology, the term stated preference methods has come to refer to any survey-based study in which respondents are

1 The tight linkage between behavior and core preferences has come into question in recent years by both behavioral economists and psychologists. See, for example, Sugden (2004), Beshears et al. (2008), and Vatn (2004).

asked questions that are designed to reveal information about their preferences or values. The term encompasses three broad types of questions. The first type involves questions that elicit monetary values for a specified commodity or environmental change. These are usually referred to as *contingent valuation* (CV) methods, as the survey responses are assumed to be "contingent" upon the presented hypothetical scenario. A wide range of elicitation formats fall within the CV framework. A relatively straightforward CV question simply asks people what value they place on a specified change in an environmental amenity or the maximum amount they would be willing to pay to have it occur. Also referred to as an "open-ended" elicitation format, the responses to these questions, if truthful, are direct expressions of value, and would be interpreted as measures of compensating surplus (*CS*). Another major type of CV question asks for a yes or no answer to the question, "Would you be willing to pay $\$X$ …?" Each individual's response reveals only an upper bound (for a no) or a lower bound (for a yes) on the relevant welfare measure. Questions of this sort are termed *binary discrete choice* (or "closed-ended") questions. By comparing the responses to such questions across individuals facing different values of X, estimated willingness to pay or indirect utility functions can be obtained.

The second and third major types of SP methods do not reveal monetary measures directly. Rather, they require some form of analytical model to derive welfare measures from responses to questions. In the second approach to questioning, respondents are asked how they would change the level of some activity in response to a change in an environmental amenity. If the activity can be interpreted in the context of some behavioral model, such as an averting behavior model or a recreation demand model, the appropriate revealed preference valuation method can be used to obtain a measure of willingness to pay. These are known as contingent behavior (CB) questions.

In the third type of SP question, individuals are given a set of hypothetical alternatives, each depicting a different bundle of environmental attributes. Respondents are asked to choose the most preferred alternative, to rank the alternatives in order of preference, or to rate them on some scale. Responses to these questions can then be analyzed to determine, in effect, the marginal rates of substitution between any pair of attributes that differentiate the alternatives. If one of the characteristics has a monetary price, then it is possible to compute the respondent's willingness to pay for the attribute on the basis of the responses. Studies based on this form of question are usually referred to as choice experiments or sometimes attribute-based methods (Holmes and Adamowicz 2003) and have their roots in the marketing literature (Louviere, Hensher, and Swait 2000).

The use of the SP methods for environmental valuation has been controversial. This controversy became especially heated after the Exxon Valdez oil spill in March 1989 when it became known that a major component of the legal claims for damages was likely to be based on CV estimates of lost passive use values. In response to the Exxon litigation, National Oceanic and Atmospheric Administration (NOAA) convened a Blue Ribbon Panel, chaired by two Nobel Prize Economists, Kenneth Arrow and Robert Solow, to review CV in the context of assessing damages to natural resources in support of litigation and provide guidelines for best practice. The Panel's report (Arrow et al. 1993), while not quelling the debate over the validity of SP techniques, had the salutary effect of stimulating a substantial body of new research on both SP practice and on the credibility or validity of SP-based estimates of value. A good overview of the issues raised at the time is contained in the three essays published as a symposium in the *Journal of Economic Perspectives* (Portney 1994; Hanemann 1994; Diamond and Hausman 1994). To get a sense of the intensity of the controversy, see the collection of essays sponsored by Exxon Corporation—especially the transcripts of the discussions by audience members and authors at the public presentation of these essays (Hausman 1993). A reprisal of this debate was recently published in the *Journal of Economic Perspectives* (Kling, Phaneuf, and Zhao 2012; Carson 2012; Hausman 2012). Interestingly, despite nearly two decades of research, some of the positions held back in 1994 have hardly changed.

The remainder of this chapter is structured as follows. The next section describes in more detail the major types of SP question formats and, where appropriate, explains how responses can be analyzed to obtain measures of welfare for changes in the environmental amenity. A major question regarding all stated preference methods concerns the validity and reliability of the data—that is, whether the hypothetical nature of the questions asked inevitably leads to some kind of bias or results in so much "noise" that the data are not useful for drawing inferences. The second section of this chapter considers approaches to assessing the validity of measures of value obtained with stated preference methods, where validity refers to the degree of correspondence between the measure obtained and the theoretical concept of value. The chapter concludes with an assessment of the current "state-of-the-art" of the stated preference methods.

The treatment in this chapter of the issues surrounding stated preference methods is selective. Relatively little attention is given to a number of design and implementation problems, such as the form of contact with respondents (mail, telephone, personal interview), survey testing, and sampling design. While these issues are essential to successful stated preference research, they are treated extensively elsewhere in the literature. Mitchell and Carson's (1989) pioneering treatise is still a primary reference for CV studies, especially for design and implementation questions. Two more recent works that focus on best practice and empirical estimation for CV, and nonmarket valuation more generically, are Bateman and Willis (1999) and Champ, Boyle, and Brown (2003), respectively. Louviere, Henscher, and Swait (2000) provided a primer for choice experiments. Other important references are: Bjornstad and Kahn (1996) for a review of theoretical and empirical issues that includes assessments by both proponents and critics of stated preference methods; Haab and McConnell (2002) for a discussion of econometric issues in non-market valuation; Bateman et al. (2002) for a user's guide to stated preference methods; Carson and Hanemann (2005) for an overview of contingent valuation; Carson (2011) for a comprehensive overview of the history of contingent valuation and an extensive bibliography of papers in the area; and Kopp, Pommerehne and Schwarz (1997).

Stated Preference Approaches to Valuation

Contingent Valuation

Contingent Valuation (CV) is perhaps the best known of the stated preference techniques employed in valuing environmental amenities, with applications dating back to Davis (1963). At their heart, these techniques seek Hicksian measures of the welfare impact of hypothetical changes in environmental conditions, or at least bounds on such measures. The specifics of the survey design vary, but as Carson and Hanemann (2005, 825) noted, based on current best practices, the survey instruments typically contain the following elements: "(1) an introductory section identifying the sponsor and general topic, (2) a section asking questions concerning prior knowledge about the good and attitudes toward it, (3) the presentation of the CV scenario including what the project was designed to accomplish, how it would be implemented and paid for, and what will happen under the current status quo situation if the project were not implemented, (4) question(s) asking for information about the respondent's WTP/WTA for the good, (5) debriefing questions to help ascertain how well respondents understood the scenario, and (6) demographic questions." In what follows, variations in the elicitation format (that is, item 4) are described.

Open-Ended Questions

Data obtained from open-ended value questions, taken at face value, are the simplest to interpret. Each respondent is typically asked to state his or her maximum willingness to pay (WTP) for an environmental improvement (compensating surplus) or to avoid a loss (equivalent surplus).[2] There are several ways to elicit this number. Many of the earliest studies used an iterative technique, which has come to be called the bidding game. In the bidding game, individuals are first asked whether they would be willing to pay $*X*. If the individual answers yes, the question is repeated with a higher "price." The procedure is repeated until the individual answers no. The highest price with a yes response is interpreted as the maximum willingness to pay. If the original response is no, the iteration proceeds downward until a yes response is received. Bidding games, however, fell out of favor in the literature when researchers found that the starting point used in the bidding process influenced the individual's purported maximum WTP. See Whitehead (2002) for a recent example. The specific source of this so-called "starting point bias" is unknown, though possible explanations include respondent fatigue (with the individual saying "yes" or "no" to simply stop the line of questioning) or the respondent's interpreting the starting point as conveying information about either the cost of provision or a "reasonable" estimate of the good's value.

2 Minimum willingness to accept (WTA), as compensation for an environmental degradation or in lieu of an environmental improvement, is sometimes also asked for, but is less common in the literature for reasons discussed later in this chapter.

An alternative elicitation technique is simply to ask open-ended questions of the form: "How much would you be willing to pay?" One concern with this approach is that it confronts people with an unfamiliar problem. In most real market settings, individuals are faced with choices among sets of goods with listed prices. Seldom are they asked to offer a one-shot bid that may be either accepted or rejected by the seller.[3] People appear to have difficulty dealing with the open-ended form of direct question as evidenced by high rates of nonresponse to the valuation question and/or high proportions of implausibly high or low stated values.

A variation on the open-ended approach is to show respondents a card with a range of alternative payment values on it and ask them either to pick a number from the card or to state their own value if that is not to be found on the card. Some authors have also experimented with payment cards that indicate the amounts that typical respondents are paying in the form of taxes for such public programs as police protection, health care, and national defense. Rowe, Schulze, and Breffle (1996) conducted a study to see if changes in the ranges covered by payment cards affected the distribution of expressed maximum WTPs. They found no effect as long as the payment card did not truncate the upper range of values.

Regardless of the specifics of the elicitation process, the key advantage of the open-ended format is that it provides a specific welfare number (W_i) for each survey respondent. If these numbers accurately reflect the individual's preferences, subsequent analysis of the data is relatively straightforward. An estimate of the total value of the welfare change for the population from which the sample is drawn can be obtained by calculating the sample mean and multiplying by the total population. Alternatively, the responses can be regressed on income $\left(M_i\right)$ and other socioeconomic characteristics ($\boldsymbol{S}_i$) to obtain a bid function for the proposed policy scenario:

$$B_i = B\left(M_i, \boldsymbol{S}_i\right). \tag{12.1}$$

If the survey design includes variation in the size or composition of the environmental changes ($\Delta\boldsymbol{Q}_i$) across the sample, then the bid function

$$B_i = B\left(\Delta\boldsymbol{Q}_i, M_i, \boldsymbol{S}_i\right) \tag{12.2}$$

can be estimated and used to calculate values for alternative scenarios of environmental or resource change.

Two econometric issues typically arise in estimating either type of bid function. First, the parameter estimates can be sensitive to outliers. Many surveys obtain at least a few bids that are so large relative to the sample mean, or individual income levels as to be of questionable validity. Varieties of procedures have been suggested in the literature to deal with these observations, including trimming or censoring

3 This statement, while generally true in much of the industrialized world, is perhaps less of a concern in countries where haggling in local market places is more commonplace. See Whittington (2002) for a discussion of the use of contingent valuation methods in developing countries.

extreme values (e.g., Mitchell and Carson 1989, 226–227). As these procedures involve judgment calls regarding what constitutes an "extreme value," sensitivity analysis should be reported. Second, open-ended questions will often yield a large number of zero values. Censored regression models, such as Tobit, or double-hurdle models, are typically used to accommodate the mass of observations at zero. See Haab and McConnell (2002) for additional details.

The primary concern regarding the open-ended elicitation format is whether the reported welfare measure W_i accurately reflects the respondent's true preferences. There are two related issues here. First, if the survey respondent does not believe that their answers will have any impact on them, either because the survey will not ultimately influence policy or because they believe they will not have to pay for the policy change, then the survey itself is *inconsequential* and they have no incentive to respond truthfully.[4] Second, even if the respondent believes that the survey is *consequential*, Carson and Groves (2007) argued that the open-ended elicitation format is *incentive incompatible*; that is, agents have an incentive to exaggerate their reported WTP. The problem arises in part because it is not credible that actual payment for the program will be tailored to the individual's reported W_i. Consider the simple case in which the respondent believes her portion of the program's cost is fixed, say at C_i. In this setting, she has an incentive to report an arbitrarily high WTP if $W_i > C_i$ (thereby encouraging the policy's adoption) and report a zero WTP if $W_i < C_i$. The direction of the bias will depend upon individual beliefs regarding C_i. While conventional wisdom early on was that open-ended questions would lead to overstated WTP estimates, the reverse appears to be the case.

Another concern with open-ended questions is that the sample could contain invalid zero responses, so-called *protest zeros*. Protest zeros occur when respondents reject some aspect of the constructed market scenario by reporting a zero value even though they place a positive value on the amenity or resource being valued. Simply trimming a portion of the zeros is not a solution to the problem, since it is not known a priori what proportion of the zeros to trim. Rather, some means must be found to identify protest zeros for deletion before applying the procedure for deleting outliers. One approach commonly used in the literature is to follow up the valuation question with a question regarding the individual's motive for his response. For example, respondents could be asked:

Which statement best expresses your reason for your response?

- I can't afford to pay for the good.
- The good is not important to me.
- I don't think that I should have to pay for the good.
- The proposed program is unrealistic.

4 The notion of consequentiality is discussed in greater detail in Carson and Groves (2007). This issue is discussed again later in the chapter.

Responses of those choosing either the third or fourth statement would be classified as protest zeros and deleted from the sample, while responses of those choosing the first and second would be considered valid responses. Unfortunately, the deletion of protest zeros can also lead to item nonresponse bias if those who protest are systematically different in some respect from those who give proper responses. At a minimum, it would be desirable to examine whether the individuals in the protest group differ in observable ways from the general population.

Binary Discrete Choice Questions

Perhaps the most commonly used elicitation format in contingent valuation is the single-shot binary discrete choice question, also known as the dichotomous choice format. The questions are typically couched in the form of a referendum. After presenting the survey participant with the proposed environmental changes and the cost (also referred to as the "bid amount") that they would bear if the changes were implemented, the individual is asked if they would vote in favor of the referendum. If a respondent answers yes, that person has indicated a willingness to pay that is greater than or equal to the specified cost. If the response is no, then that sum of money can be taken as an upper bound on true willingness to pay. Respondents are assigned randomly to different subsamples, with each subsample being asked to respond to a different bid amount. It is then possible to test the hypothesis that the proportion of yes responses decreases with an increase in the price of the environmental good. These data can then be analyzed with a model of discrete choice to obtain estimates of indirect utility functions or bid functions.

The single-shot binary choice format has at least three advantages relative to open-ended formats. First, it places people in a relatively familiar social context. Many private market transactions involve goods offered on a take-it-or-leave-it basis in which the individual decides whether or not to purchase the good at the offered price. Furthermore, if the payment vehicle is a tax, the discrete choice question simulates a true referendum of the sort found everywhere from small New England town meetings to statewide votes on highway bond issues. The second advantage is that, since only a yes or no answer is required, the discrete choice question format poses a relatively simple decision problem for individuals. This may result in lower levels of item nonresponse and fewer refusals to participate in the survey. Third, in at least some circumstances, it is incentive-compatible; that is, respondents' best strategy is to be truthful in answering the question. This point is returned to below.

The primary disadvantage of the binary choice format is that it yields relatively little information from each survey respondent. Specifically, one learns only whether the individual's WTP for the proposed program lies above or below the bid value they are presented with. Consequently, relatively large sample sizes are required in order to accurately characterize central tendencies and the distributional characteristics of WTP in the population. This is all the more the case if one is interested in recovering a bid function analogous to (12.2) above.

In order to convert data on yes or no responses to a discrete choice question into a monetary measure, it is necessary to employ some explicit utility-theoretic model of choice. The discrete choice model introduced in Chapter 4 is well suited to the task. As Cameron (1988) and Cameron and James (1987) have pointed out, the variation in prices across the sample makes it possible to explain individuals' choices in terms of a willingness-to-pay function rather than in terms of differences in indirect utility. McConnell (1990) has compared the Cameron model with the model based on utility functions outlined by Hanemann (1984) and by Sellar, Chavas, and Stoll (1986). He showed that the two models can be derived from the same underlying utility-theoretic framework. In a deterministic formulation, the two models yield the same predictions about behavior and choice. Since the stochastic forms of the models introduce random components in different ways, the two models are not in general equivalent, although there are special cases in which the two models are dual to each other. Moreover, as Cameron (1988) showed, the willingness-to-pay function model permits the straightforward calculation of marginal values for all of the arguments in the willingness-to-pay function, while this is not possible with the utility function model.

In what follows, a model of the individual's response to a single dichotomous choice question is first presented using the utility difference framework of Hanemann (1984), showing how welfare measures can be derived. This is followed by the parallel development based upon the willingness-to-pay function approach.

Consider an individual who must decide whether to answer yes or no to the following question: "Would you vote for a program to permanently increase environmental quality from q^0 to q^1 if the total cost to you was \$$T$?" Let the indirect utility function be $u(q,M,\boldsymbol{S})$, where $\boldsymbol{S}$ is a vector of individual characteristics and the vector of market prices, $\boldsymbol{P}$, is omitted since these prices are assumed to be constant. The individual responds yes if

$$u\left(q^1,M-T,\boldsymbol{S}\right)>u\left(q^0,M,\boldsymbol{S}\right) , \tag{12.3}$$

and no otherwise.

Let $v(\cdot)$ denote the analyst's representation of consumer preferences, sometimes also referred to as the "observable" component of utility. The probability of a yes response is then given by

$$\Pr(Y)=\Pr\left[v\left(q^1,M-T,\boldsymbol{S}\right)+\varepsilon^1>v\left(q^0,M,\boldsymbol{S}\right)+\varepsilon^0\right] , \tag{12.4}$$

where the $\varepsilon^j\equiv u\left(q^j,M,\boldsymbol{S}\right)-v\left(q^j,M,\boldsymbol{S}\right)\left(j=0,1\right)$ are the random, unobserved components of utility.[5] As discussed in Chapter 4, if the random terms are

5 The error terms in this random utility model capture a myriad of possible errors on the part of the analyst in characterizing preferences, including characteristics of the individual or the choice alternatives that are unobserved by the analyst, errors in the functional form used by the analyst to represent utility (i.e., v), or measurement errors on the part of the analyst. It is important to keep in mind, however, that these error terms are not capturing errors or uncertainty on the part of the respondent. We discuss this point again later in the chapter.

independently and identically distributed with a Type I Extreme Value distribution, then this probability can be expressed as

$$\Pr(Y) = \frac{\exp(\Delta v)}{1 + \exp(\Delta v)}, \tag{12.5}$$

where $\Delta v \equiv v^1 - v^0$. Also, reversing the sign on the probability difference gives the expression for the probability of rejecting the offer:

$$\Pr(N) = \frac{1}{1 + \exp(\Delta v)}. \tag{12.6}$$

The willingness to pay for q^1 (CS) is defined implicitly by

$$u(q^1, M - CS, \boldsymbol{S}) = u(q^0, M, \boldsymbol{S}), \tag{12.7}$$

or equivalently as

$$v(q^1, M - CS, \boldsymbol{S}) + \varepsilon^1 - \varepsilon^0 = v(q^0, M, \boldsymbol{S}). \tag{12.8}$$

In terms of the observable utility function, CS is thus a random variable, because of the term $\varepsilon^1 - \varepsilon^0$. The probability of accepting the offer is also, then, the probability that $CS \geq T$, and the probability of rejecting the offer is the probability that $CS < T$. This is a cumulative density function (c.d.f.), denoted here as $F(T)$, and shown in Figure 12.1, which plots the probability of a "no response," $\Pr(N)$, as a function of T. As Hanemann (1984) has pointed out, the expected value of the random variable CS can be found from the c.d.f. as follows:

$$E[CS] = \int_0^\infty [1 - F(T)]\, dT. \tag{12.9}$$

Graphically, in Figure 12.1, the expected value of CS is the shaded area above the c.d.f. and below $\Pr(N) = 1$. For example, for any given T, an individual with a lower willingness to pay for the change in q would have a higher probability of rejecting the offer, and the shaded area would be smaller. Alternatively, an offer of $q^2 > q^1$ at any T would decrease the probability of an individual rejecting the offer. So $F(T)$ would be shifted down and $E[CS]$ would be larger.

Specifying a functional form for the observable component of utility makes it possible to estimate the parameters of the utility difference in equations (12.5) and (12.6). For example (following Hanemann 1984), if $v = a + b{\cdot}\ln M + c{\cdot}\ln q$ (ignoring the $\boldsymbol{S}$ vector for simplicity), then

$$\Delta v = (a^1 - a^0) + \left[b \cdot \ln\left(1 - \frac{T}{M}\right)\right] + \left[c \cdot \ln\left(\frac{q^1}{q^0}\right)\right]. \tag{12.10}$$

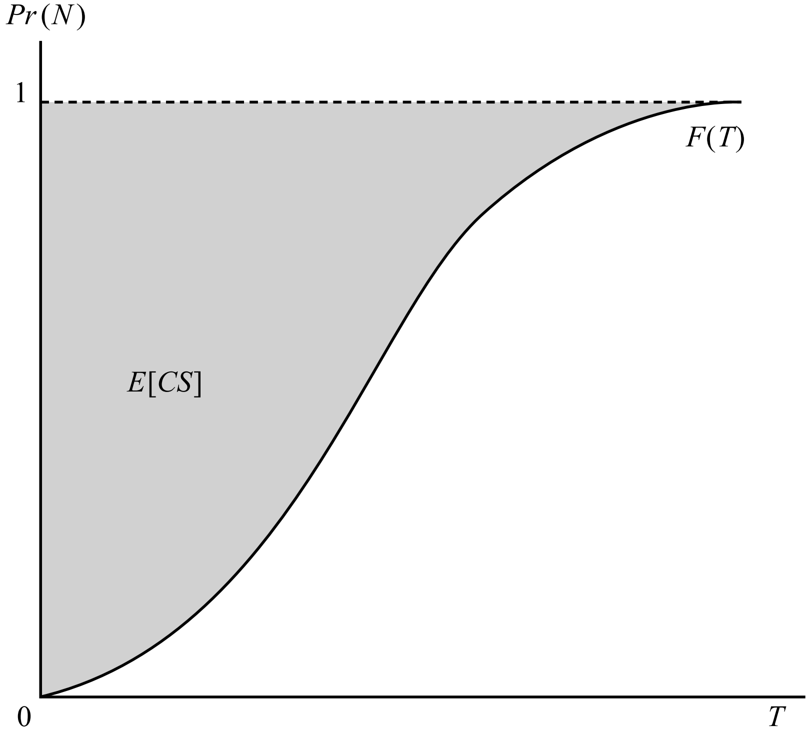

Figure 12.1 The cumulative density function for rejecting the offer of q^1 at T and the expected value of *CS*

Then with the parameters from equation (12.10), equations (12.5) or (12.6) and (12.9) can be used to calculate $E[CS]$ (for example, Hanemann 1984 and Seller, Chavas, and Stoll 1986).

As equation (12.19) shows, in principle $F(T)$ is integrated over the range to infinity. In order to avoid the implausibly high estimates of $E[CS]$ that can sometimes result, some researchers have truncated the integration of $F(T)$ at some "reasonable" finite value for T, such as some fraction of the individual's income.[6] Estimates of $E[CS]$ can be highly sensitive to the value of T_{max} chosen for truncation, at least in some cases (Bishop and Heberlein 1979; Hanemann 1984; Mitchell and Carson 1989, 103, 196–197). However, T should, at a minimum, be bounded above by income, recognizing the individual's budget constraint.

There is an alternative approach to dealing with the sensitivity of estimates of *CS* to outliers in the distribution. As Hanemann (1984) has suggested, the median willingness to pay for the sample can be found by setting $\Pr(Y)$ in equation (12.5) equal to 0.5 and solving for T. This gives the value for T that makes

6 Indeed, Δv in equation (12.10) becomes undefined for $T > M$.

the representative individual indifferent as to accepting or rejecting the offer. Generalizing to the sample, one would expect half the respondents to accept (and half to reject) the offer, other things (including income and ***S***) being held constant. This procedure is less sensitive to extreme values in the data.

The alternative model for analyzing discrete choice data is based on the bid or willingness-to-pay function. This function can be derived from the expenditure function as follows:

$$B\left(q^0, q^1, u^0, \boldsymbol{S}\right) = e\left(q^1, u^0, \boldsymbol{S}\right) - e\left(q^0, u^0, \boldsymbol{S}\right). \tag{12.11}$$

Cameron (1988) called this a valuation function, and McConnell (1990) called it a variation function, because it can be defined for either compensating or equivalent welfare measures. The formulation here gives the compensating surplus measure of value for an improvement. Other versions can easily be specified for losses and for *ES* measures. See Carson (1991, especially 143–144) for further discussion of this model.

The individual will respond yes if

$$B\left(q^0, q^1, u^0, \boldsymbol{S}\right) \geq T, \tag{12.12}$$

and no otherwise. The probability of accepting the offer of q^1 at T can be expressed in terms of the function

$$\Pr\left(Y\right) = \Pr\left[B^*\left(q^0, q^1, u^0, \boldsymbol{S}\right) + \eta \geq T\right], \tag{12.13}$$

where B^* is the observable component of the bid function and η is the unobserved random component of willingness to pay. In other words,

$$B\left(q^0, q^1, u^0, \boldsymbol{S}\right) = B^*\left(q^0, q^1, u^0, \boldsymbol{S}\right) + \eta. \tag{12.14}$$

The next step is to make some assumptions about the distribution of the random component of the bid. Cameron and James (1987) assumed that η is normally distributed with zero mean and standard deviation σ. The result is a form of probit model in which

$$\Pr\left(Y\right) = \Phi\left[\frac{B^*\left(q^0, q^1, u^0, \boldsymbol{S}\right) - T}{\sigma}\right], \tag{12.15}$$

where $\Phi(\cdot)$ is the standard normal cumulative density. It is typical in the literature to assume that $B^* = \boldsymbol{X}\beta$, where $\boldsymbol{X}$ is a row vector including characteristics of the individual (***S***), the choice alternative (e.g., $\Delta q = q^1 - q^0$) and their interactions, while β is a conformable vector of parameters. The bid function in (12.14) can then be rewritten as

$$B = \boldsymbol{X}\beta + \sigma\eta_s, \tag{12.16}$$

where $\eta_s \sim N\left(0,1\right)$ denotes a standard normal random variable. In this case, (12.15) reduces to

$$\Pr(Y)=\Phi\left[\boldsymbol{X}\left(\frac{\beta}{\sigma}\right)-\frac{1}{\sigma}T\right]. \tag{12.17}$$

Once the parameters of equation (12.17) have been estimated, calculations of total and marginal bids for individuals and in the aggregate are straightforward using (12.16). The precision with which the parameters will be estimated will depend on the sample size, the variability of preferences in the population (represented by σ), and the bid design (that is, the range and spacing of bids presented to different sub-segments of the sample, see, for example, Alberini 1995).

Alternatively, if η is assumed to be distributed as in the logistic model, with scale parameter μ , then (following Cameron 1988)

$$\Pr(Y)=\Lambda\left[\frac{B^{*}\left(q^{0},q^{1},u^{0},\boldsymbol{S}\right)-T}{\mu}\right], \tag{12.18}$$

where

$$\Lambda(x)\equiv\frac{\exp(x)}{1+\exp(x)} \tag{12.19}$$

denotes the standardized logistic cumulative distribution function. Again, estimation of the parameters of the bid function makes it possible to calculate individual total and marginal willingness to pay as well as the probabilities of accepting an offer.

The logit and probit specifications above, while convenient, impose a considerable structure on the distribution of the bid function in the population, including symmetry around the observable portion of the bid function, B^{*} . To avoid these assumptions, nonparametric procedures exist that can be used to characterize the distribution of B (see, for example, Kriström 1990), and to estimate a lower bound on the mean WTP, such as using the Turnbull estimator (Turnbull 1976). For a more thorough discussion of these and other statistical issues associated with the analysis of discrete choice responses, see Hanemann and Kanninen (1999), Haab and McConnell (2002), or Carson and Hanemann (2005).

Just as in the case of open-ended valuation questions, respondents might express rejection of the discrete choice market scenario described in the survey instrument. *Protest questions*, similar to those described at the end of the discussion on the open-ended format, can be used to identify these individuals. Respondents who signal that their vote is a protest response can be deleted from the sample, though again it would be preferable to understand how these individuals differ from the general population so as to avoid nonresponse bias.

Alternative CV Elicitation Formats

The elicitation formats discussed thus far represent extremes of sort in the CV literature. The open-ended format provides a precise, though potentially biased, value in terms of the individual's WTP for the proposed policy scenario. The single-shot binary choice referendum format, at the other end of the spectrum, yields only limited information on WTP (in the form of upper or lower bounds), but is argued to be easier for the survey respondent to answer and, under certain sets of assumptions, is incentive compatible. The choice between the two represents a classic variance-bias tradeoff, though the NOAA Blue Ribbon Panel argued that binary choice referenda are preferred.

In the years following the NOAA Panel report, researchers have sought a middle ground, trying to squeeze more information out of each survey respondent, while avoiding the problems with pure open-ended questions. One popular variation is the double-bounded discrete choice (DBDC) format. The approach augments the standard binary choice format with a follow-up question, asking the individual to further narrow the range of the willingness to pay, W_i. For example, suppose an individual responds yes to an initial question asking if they would vote for a referendum given a cost of T. The follow-up question would ask if they would still vote for the referendum given a higher cost, say $T_H > T$. On the other hand, if they respond no to the first question, they are then asked if they would vote yes if the cost were lowered to $T_L < T$. The responses, if truthful, provide tighter bounds on the individual's WTP, with

$$\begin{aligned}
&(\text{no,no}) \Rightarrow W_i \in (-\infty, T_L) \\
&(\text{no,yes}) \Rightarrow W_i \in [T_L, T) \\
&(\text{yes,no}) \Rightarrow W_i \in [T, T_H) \\
&(\text{yes,yes}) \Rightarrow W_i \in [T_H, \infty).
\end{aligned} \tag{12.20}$$

These tighter bounds, in turn, yield more precise parameter estimates.

Unfortunately, more data is not the same thing as more useful information. In early applications of the double-bounded format, researchers treated the responses to the first and second questions as independent draws from the same distribution, for example, Hanemann, Loomis, and Kanninen (1991). Subsequently, others subjected this assumption to statistical tests and found that it was not consistent with the data (Cameron and Quiggin 1994), and can lead to biased welfare estimates.[7] In general, the WTP calculated from both questions together is often less than the WTP based on responses to the first question alone. Carson and Groves (2007) offered an explanation based on an examination of the incentive properties of this

7 In some sense, this result should not be surprising. The double-bounded format is essentially a structured version of the bidding game approach used earlier in the literature and found to be potentially biased.

question format, a point returned to below. Herriges and Shogren (1996) provided an explanation based on anchoring.

Other formats have been suggested in the literature, including a one-and-one-half-bound format (Cooper, Hanemann, and Signorello 2002) and the multi-bounded discrete choice format (Welsh and Poe 1998). In each case, the formats try to elicit additional bounds on preferences. However, in doing so, the formats risk the incentive compatibility of the survey questions and open up the possibility of framing effects driven by the structure of the questions themselves. How substantial these effects are remains an empirical question.

Contingent Behavior

As noted in the introduction to this chapter, a fundamental limiting factor often encountered in applications of revealed preference techniques is the lack of sufficient and independent variation in the environmental attributes of interest. Without such variation, it is not possible to identify the causal impact that these attributes have on individual behavior and welfare without imposing additional assumptions on consumer preferences. For example, suppose one had used travel cost data to estimate a demand function for visits to a single recreation site, but one wanted to know the value of a change in one of the environmental attributes of that site. In the absence of an observed variation in the environmental attribute, it might not be possible to predict the shift in the demand curve for visits to the site. Even in those studies with substantial variation in the environmental amenity of interest, either through spatial or temporal variation, it is often difficult to isolate the impact of that amenity. For example, most recreation demand studies have data on only a small number of site attributes, such as a measure of water clarity or fish stock. The problem is that these measures are likely correlated with the many unobserved site attributes, leading to the classic omitted variables bias problem for the included variables.

Contingent behavior studies provide one resolution to this problem by asking survey respondents, not how much they would value proposed policy scenarios, as one would in contingent valuation, but rather how they would change their behavior if the policy changes occurred. For example, in the single site example above, individuals could be asked how their visitation behavior would change if the environmental attribute were to change in a specified way. This contingent change in visitation rates could then be used to estimate the shift in the demand curve for visits. McConnell (1986) used this approach to estimate the benefits to visitors of local beaches if the pollution of New Bedford Harbor, Massachusetts, by polychlorinated biphenyls (PCBs) could be eliminated. For more recent applications, see Cameron et al. (1996), Azevedo, Herriges, and Kling (2003), and Egan and Herriges (2006).

The advantage of the contingent behavior approach is that, if it is an accurate reflection of what individuals would actually do under the hypothetical scenario presented, it not only fills in information that is missing on consumer reactions to alternative environmental conditions, but also it does so in a way that is independent of other unobservable factors. Thus, while actual recreational trips may be impacted by both observed water quality conditions and those that are unobserved, making it hard to identify the effect of water quality alone, the contingent behavior questions can ask how trips would change given *just* a shift in water quality—that is, *holding fixed* all other conditions at the site (see von Haefen and Phaneuf 2008).

The concern with contingent behavior data is that the incentives for truthful responses are less clear. Consider, for example, a proposed policy scenario involving an improvement in water quality at a given site with a fixed cost to consumers (in particular, the cost is not in the form of an entrance fee). An individual who would see the policy as welfare improving might exaggerate their purported future use of the site if they perceived decision-makers would value such use through, say, its economic impact on the surrounding community. There have been relatively few studies examining the consistency between actual trips and those reported through contingent behavior questions (e.g., Azevedo, Herriges, and Kling 2003; Jeon and Herriges 2010; von Haefen and Phaneuf 2008). Generally, these studies have rejected consistency between the two data sources, though the divergence, while statistically significant, has not always been substantial. More research in this area would seem warranted.

Choice Experiments

Attribute-Based Methods

Chapter 4 described the use of the discrete choice model to estimate the parameters of a utility function from individuals' choices of one alternative from a set of J. The discrete choice contingent valuation format described above is a simplified version of this approach in that individuals make their selection of one alternative from a set of two (the alternatives of yes or no). Choice experiments or attribute-based methods of estimating values ask individuals to provide more information about their preferences by giving them more alternatives than the discrete choice approach and by asking them either to select their most preferred option or to rank alternatives in order of preference.[8] Each alternative is described in terms of a series of attributes. Normally one attribute would have a monetary dimension—for example, a price—in order to facilitate calculation of monetary values.

In this respect, choice experiments bear more than a passing resemblance to the random utility models used in the analysis of recreation demand (see Chapter 9). In both cases, the objects of choice are differentiated by embodying different levels

8 Some researchers have asked respondents to rate the alternatives on some scale. See, for example, Roe, Boyle and Teisl (1996) and Boyle et al. (2001). Holmes and Adamowicz (2003) also discussed this approach but conclude that the pick one or ranking formats are to be preferred because people may not be able to translate differences of strength of preference into numerical ratings.

of a set of attributes. By focusing on the tradeoffs among attributes, both methods yield estimates of the marginal rates of substitution between pairs of attributes and, where price is one of the attributes, the marginal willingness to pay for the attribute. Furthermore, with the stated choice method the analyst has experimental control through the design of the attributes presented in the choice set.

One of the earliest applications of choice experiments in economics was that of Beggs, Cardell, and Hausman (1981). They used an ordered logit model to estimate the values of characteristics of alternative models of cars, including electric vehicles. Their respondents were asked to rank sixteen alternative vehicle designs. Each design had nine attributes, including purchase price and fuel costs per mile. The method has subsequently been used in the environmental realm to value rural visibility (Rae 1983), water quality (Smith and Desvousges 1986), and the avoidance of diesel odor (Lareau and Rae 1989). The Smith and Desvousges study was perhaps the simplest for respondents: they were asked to rank only four alternatives, each of which involved only two attributes, a description of water quality and an annual fee. Other studies have posed the simpler task of asking respondents to indicate only their most preferred alternative. Examples include Adamowicz, Louviere and Williams (1994) and Morrison, Bennett, and Blamey (1999).

There has been a significant resurgence of interest in choice experiments in recent years. This stems in part from the rich array of data that it provides relative to the single-shot binary discrete choice contingent valuation exercise. The binary choice CV is typically designed to consider a single specific policy scenario, with each survey respondent providing a single bound on her WTP for the proposed program. In contrast, choice experiments ask respondents to compare a series of alternatives, typically presented using a sequence of pairwise or three-way comparisons (with one alternative often consisting of the status quo). By systematically varying the attributes both across the sequence of alternatives presented to any one individual and across individuals, the analyst can, in theory, estimate the marginal willingness to pay for each attribute and how that WTP interacts with the level of other attributes.

The additional data provided by choice experiments, however, does not come without a cost. Presenting individuals with a sequence of comparisons can result in respondent fatigue, with participants ignoring portions of the information provided and drawing on simpler decision rules to process the complex set of alternatives (see, for example, DeShazo and Fermo 2002). There are also concerns about order effects in the presentation of the choice attributes (see Day et al. 2011 and Boyle and Özdemir 2009). Finally, the conditions under which choice experiments are incentive compatible are more stringent than those required for the single-shot binary choice referendum (Vossler et al. 2011).

Analysis of Stated Choice Responses

The analytical model used to extract information about preferences from the stated choice responses is a straightforward extension of the discrete choice model. This model provides that set of parameter weights on the attributes that maximizes the likelihood of realizing the observed choice (the most preferred alternative or the complete rank ordering). Provided one of the attributes is a money measure, parameter weights can then be used to calculate the marginal willingness to pay for an attribute (the marginal rate of substitution between income and the attribute) or the willingness to pay for a nonmarginal change in its level.

Suppose that an individual is asked to pick the most preferred alternative or to rank a set of J alternatives, with the jth alternative described in terms of price p_j and a vector of K nonprice attributes $\boldsymbol{Q}_j = (q_{1j}, \ldots, q_{Kj})$, where q_{kj} denotes the level of the kth attribute for alternative j. The individual's indirect utility associated with alternative j consists of two components, with

$$u(p_j, \boldsymbol{Q}_j, M, \boldsymbol{S}) = v(p_j, \boldsymbol{Q}_j, M, \boldsymbol{S}) + \varepsilon_j, \tag{12.21}$$

where $v(\cdot)$ denotes the portion of utility that is a function of factors observable by the analyst, whereas ε_j captures the remaining portion of $u(\cdot)$ that is unknown to the analyst and assumed to be random. If the ε_j are independently and identically distributed with a Type I Extreme Value distribution, the probability of the individual picking alternative 1 or ranking it first is

$$\Pr[u_1 > u_j; \text{for all } j \neq 1] = \frac{e^{v_1}}{\sum_{j=1}^{J} e^{v_j}}. \tag{12.22}$$

In the case of contingent ranking, suppose that the individual values the alternatives in the following order from most preferred: $r_1, \ldots, r_J$, where r_m denotes the number of the alternative given rank m. The probability of this ranking is the probability of ranking r_1 above the remaining $J - 1$ alternatives times the probability of ranking alternative r_2 above the remaining $J - 2$ alternatives, and so forth. That is,

$$\Pr[u_{r_1} > u_{r_2} > \cdots > u_{r_J}] = \frac{e^{v_{r1}}}{\sum_{j=1}^{J} e^{v_j}} \cdot \frac{e^{v_{r2}}}{\sum_{j=2}^{J-1} e^{v_{r_j}}} \ldots = \prod_{j=1}^{J} \left[\frac{e^{v_{rj}}}{\sum_{l=j}^{J} e^{v_{r_l}}} \right]. \tag{12.23}$$

If, as is typically the case, the survey respondents are asked to make a sequence of comparisons of such choices or rankings, the corresponding probability expressions become more complex, particularly if one allows for correlations across the responses made by the same individual across different sets of comparisons (as one should). Discussion of these issues can be found in Holmes and Adamowicz (2003).

As in the other applications of the discrete choice model, once a functional form for the observable component of the indirect utility function has been specified, its parameters can be estimated from data on the choices or rankings of a sample of individuals using maximum likelihood methods. Welfare measures can then be derived by computing the marginal rate of substitution between income and any of the elements of $\boldsymbol{Q}$.

Assessing the Validity of Stated Preference Welfare Measures

As Mitchell and Carson (1989, 190) put it, "The validity of a measure is the degree to which it measures the theoretical construct under investigation." The theoretical construct of interest here is the individual's true value, her true probability of accepting an offer or selecting one of the alternatives offered, her true ranking of alternatives, or her true change in the level of an activity. Ideally, one would like to assess the validity of a stated preference measure by comparing it with the true measure. But at least in the cases of monetary values (especially those including passive use values), probabilities, and rankings, the true measure cannot be known, so this option is not available. In the case of contingent behavior questions, we will not often be able to observe "true" changes in activity levels under precisely the same conditions that prevailed when the question was asked. So, much of the discussion of validity in the literature focuses on less direct approaches to validity assessment, such as asking whether the results are consistent with economic theory (construct validity) or whether the study was conducted in accordance with generally accepted principles (content validity).

In this section, four types of validity are described along with the approaches to assessment for each type. These are not mutually exclusive alternatives but should be seen as focusing on different aspects of the way a study is conducted and of the information generated by the study. For each type of validity, a summary is provided in terms of what can be said at present about the validity of stated preference methods in general, as distinct from the validity of particular studies. For additional discussion of validity and a review of the evidence available as of the date of writing, see Mitchell and Carson (1989). The reader should also consult Kling, Phaneuf, and Zhao (2012), Carson (2012), Carson and Hanemann (2005), and some of the essays included in Bjornstad and Kahn (1996) and Kopp, Pommerehne, and Schwarz (1997). Bishop (2003) also includes an insightful discussion of concepts of validity.

Criterion Validity

The Concept

Assessing criterion validity involves comparing the value as measured by the stated preference method with some alternative measure that can be taken as the criterion for assessment. It is tempting to think of the "true" value as being the appropriate criterion. However, the "true" value is a theoretical construct: the sum that creates a state of indifference between two alternative states of the world; and as such, it is not directly observable (Bishop 2003). Instead, researchers have turned to experimental methods, in field or laboratory settings, to create simulated markets in which individuals can engage in actual transactions. In simulated or experimental markets, transactions actually take place and respondents live with the consequences of their choices. These experiments allow the comparison of responses to hypothetical questions with responses in similarly designed experimental or simulated markets. These authors have used the observed values from these transactions as the criterion for comparison with stated preference responses.

The Evidence

One of the pioneering studies of this type was conducted by Bishop and Heberlein (1979) who made hypothetical closed-ended offers to purchase special goose hunting permits from a sample of hunters. At the same time, they made actual cash offers for a limited number of permits from a different sample, so they could compare the values revealed. The mean value of a permit calculated from the responses to the hypothetical offer (a WTA measure) was about 60 percent higher than the mean revealed by the real cash offers. Similar findings have emerged in subsequent lab and field experiments, including studies by Neill et al. (1994), Cummings, Harrison, and Rutström (1995), and Cherry et al. (2004). Indeed, Harrison and Rutström (2008) noted that roughly three-fourths of such studies find that stated preference values are biased upwards.

List and Gallet (2001) provided one the first systematic analyses of this line of research, reviewing the results of 29 studies dealing with the provision of both private and public goods, including that of Bishop and Heberlein. They used a meta-analysis to explain the reported calibration ratios; that is, the ratios of minimum, median, and maximum hypothetical values to the corresponding values from the real transactions. Although the median calibration ratio for their sample was about 3.0, they found that using a WTP rather than WTA question and using a first-price sealed bid auction mechanism both had statistically significant effects in lowering the calibration ratio. A more recent meta-analysis by Murphy et al. (2005) came to similar conclusions, though they found a substantially smaller median calibration ratio of 1.35. In any case, the persistent gap between hypothetical and real cash transactions documented in both lab and field experiments has come to be referred to as "hypothetical bias" and has been a central point of concern regarding the application of stated preference methods in welfare analysis.

The response to the experimental evidence of criterion invalidity (in the form of hypothetical bias) has taken two forms in the stated preference literature. First, some authors have investigated whether changes in the design or interpretation of stated preference surveys can mitigate or eliminate hypothetical bias. Second, and more recently, a number of authors (Carson and Groves 2007; Vossler and Poe 2011) have come to question the value in comparing purely hypothetical and actual transactions, arguing that stated preference questions require an element of consequentiality in order to be incentive compatible. These two forms of responses are discussed in turn.

REDUCING HYPOTHETICAL BIAS

A number of strategies have emerged in the literature designed to ameliorate hypothetical bias in stated preference surveys. Ex ante approaches seek to alter the survey process so as to encourage truthful answers from the survey respondent, while ex post strategies use the survey data itself to adjust either individual responses or the aggregate measure of WTP derived from the survey.

The most popular of the ex ante strategies is the use of a so-called "cheap talk" script. With this procedure, survey participants are explicitly warned of the tendency for hypothetical bias in stated preference surveys and asked to treat the survey "as if" it were real. Cummings and Taylor (1999) found that, in an experimental setting, their cheap talk script was successful in inducing hypothetical valuation responses that were not statistically different from comparable responses involving real cash payments. However, subsequent applications have found the efficacy of cheap talk to be somewhat idiosyncratic, varying with the length of the cheap talk script used (Aadland and Caplan 2003), the experience of survey respondents (List 2001), and other factors. See Morrison and Brown (2009) for a useful summary. Taking a different tack, Jacquemet et al. (2013) used an oath script, asking respondents to promise to tell the truth in responding to the survey questionnaire. They found that the oath script was successful in inducing responses that are consistent with parallel real transaction responses. In the case of both "cheap talk" and "oath scripts," however, more research is needed into why it does (or does not) work before it can be reliably used to close the gap between hypothetical and actual transactions responses.

A number of ex post approaches to dealing with hypothetical bias have also emerged in the literature. One tack is to adjust values obtained from a stated preference study by a calibration factor (see, for example, Diamond and Hausman 1994, 54) to reflect the observed disparity between hypothetical and real transactions. While appealing in its apparent simplicity, it is not clear which calibration factor one would use in a given setting. Experimental comparisons of hypothetical and real transactions can provide some guidance, though these are often conducted using private goods, while stated preference work is particularly useful in the context of public goods. Moreover, the meta-analyses of List and Gallet (2001) and Murphy et al. (2005) suggest that the calibration factors can vary substantially across settings. Using data from their field experiment, List and Shogren (1997, 194) concluded that "calibration is good- and context-specific."

A second ex post approach to controlling for hypothetical bias involves recoding survey responses based upon the self-reported certainty of the respondent in answering discrete choice referendum questions. A variety of formats have been used to allow individuals to express their degree of confidence, from qualitative ratings (such as the "definitely sure" and "probably sure" categories used by Blomquist, Blumenschein, and Johannesson, 2009) to numerical scales (such as the "1 very unsure" to "10 very certain" scale used by Champ et al. 1997). Numerous studies have found that recoding some or all "uncertain" yes responses to no responses yields a better match between hypothetical surveys and their actual transactions counterpart (see, for example, Champ et al. 1997; Champ and Bishop 2001; Blumenschein et al. 1998; Poe et al. 2002), though the precise cutoff used for recoding varies across the studies.

There are a number of concerns with this type of ex post recoding. First, it is not clear why it works and which specific cutoff is appropriate in a given setting. The argument is typically that the "uncertain" yes's do not represent a true commitment to purchase the policy option being offered, but rather are a form of "yea-saying." The fact that it "works" in experimental settings is relatively little comfort, since a variety of other bases for recoding would do the same, as long as they reduced the proportion of yes responses in the survey. One might, for example, argue (facetiously) that older survey respondents are easily confused by complex survey instruments and recode all their yes's to no's. If age was not typically a factor in WTP, there would exist an age cutoff that would successfully yield a match between hypothetical and actual referenda responses, but hopefully few would actually buy the senility argument. Second, if uncertainty is indeed a cause of hypothetical bias, then the random utility maximization (RUM) model typically used to analyze such discrete choice responses becomes problematic in that the RUM framework assumes that individuals know precisely the utility (or at least the expected utility) associated with each option. Understanding the nature of respondent uncertainty, and how it varies across individuals, becomes important in modeling the decision process used in choosing among alternatives. This remains an area in need of both theoretical and empirical research.

FOCUSING ON INCENTIVES INSTEAD

In recent years, a number of authors (Carson and Groves 2007; Vossler and Poe 2011) have argued that experimental comparisons between purely hypothetical and real transactions should not be the point of departure in evaluating stated preference techniques. Rather, one should start by asking what incentives a given elicitation procedure creates for truthful revelation of the individual's preference. In particular, Carson and Groves (2007, 183) distinguished between two broad categories of survey questions:

- Consequential survey questions: If a survey's results are seen by the respondent as potentially influencing an agency's actions and she cares about the outcomes of those actions, she should treat the survey questions as an opportunity to influence those actions. In such a case, standard economic theory applies and the response to the question should be interpretable using mechanism design theory concerning incentive structures …[and]
- Inconsequential survey questions: If a survey's results are not seen as having any influence on an agency's actions or the agent is indifferent to all possible outcomes of the agency's actions, then all possible responses by the agent will be perceived as having the same influence on the agent's welfare. In such a case, economic theory makes no predictions.

There are two important implications of this dichotomy. First, purely hypothetical stated preference surveys are uninformative and should not be used. Indeed, Carson and Groves (2007, 183) went out of their way to "formally reject the notion, sometimes advanced by proponents of preference surveys, that when a respondent perceives no gain or loss from how a preference survey is answered, the respondent always answers truthfully. While such an assumption may be true, there is no basis in economic theory to either support or deny it." A corollary to this is that the experimental comparisons between purely hypothetical and actual cash transactions that have proliferated in the literature shed little, if any, light on the validity of properly designed and consequential stated preference surveys.

Second, alternative elicitation formats used in a consequential survey instrument should be evaluated in terms of well-known results from mechanism design theory. A consequential survey instrument is not necessarily incentive compatible—that is, it need not encourage truthful revelation of preferences. However, understanding the incentives of respondents is key to understanding the direction of bias, if any, one should expect from their responses. For example, Carson and Groves (2007) argued that, when asked about their WTP for a proposed *private* good, respondents have an incentive to say yes to a given bid (T), not because their WTP for the new good necessarily exceeds T, but by saying yes they encourage the supply of the good and create the option to consider its purchase at a later date. Thus, one should expect an upward bias in the estimated WTP derived from such a survey instrument.

Drawing on the results from mechanism design theory, Carson and Groves (2007) went on to examine several of the elicitation formats commonly used in the stated preference literature. They concluded that many of the formats, including open-ended, double-bounded, and multinomial choice formats, are *not* incentive compatible. However, the single-shot binary choice referendum format is incentive compatible if (a) the probability that the proposed project is implemented is increasing in the proportion of yes votes, and (b) the agency proposing the policy scenario can enforce payment at the proposed price (T). Experimental evidence to date appears to corroborate this claim (see, for example, Carson et al. 2004; Landry and List 2007; Vossler and Evans 2009). Herriges et al. (2010) provided additional support in the context of a single-shot dichotomous choice referendum question. More research, however, is needed in terms of how to design survey instruments that meet the incentive compatibility conditions. Vossler et al. (2011) provided a similar discussion regarding the conditions for incentive compatibility in the context of discrete choice experiments.

Convergent Validity

The Concept

A related form of validity assessment is the comparison of stated preference values with measures derived from some revealed preference method. This provides an assessment of convergent validity. This strategy is as old as the contingent valuation method itself (see Knetsch and Davis 1966). The problem with this measure of validity, of course, is that if a discrepancy between the stated preference value and the alternative is discovered, it could be because of the lack of validity of the stated preference measure, the lack of validity of the alternative measure derived from observation, or both (see, for example, Azevedo, Herriges, and Kling 2003).

The Evidence

Early reviews of this evidence can be found in Cummings, Brookshire, and Schulze (1986) and Mitchell and Carson (1989). More recently, Carson et al. (1996) provided a large study of convergent validity of contingent valuation estimates of WTP for quasi-public goods (goods provided by the government for which exclusion of users is possible). The estimates covered were for three types of goods: access to outdoor recreation activities, environmental amenities, and changes in health risks. They found that the ratio of CV to RP values averaged 0.89 for the whole sample with a 95 percent confidence interval of 0.81–0.96. Based on their evidence, the authors argued against systematic calibration of CV values and for convergent validity.

Construct Validity

The Concept

Construct validity asks whether stated preference responses are related to variables that economic theory suggests should be predictors of WTP, and invariant to factors that theory suggests should be irrelevant. For example, it is usually thought that values should be an increasing function of income, other things being equal, and that the estimated income elasticity of WTP should be "reasonable." When an experimental design involves facing different individuals with different quantities and/or implied prices, the choices should also be consistent with the axioms of revealed preference, such as transitivity (see, for example, Adamowicz and Graham-Tomasi 1991). Construct validity can be assessed by regressing expressed values on characteristics of the good being valued and characteristics of the respondent. Depending upon the stated preference approach, these assessments may involve comparisons *across* randomly assigned groups that were exposed to different treatments (such as the price, payment mechanism, and the scope of the proposed policy scenario). Other elicitation formats that expose participants to a series of choices (such as in discrete choice experiments) allow *within* individual validity checks.

While there are a myriad of possible validity checks one might consider, the literature has focused on three particular areas of concern, involving the sensitivity of stated preference valuations to

- income,
- scope, and
- the use of WTP versus WTA measures.

Each of these topics is considered in turn below.

Evidence on Sensitivity to Income

It is generally thought that environmental amenities represent luxury goods, with income elasticities that would exceed one. The fact that WTP is often found to be inelastic with respect to income has been cited as evidence of the construct invalidity of stated preference techniques (McFadden 1994). However, Flores and Carson (1997) showed that the relationship between the income elasticity of demand for a quantity constrained good and the corresponding income elasticity of WTP is not as simple as one might think. They showed that in the case of more than one quantity-constrained good, the income elasticity of WTP for a good depends not only on the income elasticities of demand for all the quantity-constrained goods but also on the cross-price substitution effects for those goods, and on the expenditures for market goods as a percentage of virtual income. This latter term will always be less than one, leading to a likelihood that the income elasticity of WTP will be less than the income elasticity of demand. However, they also showed by example that plausible values for the three determinants of the income elasticity of WTP can lead to values greater than the income elasticity of demand, or even negative values for so-called normal goods. So very small or even negative values for the income elasticity of WTP for a nonmarket good are not prima facie evidence of lack of construct validity.

Evidence on Sensitivity to Scope

It is generally assumed that an increase in the "quantity" of the good being valued should produce an increase in the magnitude of expressed values. This increased quantity can take the form of a literal increase in the number of environmental amenities provided (such as the number of birds saved or acres of wetlands preserved). Alternatively, the increase can be in the number of distinct subprograms proposed (going from, say, a program to provide improved rescue equipment and trained personal to a package of programs in education, safety, etc., that includes the safety equipment and personnel as a component, as in Kahneman and Knetsch 1992).[9] In either case, tests of the hypothesis that value increases with the size of the proposed policy have come to be known as scope tests. The NOAA Panel (Arrow et al. 1993) included conducting a scope test in its recommended guidelines.

Inadequate response of expressed WTP to changes in scope is often cited as a reason to reject stated preference techniques by critics of the methods, for example Diamond and Hausman (1994) and Hausman (2012). Yet, several authors have reviewed the evidence from scope tests and found that they can reject

9 This latter type of scope change is also referred to as "nesting."

the hypothesis of inadequate responsiveness of expressed values to changes on scope (see Smith and Osborne 1996, Kopp and Smith 1997, Brouwer et al. 1999, and Carson 1997). Ojea and Loureiro (2011) found scope sensitivity in their meta-analysis of biodiversity valuation studies, and Lew and Wallmo (2011) found scope sensitivity in a choice experiment setting. Moreover, a number of the early studies purporting to show scope insensitivity have come under criticism in terms of both survey design and analysis (see, for example, Harrison 1992). Powe and Bateman (2004) also provided experimental evidence that an erroneous scope insensitivity finding may result from survey respondents questioning the realism of the larger programs and not because the programs themselves are not valued.

Heberlein et al. (2005) suggested that the current emphasis on scope testing is misplaced and that, in many instances, a finding of scope insensitivity is perfectly consistent with both psychological and economic theory. Using an example of species preservation, they note that an individual may be willing to pay for the preservation and protection of 300 wolves, but would be reluctant to support a population growth to 800 wolves, fearing the increased potential for negative externalities. Through a series of such scenarios, Heberlein et al. (2005, 20–21) illustrated the fact that values can depend in complex ways on scope and how it is communicated, and that "more is better" is too simplistic a characterization of preferences. They conclude "that, by itself, failure to pass a scope test is neither a necessary nor sufficient condition to invalidate any given CV study. And certainly the scope test failures reported so far in the literature do not undermine the validity of the entire CV method."

Evidence on the WTP/WTA Gap

SP questions can be asked either in WTP or WTA formats. A large body of evidence suggests that WTP and WTA formats for stated preference values lead to quite different results. WTA responses are typically several times larger (and, in at least one case, two orders of magnitude larger) than WTP responses for the same change. See Cummings, Brookshire, and Schulze (1986, especially 35) for a review of earlier studies comparing WTP and WTA. Brookshire and Coursey (1987) also found large differences. Some have argued that the disparity between WTP and WTA responses is evidence that SP methods lack construct validity.

There are two basic responses to this concern in the literature. The first response is that, as discussed in Chapter 3, there are theoretically plausible reasons why there could be large differences between WTP and WTA, at least in the case of environmental resources and amenities. For example, Hanemann (1991) noted that the gap between WTP and WTA will be larger, all else equal, if there are few substitutes for the commodity in question, which is likely to be the case for many environmental amenities. Shogren et al. (1994) provided empirical support for this argument using laboratory experiments. Zhao and Kling (2004), on the other hand, argued that dynamic considerations, involving uncertainty and the opportunity for learning, can also drive a wedge between WTP and WTA. While both of these arguments are valid, the counterargument is that it is the size of the gaps, not their direction, that are implausible. Specifically, Sugden (1999) derived the relationship

$$1-\frac{\text{WTP}}{\text{WTA}} \approx \frac{\partial \text{WTP}}{\partial M} \tag{12.24}$$

where M denotes income. Using this result, Horowitz and McConnell (2003, 539) examined the WTA/WTP ratios available in the literature and concluded that they "imply income effects that are implausibly high, much higher than income effects found in valuation studies; and much higher than income effects estimated in the set of studies from which the WTA/WTP results are drawn. Based on this evidence, we reject the claim that observed WTA/WTP ratios are consistent with a standard neoclassical model."

The second response to the disparity between WTA and WTP is to note that such differences are not limited to stated preference settings, but have also arisen in lab and field experiments involving real transactions. Indeed, in their review of WTA/WTP ratios found in the literature, Horowitz and McConnell (2002) found that real experiments do not yield significantly lower ratios. Thus, the observed disparities between WTP and WTA are not prima facie evidence of lack of construct validity for stated preference techniques per se, but perhaps raise broader concerns regarding neoclassical theory.

Content Validity

The Concept

Content validity refers to the extent to which the design and implementation of the survey conform to the generally recognized best practice or state of the art. It should involve an examination of the survey instrument, including the scenario specification, the elicitation question—especially its incentive properties—and the payment vehicle, as well as procedural matters such as sample size and design and the analysis of data. For a more complete discussion of the assessment of content validity, see Bishop and McCollum (1997).

Practitioners in the field of stated preference valuation methods have identified a number of problem areas and issues that can affect the validity of responses. There is general agreement, for example, that the form of the valuation question is important. Discrete choice questions work better than the bidding game format. WTP questions work better than WTA questions, and a number of sources of potential bias have been identified. Thus, much can be learned from a systematic examination of the survey instrument.

Mitchell and Carson (1989, especially Chapter 11) provided a classification of ways in which scenarios can affect individuals' responses so as to introduce systematic error or bias into the value measures that are derived from these responses. Systematic errors arise from several sources. One is scenario misspecification. Were the question and the information provided with it to establish its context understood by the respondent in the way intended by the investigator? A second source is implied value cues. To what extent did the information provided to the respondent predispose her toward a particular response? A third is the choice of a payment vehicle. Did the payment vehicle tap into attitudes about government and taxation or business pricing policies that might bias responses? Finally, were there incentives for misrepresentation of values? Part of the art of asking stated preference questions involves avoiding problems such as these. In what follows, each of these sources of error is taken up in turn.

Scenario Misspecification

The various types of scenario misspecification discussed by Mitchell and Carson (1989, 246–258) all have the effect of creating a divergence between what the respondent understands about the choice situation and what the investigator intends him to understand. The investigator can increase the likelihood that the questions will be understood as intended by developing an awareness of how people think about the issues at hand and what language and terminology they use in talking about it. This can be accomplished through the use of focus groups, pilot surveys, pretests, and the like. Focus groups involve structured interviews with small groups of individuals from the population to be surveyed. Focus groups can be used to investigate perceptions of environmental problems, to determine which terminologies have particular salience, and to try out various descriptions and question formats. Pretests and pilot surveys may be useful in testing different scenarios and question formats, especially if respondents are queried afterwards to identify questions and problems they may have with specific wording.

Implied Value Cues

If a respondent has well-formed preferences regarding the hypothetical choice problem posed to him, he should find the task of formulating a response to the question fairly easy. However, if the choice problem posed to the individual is unfamiliar and the individual is not clear about her preferences, she might, quite unconsciously, seek clues regarding the "correct" choice or value from the information supplied as part of the scenario specification. If such clues are present, they may systematically bias individuals' choices and values. Such clues are termed implied value cues. The most well-studied implied value cue is the starting point in bidding game versions of direct expression of value questions. Other forms of direct value questions may avoid starting point biases.

Implied value cues can be found outside of the scenario itself. For example, respondents who are asked about the value of access to a recreation site or to a fishing experience may anchor their responses to the implied value provided by the normal admission fee or fishing license fee. The task for the investigator is to find ways to thwart this tendency, perhaps by specifying a payment vehicle that focuses attention away from the customary admission fee or license fee. For example, recreationists might be asked about their willingness to pay higher travel costs to reach a site.

Stated choice and contingent behavior questions appear to be less susceptible to implied value cues. However, there could be a tendency for the results of a stated choice question to depend upon the order in which options are initially presented for ranking. Order effects can be examined by randomizing the sequence in which alternatives are presented to respondents, allowing the researcher to test for order effects explicitly. If order effects are determined to exist, it may only be possible to bound the values associated with an individual alternative.

Payment Vehicle

Another feature of scenario design that may be important is the specification of a payment vehicle for collecting the respondents' willingness to pay. In order to make contingent market scenarios plausible to respondents, it is necessary to inform them of how their stated value will be collected. Examples of payment vehicles used in past studies include an increase in the prices of all goods resulting in a higher expenditure of $\$X$, an increase in taxes of $\$X$, and a surcharge on utility bills of $\$X$. If the vehicle involves a higher price for a market good, the question must be clear about whether it refers to a change in expenditure or a change in price. If the question is about a change in price, then the welfare calculation must take account of the reduction in quantity that a price increase will cause. In some studies where respondents were asked for a maximum increase in a price per unit, total willingness to pay was calculated by multiplying the increase in price by the observed quantity. However, assuming that respondents accurately stated the maximum price increase they would accept, the correct welfare measure is the area to the left of the demand curve between the two prices. This will be less than the original quantity times the increase in price.

There is evidence that the specification of a payment vehicle can exert an independent influence on bids, a phenomenon that has been termed "vehicle bias" in the literature. However, judging whether this influence is a bias in the usual sense of the word is a more difficult matter, since the respondent's "true" value is probably not known. The influence of the payment vehicle on bids is one example of the importance of what Fischhoff and Furby (1988) have called the social context of the hypothetical transaction. They argued that in both designing and interpreting surveys that elicit hypothetical values, researchers must be sensitive to how individuals understand and interpret the environmental good being valued, the payment they are being asked to consider, and the social context in which the exchange takes place.

Conclusions

Economists and other social scientists have made a great deal of progress in the design and implementation of stated preference methods since the earliest contingent valuation studies of the mid-1960s and early 1970s. One major advance was the development and examination of a wide range of elicitation formats, broadly grouped as contingent valuation methods, choice experiments, and contingent behavior methods. Each of these approaches has its strengths and limitations relative to the others; but together they provide a more powerful array of tools for dealing with a variety of valuation problems.

A second advance was the development of a much greater awareness of, and sensitivity to, the ways in which scenario specification can influence responses and thereby affect the validity of the value measures. Practitioners in economics owe a debt to researchers from other fields, especially those in cognitive psychology, for their contributions to our understanding of how people respond to questions and of the importance of how questions are asked.

A third advance was the development of more sophisticated approaches to the analysis of data and the calculation of welfare measures, especially with respect to discrete choice models. Fourth, there has developed a much greater sensitivity to a variety of sampling issues and related problems such as the need to deal with item nonresponse and to identify and treat outliers and protest zeros. Finally, researchers are developing a variety of methods for combining stated preference and revealed preference data into one statistical model, a topic that is covered in greater detail in Chapter 13.

All of these developments have helped to make the set of stated preference methods a more powerful tool for dealing with amenity and resource valuation problems. Yet, stated preference methods continue to be controversial. On the one hand, the NOAA "Blue Ribbon Panel" concluded that "under those conditions [that were discussed earlier in the Report], CV studies convey useful information. … [and] can produce estimates reliable enough to be the starting point of a judicial process of damage assessment" (Arrow et al. 1993, 4610). This conclusion is in sharp contrast to that of Diamond and Hausman (1994, 62), "that contingent valuation is a deeply flawed methodology for measuring nonuse values, one that does not estimate what its proponents claim to be estimating." Indeed, Hausman's position has changed little over the past two decades. He concluded recently that "'no number' is still better than a contingent valuation estimate… [U]nless or until contingent value studies resolve their long-standing problems, they should have zero weight in public decision-making" (Hausman 2012, 54).

Our own assessment of SP methods is cautiously optimistic. On the one hand, stated preference techniques should not be viewed as a panacea. They do not provide a "simple" alternative to the daunting, and sometimes impossible, task of valuing changes to environmental amenities based on revealed preference data alone. Designing, implementing, and analyzing a stated preference survey requires careful attention to detail, with an unusual mix of both art and science to be successful. Reading this chapter alone is certainly not sufficient to equip one to do a SP study well. There is no substitute for a careful reading in the now substantial and rapidly growing literature—including, but not limited to, the sources cited here. Having said that, the preponderance of evidence suggests that a carefully executed SP study provides valuable insights into the tradeoffs that individuals are willing to make to secure or avoid changes to the environment. To ignore these tradeoffs, or to rely solely on expert opinion regarding such tradeoffs (as Diamond and Hausman 1994, 56, suggested), would seem counterproductive.

References

Aadland, David, and Arthur J. Caplan. 2003. "Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias." *American Journal of Agricultural Economics* 85(2): 492–502.

Adamowicz, Wiktor L., and Theodore Graham-Tomasi. 1991. "Revealed Preference Tests of Nonmarket Goods Valuation Methods." *Journal of Environmental Economics and Management* 20(1): 29–45.

Adamowicz, Wiktor, J. Louviere, and M. Williams. 1994. "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities." *Journal of Environmental Economics and Management* 26(3): 271–292.

Alberini, Anna. 1995. "Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-Bound, Double-Bound, and Bivariate Models." *Journal of Environmental Economics and Management* 28(3): 287–306.

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58(10): 4601–4614.

Azevedo, Christopher D., Joseph A. Herriges, and Catherine L. Kling. 2003. "Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations." *American Journal of Agricultural Economics* 85(3): 525–537.

Bateman, Ian J., and Kenneth G. Willis (eds). 1999. *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*. Oxford: Oxford University Press.

Bateman, Ian J., Richard T. Carson, Brett Day, W. Michael Hanemann, Nick Hanley, Tanis Hett, Michael J. Lee, Graham Loomis, Susana Mourato, Ece Özdemiroglu, and David W. Pearce. 2002. *Economic Valuation with Stated Preference Techniques: A Manual.* Northampton: Edward Elgar.

Beggs, S., S. Cardell, and J. Hausman. 1981. "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics* 17(1): 1–19.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2008. "How Are Preferences Revealed?" *Journal of Public Economics* 92(8–9): 1787–1794.

Bishop, Richard C., 2003. "Where to from Here?" In *A Primer on Non-market Valuation*, edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Boston: Kluwer Academic.

Bishop, Richard C., and Thomas A. Heberlein. 1979. "Measuring Values of Extra Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61(5): 926–930.

Bishop, Richard C., and Daniel W. McCollum. 1997. "Assessing the Content Validity of Contingent Valuation Studies." Working paper, Department of Agricultural and Applied Economics, University of Wisconsin.

Bjornstad, David J., and James R. Kahn (eds). 1996. *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs* Cheltenham: Edward Edgar.

Blomquist, Glen C., Karen Blumenschein, and Magnus Johannesson. 2009. "Eliciting Willingness to Pay without Bias using Follow-up Certainty Statements: Comparisons between Probably/Definitely and a 10-point Certainty Scale." *Environmental & Resource Economics* 43(4): 473–502.

Blumenschein, Karen, Magnus Johannesson, Glen C. Blomquist, B. Liljas, and R.M. O'Conor. 1998. "Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation." *Southern Economic Journal* 65(1) (July): 169–177.

Boyle, Kevin J. and Semra Özdemir. 2009. "Convergent Validity of Attribute-Based, Choice Questions in Stated-Preference Studies." *Environmental & Resource Economics* 42(2): 247–264.

Boyle, Kevin J., Thomas P. Holmes, Mario Teisl, and Brian Roe. 2001. "A Comparison of Conjoint Analysis Response Formats." *American Journal of Agricultural Economics* 83(2): 441–454.

Brookshire, David S., and Don L. Coursey. 1987. "Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures." *American Economic Review* 77(4): 554–566.

Brouwer, Roy, Ian H. Langford, Ian J. Bateman, and R. Kerry Turner. 1999. "A Meta-Analysis of Wetland Contingent Valuation Studies." *Regional Environmental Change* 1(1): 47–57.

Cameron, Trudy A. 1988. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." *Journal of Environmental Economics and Management* 15(3): 355–379.

Cameron, Trudy A., and Michelle D. James. 1987. "Efficient Estimation Methods for Contingent Valuation Surveys." *Review of Economics and Statistics* 69(2): 269–276.

Cameron, Trudy Ann, and John Quiggin. 1994. "Estimation Using Contingent Valuation Data from a 'Dichotomous Choice with Follow-Up' Questionnaire." *Journal of Environmental Economics and Management* 27(3): 218–234.

Cameron, Trudy A., W. Douglass Shaw, Shannon E. Ragland, J. Mac Callaway, and Sally Keefe. 1996. "Using Actual and Contingent Behavior Data with Differing Levels of Time Aggregation to Model Recreation Demand." *Journal of Agricultural and Resource Economics* 21(1): 130–149.

Carson, Richard T. 1991. "Constructed Markets." In *Measuring the Demand for Environmental Quality*, edited by John Braden and Charles Kolstad. Amsterdam: Elsevier.

Carson, Richard T. 1997. "Contingent Valuation Surveys and Tests of Insensitivity to Scope." In *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, edited by Raymond J. Kopp, Werner W. Pommerehne, and Norbert Schwarz. Boston: Kluwer Academic.

Carson, Richard T. 2011. *Contingent Valuation: A Comprehensive History and Bibliography*. Northampton: Edward Elgar.

Carson, Richard T. 2012. "Contingent Valuation: A Practical Alternative when Prices Aren't Available." *Journal of Economic Perspectives* 26(4): 27–42.

Carson, Richard T., and Theodore Groves. 2007. "Incentive and Informational Properties of Preference Questions." *Environmental and Resource Economics* 37(1): 181–210.

Carson, Richard T., and W. Michael Hanemann. 2005. *Contingent Valuation Handbook of Environmental Economics*, edited by K.G. Mäler and J.R. Vincent. Elsevier.

Carson, Richard T., Nicholas E. Flores, Kerry M. Martin, and Jennifer L. Wright. 1996. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics* 72(1): 80–99.

Carson, Richard T., Theodore Groves, John List, Mark J. Machina. 2004. "Probabilistic Influence and Supplemental Benefits: A Field Test Of The Two Key Assumptions Underlying Stated Preferences." Paper Presented at the European Association of Environmental and Resource Economists Meeting. Budapest: EAERE.

Champ, Patricia A., and Richard C. Bishop. 2001. "Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias." *Environmental and Resource Economics* 19(4): 383–402.

Champ, Patricia A., Richard C. Bishop, Thomas C. Brown, Daniel W. McCollum 1997. "Using Donation Mechanisms to Value Nonuse Benefits From Public Goods." *Journal of Environmental Economics and Management* 33(2): 151–162.

Champ, Patricia A., Kevin J. Boyle, and Thomas C. Brown (eds). 2003. *A Primer on Non-Market Valuation.* Boston: Kluwer Academic Press.

Cherry, Todd L., Peter Frykblom, Jason F. Shogren, John A. List and Melonie B. Sullivan. 2004. "Laboratory Testbeds and Non-Market Valuation: The Case of Bidding Behavior in a Second-Price Auction with an Outside Option." *Environmental & Resource Economics* 29(3): 285–294.

Cooper, Joseph C., W. Michael Hanemann, and Giovanni Signorello. 2002. "One-and-One-Half-Bound Dichotomous Choice Data." *Review of Economics and Statistics* 84(4): 742–750.

Cummings, Ronald G., and Laura O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89(3): 649–665.

Cummings, Ronald G., David S. Brookshire, and William D. Schulze (eds). 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.* Totowa: Rowman and Allanheld.

Cummings, Ronald G., Glenn W. Harrison and E. Elisabet Rutström. 1995. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?" *American Economic Review* 85(1): 260–266.

Davis, R.K. 1963. "The Value of Outdoor Recreation: An Economic Study of the Main Woods." Dissertation, Harvard University.

Day, Brett, Ian J. Bateman, Richard T. Carson, Diane Dupont, Jordan J. Louviere, Sanae Morimoto, Riccardo Scarpa, and Paul Wang. 2011. "Ordering Effects and Choice Set Awareness in Repeat-Response Stated Preference Studies." *Journal of Environmental Economics and Management* 63(1): 73–91.

DeShazo, J.R., and German Fermo. 2002. "Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency." *Journal of Environmental Economics and Management* 44(1): 123–143.

Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better Than No Number?" *Journal of Economic Perspectives* 8(4): 45–64.

Egan, Kevin, and Joseph A. Herriges. 2006. "Multivariate Count Data Regression Models with Individual Panel Data from an On-Site Sample." *Journal of Environmental Economics and Management* 52(2): 567–581.

Fischhoff, Baruch, and Lita Furby. 1988. "Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility." *Journal of Risk and Uncertainty* 1(2): 147–184.

Flores, Nicholas E., and Richard T. Carson. 1997. "The Relationship Between the Income Elasticities of Demand and Willingness to Pay." *Journal of Environmental Economics and Management* 33(3): 287–295.

Haab, Timothy C., and Kenneth E. McConnell. 2002. *Valuing Environmental and Natural Resources.* Northampton: Edward Elgar.
Hanemann, W. Michael. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66(3): 332–341.
Hanemann, W. Michael. 1991. "Willingness To Pay and Willingness To Accept: How Much Can They Differ?" *American Economic Review* 81(3): 635–647.
Hanemann, W. Michael. 1994. "Valuing The Environment Through Contingent Valuation." *Journal of Economic Perspectives* 8(4): 19–43.
Hanemann, W. Michael, and Barbara Kanninen. 1999. "The Statistical Analysis of Discrete-Response CV Data." In *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, edited by Ian J. Bateman and Kenneth G. Wills. Oxford: Oxford University Press.
Hanemann, W. Michael., John Loomis, and Barbara Kanninen. 1991. "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation." *American Journal of Agricultural Economics* 73(4): 1255–1263.
Harrison, Glen W. 1992. "Valuing Public Goods with the Contingent Valuation Method: A Critique of Kahneman and Knetsch." *Journal of Environmental Economics and Management* 23(3): 248–257.
Harrison, Glen W., and E. Elisabet Rutström. 2008. "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods." In *Handbook of Experimental Economics Results, vol. 1, part 5*, edited by Charles R. Plott and Vernon L. Smith. Chapter 81. New York: North Holland.
Hausman, Jerry A. (ed.). 1993. *Contingent Valuation: A Critical Assessment.* Amsterdam: North Holland.
Hausman, Jerry A. 2012. "Contingent Valuation: From Dubious to Hopeless." *Journal of Economic Perspectives* 26(4): 43–56.
Heberlein, Thomas A., Mathew Wilson, Richard C. Bishop, and Nora C. Schaeffer. 2005. "Rethinking the Scope Test as a Criterion for Validity in Contingent Valuation." *Journal of Environmental Economics and Management* 50(1): 1–22.
Herriges, Joseph A., and Jason F. Shogren. 1996. "Starting Point Bias in Dichotomous Choice Valuation With Follow-Up Questioning." *Journal of Environmental Economics and Management* 30(1): 112–131.
Herriges, Joseph A., Catherine L. Kling, Chih-Chen Liu, and Justin Tobias. 2010. "What Are the Consequences of Consequentiality?" *Journal of Environmental Economics and Management* 59(1): 67– 81.
Holmes, Thomas, and Wictor Adamowicz. 2003. "Attribute-Based Methods." In *A Primer on Non-market Valuation*, edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Boston: Kluwer Academic.
Horowitz, John K., and Kenneth E. McConnell. 2002. "A Review of WTA/WTP Studies." *Journal of Environmental Economics and Management* 44(3): 426–447.
Horowitz, John K., and Kenneth E. McConnell. 2003. "Willingness to Accept, Willingness to Pay and the Income Effect." *Journal of Economic Behavior & Organization* 51(4): 537–545.
Jacquemet, Nicolas, Robert-Vincent Joule, Stephane Luchini, Jason F. Shogren. 2013. "Preference Elicitation Under Oath." *Journal of Environmental Economics and Management* 65(1): 110–132.
Jeon, Yongsik, and Joseph A. Herriges. 2010. "Convergent Validity of Contingent Behavior Responses in Models of Recreation Demand." *Environmental and Resource Economics* 45(2): 223–250.

Kahneman, Daniel, and Jack L. Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22(1): 57–70.

Kling, Catherine L., Daniel J. Phaneuf, and Jinhua Zhao. 2012. "From Exxon to BP: Has Some Number Become Better than No Number?" *Journal of Economic Perspectives* 26(4): 3–26.

Knetsch, Jack L., and Robert K. Davis. 1966. "Comparisons of Methods for Recreation Evaluation." In *Water Research*, edited by Allen V. Kneese and Stephen C. Smith. Baltimore: Johns Hopkins University Press for Resources for the Future.

Kopp, Raymond J., and V. Kerry Smith. 1997. "Constructing Measures of Economic Value." In *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, edited by Raymond J. Kopp, Werner W. Pommerehne, and Norbert Schwarz. Boston: Kluwer Academic.

Kopp, Raymond J., Werner W. Pommerehne, and Norbert Schwarz (eds). 1997. *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*. Boston: Kluwer Academic.

Kriström, Bergt. 1990. "A Nonparametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies." *Land Economics* 66(2): 135–139.

Landry, Craig F., and John A. List. 2007. "Using Ex Ante Approaches to Obtain Credible Signals for Value in Contingent Markets: Evidence from the Field." *American Journal of Agricultural Economics* 89(2): 420–429.

Lareau, Thomas J., and Douglas A. Rae. 1989. "Valuing WTP for Diesel Odor Reductions: An Application of Contingent Ranking Technique." *Southern Economic Journal* 55(3): 728–742.

Lew, Daniel K. and Kristin Wallmo. 2011. "External Tests of Scope and Embedding in Stated Preference Choice Experiments: An Application to Endangered Species Valuation." *Environmental and Resource Economics* 48(1): 1–23.

List, John A. 2001. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *American Economic Review* 91(5): 1498–1507.

List, John A., and Craig A. Gallet. 2001. "What Experimental Protocols Influence Disparities Between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics* 20(3): 241–254.

List, John A., and Jason F. Shogren. 1997. "Calibration of Willingness-to-Accept." *Journal of Environmental Economics and Management* 43(2): 219–233.

Louviere, Jordan J., David A. Hensher, and Joffre D. Swait. 2000. *Stated Choice Methods: Analysis and Applications*. Cambridge: Cambridge University Press.

McConnell, Kenneth E. 1986. *The Damages to Recreational Activities for PCBs in New Bedford Harbor*. Cambridge: Industrial Economics.

McConnell, Kenneth E. 1990. "Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation." *Journal of Environmental Economics and Management* 18(1): 19–34.

McFadden, Daniel L. 1994. "Contingent Valuation and Social Choice." *American Journal of Agricultural Economics* 76(4): 689–708.

Mitchell, Robert Cameron, and Richard T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, D.C.: Resources for the Future.

Morrison, Mark, and Thomas C. Brown. 2009. "Testing the Effectiveness of Certainty Scales, Cheap Talk, and Dissonance-Minimization in Reducing Hypothetical Bias in Contingent Valuation Studies." *Environmental and Resource Economics* 44(3): 307–326.

Morrison, Mark, Jeff Bennett, and Russell Blamey. 1999. "Valuing Improved Wetland Quality Using Choice Modeling." *Water Resources Research* 35(9): 2805–2914.

Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, and Darryl Weatherheard. 2005. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental and Resource Economics* 30(3): 313–325.

Neill, Helen R., Ronald G. Cummings, Philip T. Ganderton, Glen W. Harrison, and Thomas McGuckin. 1994. "Hypothetical Surveys and Real Economic Commitments." *Land Economics* 70(2): 145–154.

Ojea, Elena, and Maria L. Loureiro. 2011. "Identifying the Scope Effect on a Meta-Analysis of Biodiversity Valuation Studies." *Resource and Energy Economics* 33(3): 706–724.

Poe, Greg L., Jeremy E. Clark, Daniel Rondeau, and William. D. Schulze. 2002. "Provision Point Mechanisms and Field Validity Tests of Contingent Valuation." *Environmental and Resource Economics* 23(1): 105–131.

Portney, Paul R. 1994. "The Contingent Valuation Debate: Why Economists Should Care." *Journal of Economic Perspectives* 8(4): 3–17.

Powe, Neil A., and Ian J. Bateman. 2004. "Investigating Insensitivity to Scope: A Split-Sample Test of Perceived Scheme Realism." *Land Economics* 80(2): 258–271.

Rae, Douglas A. 1983. "The Value to Visitors of Improving Visibility at Mesa Verde and Great Smoky Mountain National Parks." In *Managing Air Quality and Scenic Resources and National Parks and Wilderness Areas*, edited by Robert D. Rowe and Lauraine G. Chestnut. Boulder: Westview Press.

Roe, Brian, Kevin J. Boyle, and Mario Teisl. 1996. "Using Conjoint Analysis to Derive Estimates of Compensating Variation." *Journal of Environmental Economics and Management* 31(2): 145–159.

Rowe, Robert D., William D. Schulze, and William S. Breffle. 1996. "A Test for Payment Card Biases." *Journal of Environmental Economics and Management* 31(2): 178–185.

Sellar, Christine, Jean-Paul Chavas, and John R. Stoll. 1986. "Specification of the Logit Model: The Case of Valuation of Nonmarket Goods." *Journal of Environmental Economics and Management* 13(4): 382–390.

Shogren, Jason, S.Y. Shin, Dermot J. Hayes, and James R. Kliebenstein. 1994. "Resolving Differences in Willingness to Pay and Willingness to Accept." *American Economic Review* 84(1): 255–270.

Smith, V. Kerry, and William H. Desvousges. 1986. *Measuring Water Quality Benefits*. Norwell: Kluwer-Nijhoff.

Smith, V. Kerry, and Laura L. Osborne. 1996. "Do Contingent Valuation Estimates Pass a 'Scope' Test? A Meta-analysis." *Journal of Environmental Economics and Management* 31(3): 287–301.

Sugden, Robert. 1999. "Alternatives to the Neoclassical Theory of Choice." In *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, edited by Ian J. Bateman and Kevin G. Willis. Oxford: Oxford University Press.

Sugden, Robert. 2004. "The Opportunity Criterion: Consumer Sovereignty without the Assumption of Coherent Preferences." *American Economic Review* 94(4): 1014–1033.

Turnbull, B. 1976. "The Empirical Distribution Function With Arbitrarily Grouped, Censored and Truncated Data." *Journal of the Royal Statistical Society B* 38: 290–295.

Vatn, Arild. 2004. "Environmental Valuation and Rationality." *Land Economics* 80(1): 1–18.

von Haefen, Roger, and Daniel Phaneuf. 2008. "Identifying Demand Parameters in the Presence of Unobservables: A Combined Revealed and Stated Preference Approach." *Journal of Environmental Economics and Management* 56(1): 19–32.

Vossler, Christian A., and Mary Evans. 2009. "Bridging the Gap Between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality." *Journal of Environmental Economics and Management* 58(3): 338–345.

Vossler, Christian A., and Gregory L. Poe. 2011. "Consequentiality and Contingent Values: An Emerging Paradigm." In *The International Handbook on Non-Market Environmental Valuation,* ch. 7, edited by Jeff Bennett. North Hampton: Edward Elgar.

Vossler, Christian, Maurice Doyon, Daniel Rondeau, and Fréderic Roy-Vigneault. 2011. "Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments." Working paper.

Welsh, Michael P., and Gregory L. Poe. 1998. "Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach." *Journal of Environmental Economics and Management* 36(2): 170–185.

Whitehead, John C. 2002. "Incentive Incompatibility and Starting-Point Bias in Iterative Valuation Questions." *Land Economics* 78(2): 285–297.

Whittington, Dale. 2002. "Improving the Performance of Contingent Valuation Studies in Developing Countries." *Environmental and Resource Economics* 22(1–2): 323–367.

Zhao, Jinhua, and Catherine L. Kling. 2004. "Willingness To Pay, Compensating Variation, and the Cost of Commitment." *Economic Inquiry* 42(3): 503–517.

Chapter 13

Additional Topics

This chapter covers four additional topics that do not fit easily into any of the previous chapters. The first is the use of values borrowed from studies of values in other settings to estimate the value of the environmental change of interest. This practice has come to be known as "benefits transfer." The second is the use of data from both revealed preference and stated preference analyses to estimate a single valuation model. The third is the estimation of the values of services provided by ecosystems. The fourth is a brief discussion of some of the implications of behavioral economics for the task of estimating nonmarket values for environmental goods and services.

Benefits Transfer

Benefits transfer refers to the practice of applying nonmarket values obtained from primary studies of resource or environmental changes undertaken elsewhere to the evaluation of a proposed or observed change that is of interest to the analyst. Examples include:

- using values per day for recreational angling at one lake obtained from a travel cost demand study to value an increase in the same activity at another lake;
- using the willingness to pay to avoid an asthma attack obtained from an averting behavior study to value the prevention of a day of respiratory symptoms; and
- using the willingness to pay to preserve a square mile of tropical rain forest in one country estimated from a stated preference study to value the preservation of a square mile of rain forest in another country.

The practice of benefits transfer became common in the economic analysis of environmental regulations in the United States in the mid-1980s, even before any systematic development of either terminology or procedures and protocols and certainly before any rigorous testing of the validity of the practice. All of this changed in 1992 with the publication of a set of nine papers on benefits transfer in a special issue of *Water Resources Research*. See Brookshire and Neill (1992) for an introduction and overview of these papers.

It is common to refer to the environmental policy being evaluated as the "policy site" and the source of the values being used as the "study site," even though many of the policies for which benefits transfer are used are not site specific. In principle, the values at the policy site can be different from those of the study site for two sets of reasons: differences in the characteristics of the two environmental features being valued ("supply side" factors) and differences between the populations making use of, or at least valuing, the resource change ("demand side" factors). The latter can include differences in income, tastes, and preferences, and other relevant socioeconomic characteristics. Thus, we must consider the question of how values can be adjusted in the transfer process to reflect these two types of differences. To examine this question, we look at the types of procedures available for benefits transfer from the simplest to the most sophisticated. There are several useful sources for more detailed information on how to design and implement a benefits transfer. One (Rosenberger and Loomis 2003) is a fairly basic "how to do it" that walks the reader through the steps involved, using three case studies. Likewise, Loomis and Richardson (2008) provides guidance on the process in a manual that accompanies an on-line toolkit (Loomis et al. 2007). Desvousges, Johnson, and Banzhaf (1998) provide a more comprehensive and rigorous treatment that includes discussions of several econometric issues in analyzing and summarizing study site data.

The simplest procedure is to select the study site that in the analyst's judgment is most similar to the policy site in terms of both demand-side and supply-side characteristics. If there is more than one suitable study site (or more than one estimated value for a given study site), one could use a range, calculate a mean or median value for the distribution, or perhaps subjectively weight the values based on judgments of their quality before taking the mean. If the study site and policy site values are for different years, the study site value should be adjusted for changes in prices using a suitable price index. However, adjusting for inflation in this way involves the assumption that the budget shares of expenditures of both the study site and policy site populations are similar to the weights used in constructing the price index (Eiswerth and Shaw 1997). If there is information available on the income levels of the two populations, then it would also be appropriate to adjust the study site WTP using an estimate of the income elasticity of WTP. The relevant elasticity, however, is not the income elasticity of demand for the good. As Flores and Carson (1997) showed, the relationship between the two elasticity concepts is complex.

This simple procedure does not allow for more systematic adjustments of study site values to account for differences in site characteristics and populations. To make such adjustments, it is necessary to have information on how values are related to the relevant characteristics. Some information of this sort is available if the study site data includes a valuation function or WTP function—that is, a statistical relationship between the WTPs of individuals in the sample and their socioeconomic characteristics. The adjusted policy site value can be calculated by plugging mean values for the policy site population into the valuation function.

A more sophisticated approach is to apply meta-analytical techniques to a larger sample of study site data that represents a variety of population and site characteristics. The meta-analysis equation can also include variables that reflect differences in methods, such as revealed preference versus various forms of stated preference methods. Examples include Smith and Kaoru (1990) for recreation and Mrozek and Taylor (2002) for the value of statistical life. For a more detailed analysis of meta-analysis for benefits transfer, see Bergstrom and Taylor (2006). Desvousges, Johnson, and Banzhaf (1998) and Rosenberger and Loomis (2003) contain references to additional examples.

More recently, in a series of papers, Klaus Moeltner has described several studies showing how to utilize meta-regression techniques and Bayesian models to deal with problems such as unobserved heterogeneity and small sample size in the underlying valuation studies. See Moeltner et al. (2007, 2009), Moeltner and Rosenberger (2008), and Moeltner and Woodward (2009). Phaneuf and van Houtven (forthcoming) recently added to this line of research.

Finally, Smith and others developed and applied a method, based on data from several studies using different valuation methods, to "calibrate" the parameters of an assumed underlying preference function over site characteristics. They then used the calibrated preference function to calculate values for a variety of policy site proposals. See Smith, van Houtven, and Pattanayak (2002), Smith, Pattanayak, and van Houtven (2006), and Pattanayak, Smith, and van Houtven (2007).

An important question concerns the validity of using benefit transfer values as a substitute for doing original valuation research for the policy being evaluated. Any assessment of the validity of benefits transfer requires a yardstick. The yardstick used by those who have investigated this question is an estimate of value obtained by stated preference or revealed preference methods at the policy site, so these are tests of convergent validity, not criterion validity (see Chapter 12).

Rosenberger and Loomis (2003) and Shrestha and Loomis (2001) both reviewed studies of validity (see papers for references). Also, Shrestha and Loomis (2001) performed a validity test of benefits transfer of outdoor recreation activities across countries based on a meta-analysis of more than 600 recreation values from U.S. studies. Overall, the results are mixed, but one finding stands out: convergent validity is generally higher when the transfer is done using either a valuation function or meta-analysis. This confirms the importance of using systematic methods to adjust study site values for differences in population and site characteristics. Nevertheless, significant differences between transfer values and original values often occur. See Lindhjem and Navrud (2008) for a review of the more recent literature on validity testing and their own test of convergent validity of benefits transfer involving non-timber values in Norway, Sweden, and Finland. They found substantial transfer errors, and accordingly, they were skeptical about the use of meta-analysis for international benefits transfers.

In recognition of the popularity of using benefits transfer and the potential savings in analytical costs that are possible, several organizations have created and made available on the Internet large databases of valuation studies that can be used in benefits transfer. They provide not only the estimated values from each study but also other data on study methodology in a standardized format so that studies can be evaluated for their suitability for the transfer exercise. The largest and most detailed of these is the Environmental Valuation Reference Inventory (EVRI), created and maintained by the Canadian federal environmental agency, Environment Canada. It is a searchable storehouse of empirical studies on the economic value of environmental benefits and human health effects. Access to EVRI is free for Canadian residents, and subscription information is available at the EVRI website (www.evri.ca/Global/Splash.aspx).

There are also more specialized databases. For example the University of California has created a database of economic values for beneficial uses of water (http://buvd.ucdavis.edu/); the U.S. Fish and Wildlife Service maintains a sports fishing value database (www.indecon.com/fish/); and Texas A&M has compiled a database for valuing ecosystem services relevant to the Gulf of Mexico.

Combining Revealed Preference and Stated Preference Data

It seems natural to think of revealed preference (RP) and stated preference (SP) methods as substitutes or alternative ways of estimating values for a given change in environmental or resource conditions. Thought of in these terms, RP methods have served as a natural comparison point for convergent validity tests of SP as described in Chapter 12 (see Whitehead et al. 2010 for a current example). However, several authors have shown that it is possible to use RP and SP methods as complements by combining RP and SP data to estimate a single valuation model. These combined models embody the hypothesis that the RP and SP data come from the same structure of preferences or utility function. Testing this hypothesis can be viewed as a kind of convergent validity test.

In the first published study using this approach, Cameron (1992) specified a travel cost demand model to explain the numbers of fishing trips taken in a year by individuals in her sample of Texas anglers. She also specified a stochastic utility difference model to explain the same individuals' responses to a contingent behavior question of the form "If the travel cost of your trips were \$*X* more, would you stop visiting these sites altogether?" The functional forms of the two equations were derived from an assumed underlying quadratic utility function. She constrained the common parameters of the two equations to be equal, estimating them using a simultaneous equation technique. She then used the estimated utility parameters to generate welfare measures for various changes in the conditions of access to the sites to illustrate the method.

Adamowicz, Louviere, and Williams (1994) designed a survey explicitly to generate both RP and SP data of recreation choices. Their RP data were used to estimate a random utility model of recreation site choices. This was then combined with responses to a set of SP questions about site choice based on the same set of site attributes. The SP questions were designed to eliminate collinearity among attributes. They estimated the random utility model and SP models separately and then combined the two data sets for a joint estimation. Their results supported the hypothesis that both sets of data came from the same underlying preference structures. However, von Haefen and Phaneuf (2008) have since re-examined these data using more recent econometric techniques and found inconsistencies in preferences implied by the RP and SP data sources.

McConnell, Weninger, and Strand (1999) also combined RP and SP data on recreation activities to estimate willingness to pay for a recreation trip. Their method allowed differences in the preferences generating the RP and SP data because the SP responses came after individuals had visited the sites in question. However, they could not reject the hypothesis that the parameters in the two models were equal. More recently Earnhart (2001, 2002) combined SP data with hedonic housing discrete choice data to examine housing choices and to estimate values for environmental amenities linked to residential locations.

As Kling (1997) showed through simulation studies, combining SP and RP data can reduce bias and improve the precision of welfare estimates, assuming of course that the data generating processes (including preferences) are the same for both data sources. Also, the design of the choice sets for the SP part of the study presents an opportunity for examining portions of individuals' response surfaces or preference structures for which there is no observed behavior.

A review of the literature by Whitehead et al. (2008) showed there has been extensive use of combinations of various types of RP and SP data in the fields of marketing (predicting demands for new or changed products) and transportation policy. This review also covers the types of models and econometric techniques that can be used to combine RP and SP data and discusses ways in which combined RP and SP data can help to overcome some of the problems and limitations of relying on either RP or SP data alone. Combining RP and SP has become increasingly popular with recent applications to wetlands (Grossmann 2011), professional sports (Whitehead et al. 2013) and food safety (Morgan et al. 2013).

Valuing the Services of Ecosystems

An ecosystem can be defined as "an assemblage of organisms interacting with its associated physical environment in a specific place" (Barbier 2011, 28, and references therein). It is a human construct whose size and scope depend on how an investigator draws the spatially explicit boundaries around the objects of interest. The unit of analysis can be as small as the gut of an insect and as large as the delta area of a major river system such as the Amazon or Mississippi. Each ecosystem is characterized by its spatial dimensions, its species composition, the functions or processes that it carries out, and the services that it provides to people. Ecosystem functions or processes are the normal characteristic actions or activities of the system that are necessary for its self-maintenance (Whigham 1997, 231). Ecosystem services have been defined in various ways in the literature.

EPA has defined them as "those ecological functions or processes that directly or indirectly contribute to human well-being or have the potential to do so in the future" (U.S. EPA 2004, 4). Ecologist Gretchen Daily defined them as "the conditions and processes through which natural ecosystems sustain and fulfill human life. They maintain biodiversity and the production of *ecosystem goods*, such as seafood, forage, timber … and their precursors … In addition to the production of goods, ecosystem services are the actual life-support functions, such as cleansing, recycling, and renewal, and they confer many intangible aesthetic and cultural benefits as well" (Daily 1997, 3).

A number of authors have provided lists of ecosystem services. Gretchen Daily's (1997, 3–4) list is representative and includes:

- purification of air and water;
- mitigation of floods and droughts by regulation of hydrological cycles;
- detoxification and decomposition of wastes;
- generation and renewal of soils;
- pollination of crops and natural vegetation;
- control of agricultural and other pests;
- nutrient recycling;
- partial climate stabilization; and
- providing aesthetic beauty and intellectual stimulation.

Other authors (for example, Costanza et al. 1997) have added such things as erosion control, habitat or refugia for species and preservation of biodiversity, and production of food and raw materials. Some of these services affect humans directly as in the cases of food and raw material production, and these direct services can be either market or nonmarket services. Other services affect people only indirectly as in the cases of nutrient recycling and the regeneration of soil. In these cases, it is necessary to establish the link between the ecosystem service being valued and the channel through which it affects people. More will be said on this point later.

The Millennium Ecosystem Assessment (2003, 8) added to this conceptual framework by identifying four categories of service streams: provisioning, regulating, supporting, and cultural. Provisioning services are the flows of goods such as food, fiber, fuels, and so forth that stem from the primary and secondary productivity of ecological systems. These service streams are usually easily identified and the most apt to be in the form of private goods and to be governed by market transactions. Regulating services are the benefits people obtain from the regulation of ecosystem processes such as the maintenance of air quality and climate regulation. Supporting services are those that are necessary for all other ecosystem services. Supporting services provide the tools (for example, oxygen and nutrients), while regulating services do the work (atmospheric regulation). Supporting services are sometimes called "ecosystem functions." Finally, cultural services provide benefits through recreation and other highly subjective activities like spiritual enrichment and aesthetic experience.

A careful inspection of this list and these definitions shows a strong possibility of double counting of service flows, especially between some of the supporting services and the provisioning services. For example, waste decomposition is the source of the nutrients that support the bottom of the food chain and whose value is eventually embodied in the value of the plant and animal species that are harvested for human consumption. Similarly, the value of pollination services is eventually embodied in the value of harvested plant crops. In recognition of this potential for double counting, Boyd and Banzhaf offered a more limited definition:

> Ecosystem services are components of nature, directly enjoyed, consumed, or used to yield human well-being.
>
> (Boyd and Banzhaf 2007, 619)

For example, while many economists recognize nature-based outdoor recreation as an ecosystem service, Boyd and Banzhaf pointed out that what we count as recreation (a period of time spent at a site engaged in some activity) involves the combining of human inputs (time and perhaps equipment) with inputs from the ecosystem—for example, populations of flora and fauna that add value to the experience. Not all economists accept this narrow approach to defining ecosystem services. See, for example, Polasky and Segerson (2009, 412) for further discussion.

This emphasis on final *versus* intermediate services also calls attention to another issue to be addressed in the identification and valuation of ecosystem services. Most of the provisioning and cultural services identified by the Millennium Ecosystem Assessment are actually produced by combining ecological outputs with human labor and capital. The costs of these nonecological inputs must be deducted from the value of the food, fiber, fuels, and recreation in determining the value of the ecosystem service itself. This deduction is necessary in arriving at the *value added* by the ecosystem service. Thus, the value of the ecosystem service is essentially the net economic rent attributable to this dimension of nature; and those economists estimating the values of fish and forest harvests have been engaged in valuing one type of ecosystem service.

Valuing Service Flows

Given the anthropocentric definition of ecosystem services, the economic concepts of willingness to pay and willingness to accept compensation provide the conceptual basis for defining the economic values of these services; and the methods and models described in this book provide the means for estimating these values, at least in principle. Estimating the economic value of an ecosystem service involves three steps (see for example, Barbier 2011, 33). The first is determining the nature and size of the environmental change affecting ecosystem structure and function. The change could be in the spatial area of a particular type of habitat, for example, freshwater wetlands, in the populations of species present, or in the fluxes of energy or nutrients through the system. The second step involves determining how these changes affect the quantities and qualities of ecosystem service flows to people. The third step involves using existing economic methods where available to assess the changes in people's well-being, as measured in dollars.

When an ecosystem service supports the production of a marketed commodity, the value of a change in that service is the sum of the changes in consumers' and producers' surpluses in that market. For example, an increase in the population of a pollinating insect could increase the output of agricultural crops, resulting in lower prices to consumers and/or greater quasi-rents to producers. For example, Ricketts et al. (2004) examined the economic value of the pollination services provided by feral bees to coffee planters in Costa Rica. They measured the variation in yield in areas adjacent to, proximate to, and distant from patches of forest that provided habitat for pollinating bees. The increased yield valued at the price received by planters provided an estimate of the added producers' surplus generated by the free pollination service of the bees.

Tidal wetlands are known to shelter the young of commercially valuable fish species. Changes in the area of wetlands have been related to changes in commercial harvest of blue crabs and finfish (Lynne, Conroy, and Prochaska 1981; Bell 1989). Barbier, Strand, and Sathirathai (2002) found a similar relationship between areas of mangroves and fisheries harvests in Thailand. The methods described in Chapter 8 are available for this type of ecosystem service.

When ecosystem functions support nonmarket environmental services, we may be able to draw on the tool kit of nonmarket valuation methods to determine the economic values of changes in these service flows. For example, when a change in an ecosystem service results in an improvement in the quality of outdoor recreational experiences, recreational demand models can be used to estimate the value of the service flow (see for example, Adamowicz et al. 2011), and stated preference methods might be used to value aesthetic services.

In order to estimate the economic value of a basic ecosystem function, we need to know the link between that function and the ecosystem service flows that it supports, which will not always be easy to uncover. One approach to establishing this link is to think of the relevant components of the ecosystem as being involved in a production process. For discussions of the production function approach, see Barbier (1994, 2000, and 2011, 53–54). Under this approach, the ecosystem is assumed to be an equilibrium system that can be subjected to comparative static analysis to determine changes in service flows in response to changes in ecosystem conditions. Crocker and Tschirhart (1992) developed a formal model of an ecosystem and economic system to demonstrate this approach. More recently, researchers have been developing empirically based integrated assessment models of spatially explicit multiple ecosystem services to investigate the effects of land use changes and policy interventions on the flows and values of ecosystem services. For examples, see Kareiva et al. (2011).

One complication in developing such models is that in terms of production theory ecosystems are multiproduct production systems in which jointness in production is likely to be a dominant feature. For example, a species of bird might be valued both for its pollination of a commercial fruit species and for its control of insects that damage some other commercially valuable plant. The value of the bird species is the sum of the values of all of its services. However, the jointness in production must be taken into account when estimating the values of individual service flows (see Chapter 8). Complex spatially explicit models are being developed to generate predictions of the supplies of a variety of ecosystem services in response to changes in such things as land use management plans. For an example, see Polasky et al. (2011).

Another complication is that the responses of ecosystems to perturbations might display nonlinearities, discontinuities, multiple end points, and even chaotic behavior, especially for changes in the populations of species and fluxes of energy or nutrients (Levin and Pacala 2003; Dasgupta and Mäler 2004). In fact some aspects of ecosystem behavior might be fundamentally unpredictable (Huisman and Weissing 2001). See also Botkin (2012) who challenged the whole notion of stability and equilibrium of ecosystems generally. For these reasons, economists may have more success in estimating the values of changes in the spatial extent of an ecosystem than changes in other characteristics of the system.

A recent and ambitious example of incorporating the value of ecosystem service flows into land use alternatives demonstrates both the complexity of doing so and the potential policy importance that such an effort may produce. Bateman et al. (2013) used a set of spatially explicit biophysical models for the United Kingdom in conjunction with nonmarket values estimated for recreation, open space, carbon emissions and sequestration, and wildlife diversity to evaluate economically optimal land use. Their findings make clear that using market prices alone can lead to suboptimal decision making and that the tools to evaluate a wide range of ecosystem services at a relatively fine spatial scale are available.

It is sometimes suggested that the cost of replacing a function of an ecological system with a human engineered system can be used as a measure of the economic value of the function itself. Replacement cost can be a valid measure of economic value only if three conditions are met: the human-engineered system must provide services of equivalent quality and magnitude, the human-engineered system must be the least costly alternative, and individuals in aggregate must, in fact, be willing to incur these costs if the natural service were not available (Shabman and Batie 1978). Note that when these conditions are not met, there is no presumption that replacement cost is either an overestimate or an underestimate of true economic value—all that can be said is that the two numbers are measures of different things.

In a classic example of the replacement cost approach, Gosselink, Odum, and Pope (1974) used an estimate of the cost of a tertiary sewage treatment as the economic value of the nutrient removal function of a tidal wetland. In another widely cited article, Chichilnisky and Heal (1998) used this method to estimate the value of water purification service provided by a large protected watershed area in the Catskill Mountains of New York to be $6–8 billion. Chichilnisky and Heal argued that in the Catskill Mountain case, these conditions were met, since the City of New York would have been forced to spend $6–8 billion building a water filtration plant for the city water supply if it did not invest in watershed protection. However, the philosopher Mark Sagoff (2005) convincingly argued that the record does not support the assumption that the $6–8 billion filtration plant was actually required.

Valuing the Wealth of Nature

The value of that part of the natural world that contributes to human well-being or welfare through the provision of ecological goods and services can be called the wealth of nature, or the value of natural capital (Costanza et al. 1997). A large-scale collaborative effort to operationalize the concept of natural capital is the Natural Capital Project at Stanford University (Kareiva et al. 2011, 37).[1] There are two ways to think about how to measure the value of the stock of natural capital. The first is to think of the value of this stock as the discounted present value of the economic values of all of the streams of service flows from all of the ecosystems of the world. The problem is how to measure the total value of each service flow—that is, the value of moving from a zero flow to the current set of flows.

The second way to think about the value of the stock of natural capital is borrowed from the way the Bureau of Economic Analysis in the U.S. Department of Commerce develops estimates of the tangible wealth of the U.S. economy (for example, Katz and Herman 1997). These estimates are based, in principle, on counts of the numbers of each type of productive asset and a set of unit values or "prices" for each asset that are assumed to be constant. The measure of wealth is the summation of the price times quantity calculations for each asset type. It has long been understood that the result of the price times quantity calculation does not represent the total value of the stock of wealth any more than gross domestic product (also a price times quantity measure) represents the value that people place on the nation's output. Both calculations use marginal values or unit prices and have no way to capture the value of inframarginal units.

Behavioral Economics

The term behavioral economics refers to the study of behavior that is inconsistent with the standard economic model of rational choice with stable and purely self-interested preferences, and to efforts to explain these anomalies with findings from other fields, especially psychology and sociology. For an overview of this

1 Its web page states, "Our mission is to align economic forces with conservation by mainstreaming natural capital into decisions." See: http://naturalcapitalproject.org/home04.html.

relatively new field, see Camerer, Loewenstein, and Rabin (2004). Shogren and Taylor (2008) provided a cogent discussion of the ways in which behavioral economics ideas relate to environmental economics with a particular emphasis on the divergence between willingness to pay and accept.

This section raises the question of the possible implications of behavioral economics for the methods for estimating environmental and resource values that are discussed in this book. Briefly, the methods for estimating individuals' values from observations of revealed or stated choices are all based on models of behavior and choice that start with the assumptions of stable preferences and rational behavior. However, if choices are not consistent with preferences and do not reflect increases in individuals' well-being, this could call into question the values obtained from using these methods.

Examples of the kinds of behavioral anomalies that could affect estimates of values include the following: the endowment effect leading to large differences between WTP and WTA (see Chapter 3) and hyperbolic discounting (see Chapter 6). In addition to these, Shogren and Taylor add coherent arbitrariness (the anchoring of values on arbitrary levels), problems in self-control, and altruistic behavior to the list of deviations from rational choice theory that may have important implications for nonmarket valuation. Another way to categorize the issues is suggested by Kling, Phaneuf, and Zhao (2012) who noted that there are two types of findings from behavioral economics that are relevant to nonmarket valuation: (a) preferences that deviate from neo-classical preferences, and (b) failure of individuals to optimize. The endowment effect, hyperbolic discounting, framing effects, and the issues identified by Shogren and Taylor above largely fit within the first category. An example of a behavioral anomaly consistent with the second is the observation that consumers sometimes appear to hold "mental accounts" where they arbitrarily constrain their spending on goods to stay within predetermined bounds.

In response to the observed anomalies, some analysts have suggested that policy analysts should reject the premise of revealed preference and substitute expert judgments about what will really enhance individuals' well-being when conducting policy evaluations. For example, Bernheim and Rangel (2007, 9) argued that "if one knows enough about the nature of decision making malfunctions, it may be possible to recover tastes by relying on a selective application of the revealed preference principle." Kerry Smith (2007, 154), after quoting from Bernheim and Rangel, countered with, "Unfortunately we cannot resolve the problems that these authors suggest arise when people do not make what they (the authors) think is the appropriate choice. There is one small detail. Who decides what is a *suitably defined* welfare improvement—elected officials, policymakers, experts?" He goes on to say, "The skeptic should ask who decides what choices are mistakes and when enough are made so that the evidence compels a conclusion that preferences in these dimensions are incoherent" (2007, 156). There is now a growing literature dealing with these differing perspectives. In addition to Bernheim and Rangel and Smith, see, for example, Robinson and Hammitt (2011), Smith and Moore (2010), and Hammitt (2013), as well as references therein. Some authors point to the fact that observed behavioral anomalies often appear to depend on the context in which the choices are observed and the constraints that people face (for example, Smith and Moore, 2010); and this raises questions about their significance for environmental valuation.

There is no doubt that research in behavioral economics has raised important questions for nonmarket valuation, and welfare economics in general, and students of nonmarket valuation should remain engaged in this fascinating literature. However, as Shogren and Taylor point out "the evidence from behavioral economics remains insufficient to support the wholesale rejection of rational choice theory" (2008, 41), and until there are compelling alternatives, benefit-cost analysis will continue to rely on its neoclassical roots (see the work of Bernheim and Rangel 2009, for one avenue). As the literature continues to develop, welfare analysts should recognize that in addition to challenges, there are likely to be many useful insights from behavioral economics findings that can strengthen the methods used to elicit preference information (from stated preference surveys for example) and for the interpretation of our data (comparisons between revealed and stated preference information for example).

References

Adamowicz, W., J. Louviere, and M. Williams. 1994. "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities." *Journal of Environmental Economics and Management* 26(3): 271–292.

Adamowicz, W.L. (Vic), Robin Naidoo, Erik Nelson, Stephen Polasky, and Jing Zhang. 2011. "Nature-Based Tourism and Recreation." In *Natural Capital: Theory and Practice of Mapping Ecosystem Services*, edited by Peter Kareiva, Heather Tallis, Taylor H. Ricketts, Gretchen C. Daily, and Stephen Polasky. Oxford: Oxford University Press.

Barbier, Edward B. 1994. "Valuing Environmental Functions: Tropical Wetlands." *Land Economics* 70(2): 155–173.

Barbier, Edward B. 2000. "Valuing the Environment as Input: Review of Applications to Mangrove–Fishery Linkages." *Ecological Economics* 35(1): 47–61.

Barbier, Edward B. 2011. *Capitalizing on Nature: Ecosystems as Natural Assets*. Cambridge: Cambridge University Press.

Barbier, Edward B., Ivar Strand, and Suthawan Sathirathai. 2002. "Do Open Access Conditions Affect the Valuation of an Externality? Estimating the Welfare Effects of Mangrove–Fishery Linkages in Thailand." *Environmental and Resource Economics* 21(4): 343–367.

Bateman, Ian J., Amii R. Harwood, Georgina M. Mace, Robert T. Watson, David J. Abson, Barnaby Andrews, Amy Binner, Andrew Crow, Brett H. Day, Steve Dugdale, Carlo Fezzi, Jo Foden, David Hadley, Roy Haines-Young, Mark Hulme, Andrea Kontoleon, Andrew A. Lovett, Paul Munday, Unai Pascual, James Paterson, Grisca Perino, Antara Setn, Gavin Siriwardena, Daan van Soest, and Mette Termansen. 2013. "Bringing Ecosystem Services into Economic Decision-Making: Land Use in the United Kingdom" *Science* 341(611): 45–50.

Bell, Frederick W. 1989. *Application of Wetland Evaluation Theory to Florida Fisheries*. Tallahassee: The Florida Sea Grant College.

Bergstrom, John C., and Laura O. Taylor. 2006. "Using Meta-Analysis for Benefits Transfer: Theory and Practice." *Ecological Economics* 60(2): 351–360.

Bernheim, B. Douglas, and Antonio Rangel. 2007. "Behavioral Economics: Welfare and Policy Analysis with Nonstandard Decision-Makers." In *Behavioral Economics and Its Applications*, edited by Peter Diamond and Hannu Vartianen. Princeton: Princeton University Press.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics" *Quarterly Journal of Economics* 124(1): 51–104.

Botkin, Daniel B. 2012. *The Moon in the Nautilus Shell: Discordant Harmonies Reconsidered*. Oxford: Oxford University Press.

Boyd, James, and Spencer Banzhaf. 2007. "What are Ecosystem Services? The Need for Standardized Environmental Accounting Units." *Ecological Economics* 63(2–3): 616–626.

Brookshire, David S., and Helen R. Neill. 1992. "Benefits Transfer: Conceptual and Empirical Issues." *Water Resources Research* 28(3): 651–655.

Camerer, Colin F., George Loewenstein, and Matthew Rabin. 2004. *Advances in Behavioral Economics*. New York: Russell Sage Foundation; Princeton: Princeton University Press.

Cameron, Trudy A. 1992. "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods." *Land Economics* 68(3): 302–317.

Chichilnisky, Graciela, and Geoffrey Heal. 1998. "Economic Returns from the Biosphere." *Nature* 391(6668): 629–630.

Costanza, Robert, Ralph d'Arge, Rudolph de Groot, Stephen Farber, Monica Grasso, Bruce Hannon, Karin Limburg, Shahid Naeem, Robert V. O'Neill, Jose Paruelo, Robert G. Raskin, Paul Sutton, and Marjan van den Belt. 1997. "The Value of the World's Ecosystem Services and Natural Capital." *Nature* 387(6630): 253–260.

Crocker, Thomas D., and John Tschirhart. 1992. "Ecosystems, Externalities, and Economics." *Environmental and Resource Economics* 2(6): 551–567.

Daily, Gretchen C. 1997. "Introduction: What Are Ecosystem Services?" In *Nature's Services: Societal Dependence on Natural Ecosystems*, edited by Gretchen C. Daily. Washington, D.C.: Island Press.

Dasgupta, Partha, and Karl-Göran Mäler. 2004. *The Economics of Non-Convex Ecosystems*. Dordrecht: Kluwer Academic.

Desvousges, William H., F. Reed Johnson, and H. Spencer Banzhaf. 1998. *Environmental Policy Analysis with Limited Information: Principles and Applications of the Transfer Method*. Cheltenham: Edward Elgar.

Eiswerth, Mark E., and W. Douglass Shaw. 1997. "Adjusting Benefits Transfer Values for Inflation." *Water Resources Research* 33(10): 2381–2385.

Earnhart, Dietrich. 2001. "Combining Revealed and Stated Preference Methods to Value Environmental Amenities at Residential Locations." *Land Economics* 77(1): 12–29.

Earnhart, Dietrich. 2002. "Combining Revealed and Stated Data to Examine Housing Decisions Using Discrete Choice Analysis." *Journal of Urban Economics* 51(1): 143–169.

Flores, Nicholas E., and Richard T. Carson. 1997. "The Relationship between the Income Elasticities of Demand and Willingness to Pay." *Journal of Environmental Economics and Management* 33(3): 287–295.

Gosselink, James C., Eugene P. Odum, and R.M. Pope. 1974. *The Value of the Tidal Marsh*. Baton Rouge: Department of Marine Sciences, Louisiana State University.

Grossmann, Malte. 2011. "Impacts of Boating Trip Limitations on the Recreational Value of the Spreewald Wetland: A Pooled Revealed/Contingent Behaviour Application of the Travel Cost Method." *Journal of Environmental Planning and Management* 54(2): 211–226.

Hammitt, James K. 2013. "Positive versus Normative Justifications for Benefit-Cost Analysis: Implications for Interpretation and Policy." *Review of Environmental Economics and Policy* 7(2), 199–218.

Huisman, Jef, and Franz J. Weissing. 2001. "Fundamental Unpredictability in Multispecies Competition." *American Naturalist* 157(5): 488–494.

Kareiva, Peter, Heather Tallis, Taylor H. Ricketts, Gretchen C. Daily, and Stephen Polasky (eds). 2011. *Natural Capital: Theory and Practice of Mapping Ecosystem Services*. Oxford: Oxford University Press.

Katz, Arnold J., and Shelby W. Herman. 1997. "Improved Estimates of Fixed Reproducible Tangible Wealth, 1929–95." *Survey of Current Business* May: 69–92.

Kling, Catherine L. 1997. "The Gains from Combining Travel Cost and Contingent Valuation Data to Value Nonmarket Goods." *Land Economics* 73(3): 428–439.

Kling, Catherine L., Daniel J. Phaneuf, and Jinhua Zhao. 2012. "From Exxon to BP: Has Some Number Become Better than No Number?" *Journal of Economic Perspectives* 26(4): 3–26.

Levin, Simon A., and Stephen W. Pacala. 2003. "Ecosystem Dynamics." In *Handbook of Environmental Economics*, edited by Karl-Göran Mäler and Jeffery R. Vincent. Amsterdam: Elsevier.

Lindhjem, Henrik, and Stale Navrud. 2008. "How Reliable are Meta-Analyses for International Benefit Transfers?" *Ecological Economics* 66(2–3): 425–435.

Loomis, John, and Leslie Richardson. 2008. "Benefit Transfer and Visitor Use Estimating Models of Wildlife Recreation, Species and Habitats." National Council for Science and the Environment. 2006 Wildlife Habitat Policy Research Program, Defenders of Wildlife.

Loomis, John, Rim Kroeer, Leslie Richardson, and Frank Casey. 2007. "The Benefit Transfer and Use Estimating Model Toolkit." Available at: http://dare.agsci.colostate.edu/outreach/tools/.

Lynne, Gary D., Patricia Conroy, and Frederick J. Prochaska. 1981. "Economic Valuation of Marsh Areas for Marine Production Processes." *Journal of Environmental Economics and Management* 8(2): 175–186.

McConnell, Kenneth E., Quinn Weninger, and Ivar E. Strand. 1999. "Joint Estimation of Contingent Valuation and Truncated Recreational Demands." In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*, edited by Joseph A. Herriges and Catherine L. Kling. Cheltenham: Edward Elgar.

Millennium Ecosystem Assessment. 2003. *Ecosystems and Human Well-being: A Framework for Assessment.* Washington, D.C.: Island Press.

Moeltner, Klaus, and Randall S. Rosenberger. 2008. "Meta-Regression and Benefit Transfer: Data Space, Model Space, and the Quest for 'Optimal Scope.'" *Berkeley Electronic Journal of Economic Analysis and Policy* 8(1). Available at: www.bepress.com/bejeap/vol8/iss1/art31/

Moeltner, Klaus, and Richard Woodward. 2009. "Meta-Functional Benefit Transfer for Wetland Valuation: Making the Most of Small Samples." *Environmental and Resource Economics* 42(1): 89–109.

Moeltner, Klaus, Kevin J. Boyle, and Robert W. Paterson. 2007. "Meta-Analysis and Benefit-Transfer for Resource Valuation: Addressing Classical Challenges with Bayesian Modeling." *Journal of Environmental Economics and Management* 53(2): 250–269.

Moeltner, Klaus, Robert J. Johnston, Randall S. Rosenberger, and Joshua M. Duke. 2009. "Benefit Transfer from Multiple Contingent Experiments: A Flexible Two-Step Model Combining Individual Choice Data with Community Characteristics." *American Journal of Agricultural Economics* 91(5): 1335–1342.

Morgan, O. Ashton, John C. Whitehead, William L. Huth, Greg S. Martin, and Richard Sjolander. 2013. "A Split-Sample Revealed and Stated Preference Demand Model to Examine Homogenous Subgroup Consumer Behavior Responses to Information and Food Safety Technology Treatments." *Environmental & Resource Economics* 54(4): 593–611.

Mrozek, Janusz R., and Laura O. Taylor. 2002. "What Determines the Value of Life? A Meta-Analysis." *Journal of Policy Analysis and Management* 21(2): 269–286.

Pattanayak, Subhrendu K., V. Kerry Smith, and George L. Van Houtven. 2007. "Improving the Practice of Benefits Transfer: A Preference Calibration Approach." In *Environmental Value Transfer: Issues and Methods*, edited by Stale Navrud and Richard Ready. Dordrecht: Springer.

Phaneuf, Daniel J., and George Van Houtven. Forthcoming. "Structural Benefits Transfer Using Bayesian Econometrics." In *Benefit Transfer of Environmental and Resource Values: A Handbook for Researchers and Practitioners*, edited by R. Johnston and R. Rosenberger.

Polasky, Stephen, and Kathleen Segerson. 2009. "Integrating Ecology and Economics in the Study of Ecosystem Services: Some Lessons Learned." *Annual Review of Resource Economics* 1: 409–434.

Polasky, Stephen, Erik Nelson, Derric Pennington, and Kris Johnson. 2011. "The Impact of Land-Use Change on Ecosystem Services, Biodiversity and Returns to Landowners: A Case Study in the State of Minnesota." *Environmental and Resource Economics* 48(2): 219–242.

Ricketts, Taylor H., Gretchen C. Daily, Paul R. Ehrlich, and Charles D. Michener. 2004. "Economic Value of Tropical Forest to Coffee Production." *Proceedings of National Academy of Science* 101(34): 12579–12582.

Robinson, Lisa A., and James K. Hammitt. 2011. "Behavioral Economics and Regulatory Analysis." *Risk Analysis* 31(9): 1408–1422.

Rosenberger, Randall S., and John B. Loomis. 2003. "Benefits Transfer." In *A Primer on Nonmarket Valuation*, edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown. Dordrecht: Kluwer Academic Press.

Sagoff, Mark. 2005. *The Catskill Parable: A Billion Dollar Misunderstanding*. PERC Reports 23(2). Available at: www.perc.org/articles/article547.php

Shabman, Leonard A., and Sandra S. Batie. 1978. "The Economic Value of Coastal Wetlands: A Critique." *Coastal Zone Management Journal* 4(3): 231–237.

Shogren, Jason F. and Laura O. Taylor. 2008. "On Behavioral-Environmental Economics." *Review of Environmental Economics and Policy* 2(1): 26–44.

Shrestha, Ram K., and John B. Loomis. 2001. "Testing a Meta-analysis Model for Benefit Transfer in International Outdoor Recreation." *Ecological Economics* 39(1): 67–83.

Smith, V. Kerry. 2007. "Reflections on the Literature." *Review of Environmental Economics and Policy* 1(1): 152–165.

Smith, V. Kerry, and Yoshiaki Kaoru. 1990. "Signals or Noise? Explaining the Variation in Recreation Benefit Estimates." *American Journal of Agricultural Economics* 72(2): 419–433.

Smith, V. Kerry, and Eric M. Moore. 2010. "Behavioral Economics and Benefit Cost Analysis." *Environmental and Resource Economics* 46(2): 217–234.

Smith, V. Kerry, Subhrendu K. Pattanayak, and George L. Van Houtven. 2006. "Structural Benefit Transfer: An Example Using VSL Estimates." *Ecological Economics* 60(2): 361–371.

Smith, V. Kerry, George Van Houtven, and Subhrendu K. Pattanayak. 2002. "Benefits Transfer via Preference Calibration: 'Prudential Algebra' for Policy." *Land Economics* 78(1): 132–152.

U.S. Environmental Protection Agency. 2004. Ecological Benefits Assessment Strategic Plan, SAB Review Draft. Washington, D.C.

von Haefen, Roger, and Daniel J. Phaneuf. 2008. "Identifying Demand Parameters in the Presence of Unobservables: A Combined Revealed and Stated Preference Approach." *Journal of Environmental Economics and Management* 56(1): 19–32.

Whigham, Dennis F. 1997. "Ecosystem Functions and Ecosystem Values." In *Ecosystem Function and Human Activities: Reconciling Economics and Ecology*, edited by R. David Simpson and Norman L. Christensen, Jr. New York: Chapman and Hall.

Whitehead, John. C., Bruce K. Johnson, Daniel S. Mason, and Gordon J. Walker. 2013. "Consumption Benefits of National Hockey League Game Trips Estimated from Revealed and Stated Preference Demand Data." *Economic Inquiry* 51(1): 1012–1025.

Whitehead, John C., Subhrendu K. Pattanayak, George L. Van Houtven, and Brett R. Gelso. 2008. "Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: An Assessment of the State of the Science." *Journal of Economic Surveys* 22(5): 872–908.

Whitehead, John. C., Daniel J. Phaneuf, Christopher F. Dumas, Jim Herstine, Jeffery Hill, and Bob Buerger. 2010. "Convergent Validity of Revealed and Stated Recreation Behavior with Quality Change: A Comparison of Multiple and Single Site Demands." *Environmental & Resource Economics* 45(1): 91–112.

Chapter 14

Conclusions

State of the Art and Research Needs

The State of the Art

In the concluding chapter of *The Benefits of Environmental Improvement: Theory and Practice*, Freeman claimed:

> Suppose the administrator of EPA wished to know the magnitude of the benefits accruing from a given pollution control policy. If asked, I believe an economist could specify the economic theory and models he would use, the data he would like to have, and the empirical techniques he would apply to the data to obtain measures of benefits.
>
> (Freeman 1979, 248)

Freeman added five qualifications to this optimistic assessment of the state of the art. Briefly, they were:

1 Where RP methods were not available (for example, for measuring non-use values), the economist would have to resort to SP measures. He was concerned about "the accuracy of responses of individuals in the necessarily hypothetical situations they pose" (Freeman 1979, 249).
2 In the valuation of reductions in mortality risks, there did not appear to be broad acceptance of the idea that individuals' behavior could or should be the basis of welfare measures.
3 Because all welfare measures are conditioned upon the existing distribution of income, to accept these welfare measures for making public policy choices involved accepting the existing distribution as satisfactory.
4 Economic measures of welfare change "must be built upon noneconomic data such as dose–response functions" (Freeman 1979, 250), which often were not available.
5 The data required to implement many of the models and methods would often be difficult and costly to obtain.

In looking back over this list of qualifications nearly 35 years later, there seems to be much less reason to be concerned about the first two of them. Regarding the first qualification, substantial progress has been made in the development of SP methods. We can now be more optimistic about the ability of properly framed questions to generate meaningful data on preferences and values. This does not mean that all the investigator has to do is go out and ask people some questions. However, well-designed SP studies have a place in economists' toolboxes. As to the second qualification, although there is still some debate about it, there is now fairly wide acceptance of the use of RP measures of the value of risk reduction for evaluating public policies that affect health and safety. Such measures are sanctioned, for example, by the U.S. Environmental Protection Agency for use in performing regulatory impact analyses (U.S. EPA 2000). As noted in Chapters 8 and 11, there is a substantial body of evidence on the magnitude of risk premiums revealed in various markets. Rather, the terms of the debate have shifted; the debate now revolves around whether, in valuing risk reduction policies, differences in age, income, health status, cause of death, and the level of baseline risk should be taken into account, and if so, how. The third qualification remains true (as it does for any economic analysis). However, the welfare measures themselves can be used to assess the distributional impacts of any given policy.

Having said this, however, there is also a new qualification to be added to the list. It has to do with the process of fitting data to a model to generate a welfare measure and the effects of the choice of a model and functional form on the welfare measure. There has been a virtual explosion in the number of models available to be exploited, and a number of rich data sets have been created. Several studies show that welfare measures can be sensitive to the choice of a model or the choice of a functional form for a specific model. Some of these studies involve simulations in which the "true" welfare measure is known and can be compared with estimates derived by fitting the simulated data to alternative models. For examples, see Cropper et al. (1988, 1993), Kuminoff, Parmeter, and Pope (2010) and Kling (1988, 1989). Other studies have involved examining the sensitivity of welfare estimates to changes in some aspect of the specification of the model being applied to real data; for an example, see Cooper and Loomis (1992) and von Haefen and Phaneuf (2003). Because the true model cannot be known, we must add model uncertainty to the list of sources of uncertainty in welfare measures. An important area for future research is to obtain a better understanding of the sources and properties of model uncertainty.

The first chapter of this book discussed several ways of classifying the types of environmental and resource service flows for which value measures might be desired. The book concludes here by briefly outlining the types of models that can be applied to value each type of service flow. This information is summarized in Table 14.1. The first thing to note is that SP models apply to the estimation of all types of service flows except those that work through changes in market prices (producers' and consumers' surpluses). SP methods are all that is available for estimating nonuse values.

Table 14.1 Models for estimating the values of environmental services

	Type of Model						
Type of environmental service flow	*Averting Behavior (Chapters 4 and 7)*	*Health production function (Chapter 7)*	*Changes in producers' and consumers' surpluses (Chapter 8)*	*Recreation Demand Models (Chapter 9)*	*Property Value Models (Chapter 10)*	*Hedonic Wage Models (Chapter 11)*	*Stated Preferences (Chapter 12)*
Human health:							
Mortality	X				X	X	X
Chronic morbidity	X				X		X
Acute morbidity	X	X			X		X
Other direct impacts on humans (amenities, visibility, noise)	X				X	X	X
Economic productivity of ecological systems			X				
Other ecological services (recreation)				X			X
Effects on nonliving systems, such as materials	X		X		X		X
Nonuse values such as ecological stability and biodiversity							X

For valuing mortality risks, some version of the hedonic wage model is probably the method of choice, although averting behavior and SP studies have been conducted in this area with some success. Some form of averting behavior model would be most appropriate for valuing the nonfatal health effects as well. Even the health production function model can be considered a special case of the averting behavior models in that it is based on substitutability among market and nonmarket inputs into the production of health.

Hedonic property value models capture the value of differences in amenity and disamenity levels across residential locations, but they might also measure other environmental effects caused by the same environmental agent. For example, air pollution may reduce visibility and cause adverse health effects and damages to household materials, all of which would be reflected in property prices. Hedonic wage models can be used to value differences in the average levels of amenities or disamenities among cities or regions.

Human exploitation of natural and managed ecosystems in agriculture, commercial fisheries, and forestry is largely managed through market mechanisms. Thus, the values of changes in these ecological services will be reflected in changes in prices and incomes. The cost function and production function models described in Chapter 7 provide the basis for estimating the resulting changes in producers' and consumers' surpluses. However, these models are also applicable where changes in other environmental services affect the productivity and costs of firms. Examples include the effects of water quality on processing costs and the effects of air pollution on manufacturing costs and on the costs of repair and maintenance of structures.

Finally, humans use both natural and managed ecosystems for various forms of recreation. A suite of recreation demand models, detailed in Chapter 9, are available to exploit information on the implicit price of access and on the cost of traveling to a site to estimate the values of sites and changes in their characteristics.

Research Needs

In many ways, both the theory and the empirical methods used in measuring environmental and resource values have matured significantly since the first edition of this book. Stated preference techniques have undergone rigorous evaluation and evolution and have become an essential tool in the area. Recreation demand models have evolved from simple site selection and participation models to complex representations of consumer preferences often requiring elaborate econometric techniques to estimate them. Property and wage models have seen both advances in the underlying theory used to model quality differentiated products, such as those emerging from the equilibrium sorting literature, and a rapid growth in the data available for use in estimation. Yet, there remain numerous areas in which additional research is needed. In this last section of the book, we seek to identify a few of these areas, in the hopes of spurring further research into the field. The list, of course, is not exhaustive, but rather suggestive and in no particular order of importance.

Behavioral Economics – Theory and Implementation

Documenting and understanding departures from neoclassical theory, both in the lab and in the field, remains an active area of research. The implications of behavioral economics for economics broadly, and for welfare analysis in particular, is still unclear; but work in this area should and will undoubtedly continue. Of particular interest from the perspective of this book is the conceptual work on behavioral welfare analysis by Bernheim and Rangel (2009), and more recently Fleurbaey and Schokkaert (2013), with an accompanying need to develop practical approaches for implementing the welfare metrics being proposed.

Modeling Choice Under Uncertainty and Risk

Many of the models used in nonmarket valuation are based on the Random Utility Maximization (RUM) framework. An underlying assumption of this model is that the decision-maker *knows* the conditional utility associated with a given choice alternative. If there is risk (i.e., a variety of possible states of the world, each with a known subjective probability of occurring), the structure implicitly assumes that the individual is able to integrate out that risk to form an aggregate value for each available choice alternative. This can be done on the basis of expected utility (EU) theory or some other non-EU model (such as loss aversion). The difficulty here is in choosing or discerning which aggregate value and decision rule is being used by a survey respondent and eliciting relevant information that they use in making that decision, including what they perceive to be possible states of the world and what subjective probabilities they associate with those states. The challenges become even more daunting in the setting of uncertainty (or ambiguity), when the decision-maker is no longer assumed to know subjective probabilities for the different states of the world. It is clear that more research is needed into understanding the decision-making process under both risk and uncertainty and in developing practical methods for gathering the requisite information from decision-makers. De Palma et al. (2008) provide an excellent introduction to these issues.

Stated Preference Surveys and the Role of Consequentiality

As noted in Chapter 12, there has been a paradigm shift in evaluating stated preference approaches to nonmarket valuation. Whereas much of the research to date assessing the criterion validity of SP methods has compared purely hypothetical surveys with real transactions in experimental settings, Carson and Groves (2007) emphasized the need instead to understand and evaluate the underlying incentives provided by the survey instrument to induce preference revelation. An important component in making a survey incentive compatible is making it consequential from the perspective of the survey respondent. To date, relatively few studies have tested the role of consequentiality in SP survey responses and none that we are aware of have done so using the increasingly popular choice experiment format. Moreover, additional research is needed into how best to convey consequentiality and assess the efficacy of a particular consequentiality script. More research is also needed to understand the role of incentive compatibility in the choice experiment framework.

Validity Tests of Revealed Preference Methods

Kling, Phaneuf, and Zhao (2012) suggest that much could be learned by applying to revealed preference methods the same type of validity tests that stated preference methods have been subjected to. With the exception of Bishop and Heberlein's (1979) early validity tests using goose hunting permits (where recreation demand estimates were included with contingent valuation estimates to compare with actual transactions), almost all such tests have focused exclusively on stated preference methods. Given the many sources of potential error in hedonic methods, recreation demand models, averting behavior estimates, and other revealed preference models, it is somewhat surprising that so little attention has been paid to this question.

Defining the Choice Set

In modeling the demand for recreation, analysts typically assume that the choice set being considered by the individual is either the complete choice set available or one limited in a consistent fashion across individuals (e.g., alternatives within a fixed distance of the individuals' home). However, knowledge of the full choice set is likely to differ across individuals (for example, due to their time living in a given region of the country) and to evolve endogenously over time as they seek to learn more about the available alternatives. One strand of research in this area seeks to distinguish the full set of alternatives from the options the decision-makers seriously consider (that is, the so-called "consideration set"). In some cases, authors have tried to elicit the consideration set from survey respondents (see Hicks and Strand 2000), while others have treated an alternative's inclusion in the consideration set as a latent variable to be estimated econometrically (see Ben-Akiva and Boccara 1995; Haab and Hicks 1997; von Haefen 2008). While this represents a potentially useful line of research, there are concerns regarding the feasibility of extracting the desired choice set through a survey instrument (see, for example, Horowitz and Louviere 1995). At the same time, econometrically characterizing the choice set without such information relies heavily on distributional assumptions in order to identify the consideration set. More research and data development are clearly needed in order to better understand the dynamics underlying the formation of the relevant choice set. Analogous problems exist in transportation, marketing, and labor market research—areas that represent potentially fruitful literatures for the cross-fertilization of ideas. Finally, it should be noted that the problem of identifying the choice set is not restricted to recreation demand models; it is an issue for property value models (Banzhaf and Smith 2007) and in modeling stated preference survey responses.

The Opportunity Cost of Time

In inferring the value associated with recreational sites and changes in the associated amenities, one of the key elements is the opportunity cost of time. As noted in Chapter 9, this topic has received considerable attention in the literature, with a variety of approaches developed to account for differences in the constraints and opportunities facing an individual. Nonetheless, this an area where additional research would seem worthwhile given the pivotal role it plays in recreation demand modeling. Palmquist, Phaneuf, and Smith (2010), for example, consider the fact that leisure time is often broken down into component parts that cannot easily be combined to create the large blocks of time needed for certain types of recreation. This in turn has implications for the opportunity cost of time. Essentially, rather than facing a single time constraint, households face a series of constraints, with elemental components of time that are not readily substitutable for one another.

Sustainability and Resilience

The concepts of sustainability and its cousin, resilience, are quickly gaining a foothold in many government agencies and private businesses as organizing principles for planning. A commonly cited working definition of sustainability in the context of development comes from the Brundtland Report (World Commission on Environment and Development, 1987): "sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs." Others discuss sustainability in the context of three pillars: economic, environmental, and social. The term resilience has not received as much attention to date, but is increasingly discussed as a beneficial characteristic of a system. As economists, a focus on the sustainability and/or resilience of a system seems quite incomplete without considering them in the context of a broader objective addressing overall social welfare. Research on the role of sustainability and/or resilience in evaluating the optimality of a development path or use of the environment is needed. The tools of nonmarket valuation should be able to provide insight on possible tradeoffs between sustainability, resilience, and conventional social welfare optimization. Further, the tools of nonmarket valuation may prove useful in constructing metrics of sustainability of resilience.

Valuing Risk Reduction (Mortality/Morbidity)

As noted in Chapters 7 and 11, significant work is needed to improve and update the values of reducing risks that lead to morbidity and/or mortality. The extremely old studies being relied upon for many U.S. government benefit-cost analyses of regulations are sadly out of date and reflect risk and population characteristics that are often not policy relevant. Further, it is now generally appreciated that the value of risk reduction will be a function of many features of the type of risk faced and the populations subjected to that risk. Finally, the puzzle concerning why stated preference estimates of risk reduction lie almost universally below the revealed preference estimates is sorely in need of study.

Ecosystem Services

As noted in Chapter 13, the concept of ecosystem services has gained ground over the past decade. The notion that the environment and ecosystems provide services to human beings is compelling and clear from a general perspective, but clearly identifying the full suite of services provided by a particular environmental change or alteration of an ecosystem remains elusive. This is an area where economists must work closely with ecologists to identify the services and develop methods to measures the services in ways that are meaningful to human usage and values.

General Equilibrium

In nonmarket valuation, and in environmental economics more generally, there has been increased interest in the general equilibrium effects stemming from environmental changes (driven either by policy or by other forces). This has clearly been the case in the property valuation literature, with the emergence of the equilibrium sorting literature (see Kuminoff, Smith, and Timmins, 2013), where the models seek to characterize the equilibrium process in a housing market given a shock to the system. These equilibrating factors need not be restricted to market forces, but can also include feedback effects from environmental and biological factors, such as fishing stocks or habitat conditions (see, for example, Phaneuf, Carbone, and Herriges 2009, and Smith et al. 2010). Further research along these lines would seem beneficial, both because policy programs often induce feedback effects that cannot be adequately captured by marginal analysis and because these effects often require an understanding of the physical processes impacted by large-scale policy initiatives. Incorporating both wage and housing market equilibrium processes may be an important area for such research—see Kuminoff (2012).

Combining Revealed and Stated Preference Data

As noted in Chapter 13, a large and growing literature has been developing in which economists combine revealed and stated preference information to jointly estimate the parameters of a common utility or demand function. Whitehead et al. (2008) provide an excellent review of this work. One area that seems to us to be in need of additional work is the combination of stated and revealed preference data to estimate nonmarket values when the property of weak complementarity cannot be reasonably assumed to hold. As discussed in Herriges, Kling, and Phaneuf (2004), when the property of weak complementarity does not hold, the traditional approach of computing use value from revealed preference data and identifying it as a lower bound cannot be assumed to be accurate. However, including appropriately designed stated preference data has the potential for providing the missing components of information (Ebert 1998; Eom and Larson 2006). Returns to research in this area could be quite substantive.

Valuation in Developing Countries

Much, though not all, of the valuation literature has evolved in industrialized countries, where market economies are the norm. Yet there are many environmental issues that impact developing countries, such as those stemming from climate change, and an understanding of the costs and benefits of alternative initiatives would facilitate policy design. Adapting both revealed and stated preference methods to these settings will require understanding how the local economy and social structures alter both individuals' incentives and their perceptions regarding the tradeoffs available to them.

Dynamic Welfare Measures

There are several places in the book where dynamic aspects of welfare measurement are referred to, including the theoretical basis for welfare measures that incorporate uncertainty and learning in Chapter 5, empirical recreation models that attempt to incorporate dynamic behavior in Chapter 9, and equilibrium sorting models in housing and wage markets that recognize to some degree a temporal response to a change in environmental conditions. In many cases, modeling and addressing the dynamic nature of welfare measures may be unnecessary as when it can reasonably be expected that markets are in long-run equilibrium. However, when working with individual observation data such as typically collected in stated preference surveys or recreation demand data sets, there may be elements of dynamic behavior that are important to understand and identify with respect to policy. It will be necessary to carefully design data collection efforts to study the dynamic components of welfare measures, but the potential for improved understanding of fundamental valuation puzzles such as the willingness-to-pay and willingness-to-accept disparity seems high (Kling, List, and Zhao 2013).

Valuing Ecosystem and Other Environmental Effects of Climate Change

Climate change and its impacts on ecosystem services, invasive species, biodiversity, and a host of other ecological endpoints are poorly understood in the scientific community. This makes nonmarket valuation of these effects even harder.

Economic models that assess welfare losses from climate change do not include more than a cursory assessment of nonmarket values (Nordhaus 2008; Stern 2009). Yet, without such measures, assessment of the impacts of climate change on the welfare of current and future generations cannot be fully understood. While daunting, the returns to research that better quantifies the welfare losses from climate change via its impact on the environmental are likely to be high.

References

Banzhaf, H. Spencer, and V. Kerry Smith. 2007. "Meta-analysis in Model Implementation: Choice Sets and the Valuation of Air Quality Improvements." *Journal of Applied Econometrics* 22(6): 1013–1031.

Ben-Akiva, Moshe, and Bruno Boccara. 1995. "Discrete Choice Models and Latent Consideration Sets." *International Journal of Research in Marketing* 12(1): 9–24.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124(1): 51–104.

Bishop, Richard C., and Thomas A. Heberlein. 1979. "Measuring Values of Extra Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61(5): 926–930.

Carson, Richard T., and Theodore Groves. 2007. "Incentive and Informational Properties of Preference Questions." *Environmental and Resource Economics* 37(1): 181–210.

Cooper, John, and John Loomis. 1992. "Sensitivity of Willingness-to-Pay Estimates Bid Design in Dichotomous Choice Contingent Valuation Models." *Land Economics* 68(2): 211–224.

Cropper, Maureen L., Leland B. Deck, Nalin Kishor, and Kenneth E. McConnell. 1993. "Valuing Product Attributes Using Single Market Data: A Comparison of Hedonic and Discrete Choice Approaches." *Review of Economics and Statistics* 75(2): 225–232.

Cropper, Maureen L., Leland B. Deck, and Kenneth E. McConnell. 1988. "On the Choice of Functional Form for Hedonic Price Functions." *Review of Economics and Statistics* 70(4): 668–675.

De Palma, Andre, Moshe Ben-Akiva, David Brownstone, Charles Holt, Thierry Magnac, Daniel McFadden, Peter Moffatt, Nathalie Picard, Kenneth Train, Peter Wakker, and Joan Walker. 2008. "Risk, Uncertainty, and Discrete Choice Models." *Marketing Letters* 19(3–4): 269–285.

Ebert, Udo. 1998. "Evaluation of Nonmarket Goods: Recovering Unconditional Preferences." *American Journal of Agricultural Economics* 80(2): 241–254.

Eom, Young-Sook, and Douglas Larson. 2006. "Improving Environmental Valuation Estimates through Consistent Use of Multiple Information." *Journal of Environmental Economics and Management* 52(1): 501–516.

Fleurbaey, Marc, and Erik Schokkaert. 2013. "Behavioral Welfare Economics and Redistribution." *American Economic Journal: Microeconomics* 5(3): 180–205.

Freeman, A. Myrick, III. 1979. *The Benefits of Environmental Improvement: Theory and Practice*. Baltimore: Johns Hopkins University Press for Resources for the Future.

Haab, Timothy C., and Robert L. Hicks. 1997. "Accounting for Choice Set Endogeneity in Random Utility Models of Recreation Demand." *Journal of Environmental Economics and Management* 34(2): 127–147.

Herriges, Joseph A., Catherine L. Kling, and Daniel J. Phaneuf. 2004. "What's the Use: Welfare Estimates from Revealed Preference Models in the Absence of Weak Complementarity." *Journal of Environmental Economics and Management* 47(1): 55–70.

Hicks, Robert L., and Ivar E. Strand. 2000. "The Extent of Information: Its Relevance for Random Utility Models." *Land Economics* 76(3): 374–385.

Horowitz, Joel L., and Jordan J. Louviere. 1995. "What is the Role of Consideration Sets in Choice Modeling?" *International Journal of Research in Marketing* 12(1): 39–54.

Kling, Catherine L. 1988. "The Reliability of Estimates of Environmental Benefits from Recreation Demand Models." *American Journal of Agricultural Economics* 70(4): 892–901.

Kling, Catherine L. 1989. "The Importance of Functional Form in the Estimation of Welfare." *Western Journal of Agricultural Economics* 14(1): 168–174.

Kling, Catherine L., John List, and Jinhua Zhao. 2013. "A Dynamic Explanation of the Willingness to Pay and Willingness to Accept Disparity." *Economic Inquiry* 51(1): 909–921.

Kling, Catherine L., Daniel J. Phaneuf, and Jinhua Zhao. 2012. "From Exxon to BP: Has Some Number Become Better than No Number?" *Journal of Economic Perspectives* 26(4): 3–26.

Kuminoff, Nicholai. 2012. "Partial Identification of Preferences in a Dual-Market Sorting Equilibrium." Working paper, Arizona State University.

Kuminoff, Nicholai V., Christopher F. Parmeter, and Jaren C. Pope. 2010. "Which Hedonic Models Can We Trust to Recover the Marginal Willingness to Pay for Environmental Amenities." *Journal of Environmental Economics and Management* 60(3): 145–160.

Kuminoff, Nicholai, V. Kerry Smith, and Christopher Timmins. 2013. "The New Economics of Equilibrium Sorting and Policy Evaluation Using Hedonic Markets." *Journal of Economic Literature* 51(4): 1007–1062..

Nordhaus, William. 2008. *A Question of Balance: Weighing the Options on Global Warming Policies.* New Haven and London: Yale University Press.

Palmquist, Raymond B., Daniel J. Phaneuf, and V. Kerry Smith. 2010. "Short Run Constraints and the Increasing Marginal Value of Time in Recreation." *Environmental & Resource Economics* 46(1): 19–41.

Phaneuf, Daniel J., Jared C. Carbone, and Joseph A. Herriges. 2009. "Non-Price Equilibria for Non-Marketed Goods." *Journal of Environmental Economics and Management* 57(1): 45–64.

Smith, Martin D., John Lynham, James N. Sanchirico, James A. Wilson, and Steven D. Gaines. 2010. "The Political Economy of Marine Reserves: Understanding the Role of Opportunity Costs." *Proceedings of the National Academy of Sciences* 107(43): 18300–18305.

Stern, Nicholas. 2009. *The Global Deal: Climate Change and the Creation of a New Era of Progress and Prosperity*. New York: Public Affairs.

U.S. EPA. 2000. *Guidelines for Preparing Economic Analyses*. Washington, D.C.: U.S. Environmental Protection Agency.

von Haefen, Roger H. 2008. "Latent Consideration Sets and Continuous Demand System Models." *Environmental and Resource Economics* 41(3): 363–379.

von Haefen, Roger H., and Daniel J. Phaneuf. 2003. "Estimating Preferences for Outdoor Recreation: A Comparison of Continuous and Count Data Demand System Frameworks." *Journal of Environmental Economics & Management* 45(3): 612–630.

Whitehead, John C., Subhrendu K. Pattanayak, George L. Van Houtven, and Brett R. Gelso. 2008. "Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: An Assessment of the State of the Science." *Journal of Economic Surveys* 22(5): 872–908.

World Commission on Environment and Development. 1987. *Our Common Future.* Oxford: Oxford University Press.

Author Index

Subject Index