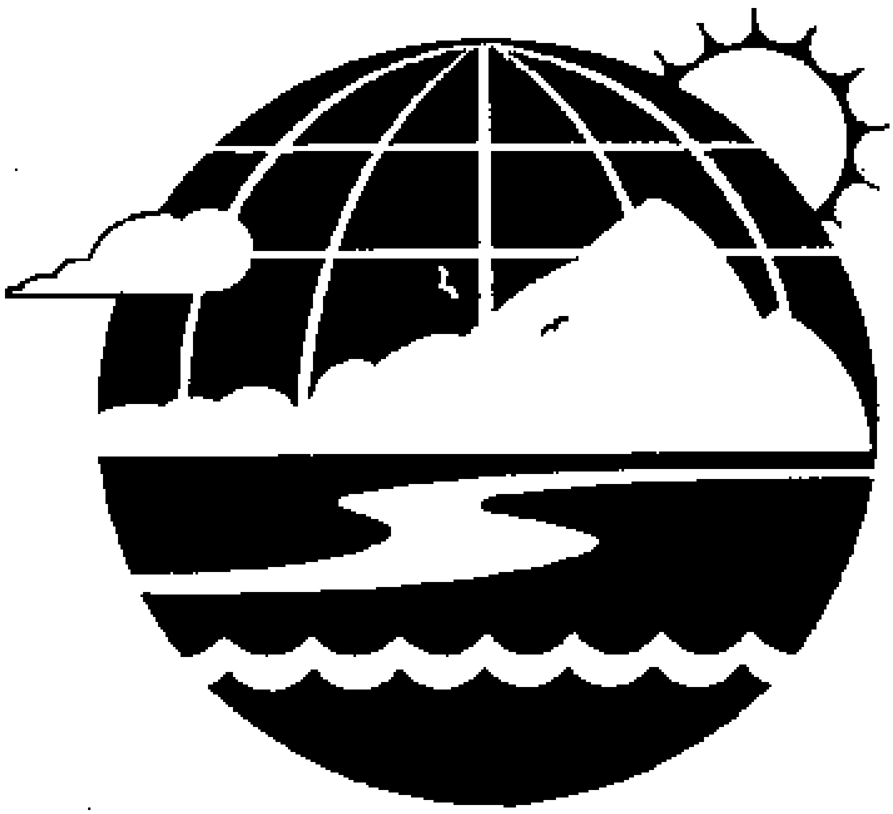
United States Environmental Protection Agency
Policy, Planning and Evaluation (2127)
EPA-230-10-89-069
October 1989

Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models: Part I



MEASURING THE BENEFITS

OF WATER QUALITY IMPROVEMENTS

USING RECREATION DEMAND MODELS

Volume II

of

BENEFIT ANALYSIS USING

INDIRECT OR IMPUTED MARKET METHODS

Prepared and Edited by

Nancy E. Bockstael
University of Maryland

W. Michael Hanemann
University of California

Ivar E. Strand, Jr.
University of Maryland

Principal Investigators

Kenneth E. McConnell and Nancy E. Bockstael
Agricultural and Resource Economics
University bf Maryland

EPA Contract No. CR-811043-O1-O

Project Officer

Dr. Alan Carlin
Office of Policy Analysis
Office of Policy and Resource Management
U. S. Environmental Protection Agency
Washington, D.C. 20460

The information in this document has been funded wholly or in part by the United States Environmental Protection Agency under Cooperative Agreement No. 811043-01-0. It has been subjected to the Agency's peer and administrative review, and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ACKNOWLEDGEMENTS

Kenneth E. McConnell, Terrence P. Smith, and Catherine L. Kling were major contributors to this volume, providing both original contributions and beneficial comments. The authors have also benefited from comments of EPA staff members, including Peter Caulkins and George Parsons, and from reviewers Edward Morey and Clifford Russell. Technical assistance was provided by Jeffrey Cunningham and Chester Hall. Additionally, both credit and appreciation are due Alan Carlin, our project officer in the Benefits Staff of EPA. Throughout the research, Diane Walbesser supplied invaluable secretarial and technical assistance. Finally, Linda Griffin of ADEA Wordprocessing and Patricia Sinclair of the University of Maryland deserve special thanks for undertaking the arduous task of typing this manuscript.

All opinions and remaining errors are the sole responsibility of the editors. This effort was funded by US EPA Cooperative Agreement number CR-811043-01-0.

FOREWARD

This is the second of two volumes constituting the final report for budget period I of Cooperative Agreement #811043-01-0, which was initiated and supported by the Benefits Staff in the Office of Policy Analysis at the U.S. Environmental Protection Agency (EPA). The two volumes, whi1e encom-passed under the same cooperative agreement, are distinct in nature. The topic of Volume 11 is the use of recreational demand models in estimating the benefits of water quality improvements.

The research reported here is the result of interaction among the principal investigators of the project, the editors of the volume, individual contributors at the University of Maryland, and outside reviewers. In addition to the team of editors, Kenneth E. McConnell, Terrence P. Smith, and Catherine L. Kling were major contributors, providing both original research and invaluable review.

The editors benefited considerably from comments by outside reviewers, Edward Morey of University of Colorado and Clifford Russell, now of Vanderbilt University. Important contributions were also made by EPA staff including Alan Carlin, Peter Caulkins, George Parsons and Walter Milon. It would be impossible to cite all the individuals who had an influence on the ideas presented here, but two of these must be mentioned, V. Kerry Smith of Vanderbilt University and Richard Bishop of the University of Wisconsin.

Progress made in this volume toward the resolution of the problems and dilemmas which plague the assessment of environmental quality improvements must be attributed to a wide range of sources. In large part the work reflects the cumulative efforts of a decade or two of researchers in this area. And, it is itself merely a transitionary stage in the development and synthesis of the answers to those problems. More progress has already been made on many of these issues - both by the authors and by other economists working in the field. This new work will be reflected in future cooperative agreement reports.

Also, included in the next budget period’s report will be discussion and analysis of survey data collected during budget period I. The survey, designed by Strand, McConnell and Bockstael in conjunction with Research Triangle Institute (RTI), was administered by RTI. It includes a telephone survey of households in the Baltimore-Washington SMSA's and a field survey conducted during the summer of 1984 at public beaches on the Western shore of the Chesapeake Bay. The survey provides data on swimming behavior which is being analyzed using some of the developments discussed in this volume. The survey instrument, the data, and the analysis will be presented in the next cooperative agreement report.

EXECUTIVE SUMMARY

In an era of growing Federal accountability, those programs which cannot substantiate returns commensurate with budgets are severely disadvantaged. Expressions such as Executive Order 12291 require an. account of the benefits of public interventions. Inability to provide, or inaccuracy in the provision of, those estimates undermines the credibility of programs and may cause their untimely demise.

The public provision of improvements in water quality is an activity endangered by the complexities involved in the accounting of benefits. The lack of markets and observed prices in water-related recreational activity has necessitated the use of surrogate prices in benefit assessment. Moreover, a formal regime (i.e. The Principles and Standards for Water Quality) articulates the assessment procedure. Unfortunately, the regime still contains ambiguities, inconsistencies and slippage sufficient to raise potential controversy over any estimate of benefits from water quality improvements.

The purpose of Volume 11 is to address some of those ambiguities and inconsistencies and, in so doing, provide a more comprehensive, credible approach to the valuation of benefits from water quality improvements. Substantial progress is made in improving valuation techniques by linking the fundamental concepts of the “travel cost” model with cutting-edge advances in the labor supply, welfare, and econometrics literature.

At the heart of the research is the study of individual recreation behavior. As water quality improves, individual behavior changes, reflecting improvements in welfare. Misconceptions and inaccuracies may’ arise if benefit evaluations are based on inappropriate aggregation of individual’s behavior. An analysis of the “zonal” (an aggregate) approach represents one contribution of Volume II. Alternatives to the zonal approach are offered. The new approaches are based on advances in the statistical analysis of limited dependent variables.

The realities of recreational choice encompass more dimensions than traditional demand analysis. Time is critical - over 50% of respondents in a recent national survey replied that “not enough time” was the reason they did not participate more often in their favorite recreation, while only 20% replied “not enough money.” Drawing on labor supply literature, an extension of traditional demand analysis to include time constraints is developed in Volume XI. The extension, which is made operational, captures the true nature of recreational decisions which are affected as much by individuals’ time constraints as their money constraints.

Statistical analysis is emphasized throughout the volume. One example is an examination of the properties of welfare estimates. Because typical welfare estimates are derived from numbers with random components, they have random components themselves. Thus it is important to study the statistical properties of typically used estimators for welfare measures. These properties, such as biasedness, are shown to be undesirable in several instances. More credible estimators are provided. Another statistical issue, causes of randomness in estimates, is shown to influence the magnitude of welfare estimates. Ways in which information about the source of randomness can be used to improve accuracy are discussed.

Part II of Volume addresses problems specifically associated with introducing aspects of water quality into the fundamental model developed in Part I. The desire to incorporate environmental characteristics (such as water quality) has prompted the treatment of an additional dimension to the recreational model. Data collected for one recreational site do not, by their nature, exhibit variation in the quality characteristics of that site, preventing the researcher from deducing anything about how demand changes with changes in quality characteristics. The only reliable means of incorporating quality is to model the demand for an array of sites of differing qualities. However, the need to develop models of multiple site decisions has been a blessing in disguise, for it has forced modelers to recognize that recreational decisions are frequently made among an array of competing, quality-differentiated resources.

A major share of Part II of this volume is devoted to the discussion of models which can incorporate quality characteristics in multiple site recreational demand decisions. While a theoretically consistent model can be developed, it is not empirically feasible, and several second best models are presented. Criteria for evaluating these alternative models includes their ability to capture the nature of recreational decisions and to respond to the research goal of valuing environmental quality changes.

TABLE OF CONTENTS

Part I

ADVANCES IN THE USE OF RECREATIONAL DEMAND MODELS FOR BENEFIT VALUATION

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS

Part II

MULTIPLE SITE DEMAND MODELS AND THE MEASUREMENT OF BENEFITS FROM WATER QUALITY IMPROVEMENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

PART I

ADVANCES IN THE USE OF
RECREATIONAL DEMAND MODELS
FOR BENEFIT VALUATION

CHAPTER 1

INTRODUCTION

Volumes I and II of this report are the result of one year's research conducted under EPA Cooperative Agreement CR-811043-O1-O. The particular methods designated by EPA to be of primary interest in this cooperative agreement are "imputed or indirect market methods," i.e. methods which depend on observed behavior in related markets rather than direct hypothetical questioning. Despite their similar themes, the two volumes are distinct in many respects. Volume I addresses a specific technical issue (the identification problem) associated with the hedonic method of valuing goods. The second volume discusses a wider range of technical issues associated with the use of recreational demand models to value environmental quality changes. The primary purpose of the agreement has been to develop and demonstrate improved methods for estimating the regional benefits from environmental improvements.

Within this volume dedicated to recreation demand models, Part I is restricted to a set of issues which arise in benefit valuation using the conventional single site recreational model. The topic of Part 11 is the application of recreation demand models for the specific task of measuring the benefits associated with changes in the quality of the recreational experience. Attention is given, in particular, to water quality improvements. In this spirit, Part II explores a broad range of models based on individual behavior which can be used to reveal valuations of environmental improvements. These models attempt to establish the relationship between use activities (specifically recreation) and water quality and can be used to devise welfare measures to assess benefits.

The emphasis this volume gives to recreation behavior is not misplaced. A 1979 report by Freeman (1979b) to the Council of Environmental Quality estimated that over fifty percent of the returns from air and water quality improvements would accrue through recreational uses of the environment. When considering water quality improvements alone, the percentage was even higher. One of the earliest studies attempting to quantify such effects (Federal Water Pollution Control, 1966) estimated that recreationists would receive more than 95% of the benefits derived from water quality improvements in the Delaware estuary. These sentiments were further supported by the U.S. National Commission on Water Quality (1975) which maintained that water based recreators would be the major beneficiaries of the 1972 Federal Water Pollution Control Act.

Thus, the emphasis in these two volumes is on recreation, but the tasks are wide-ranging. The initial charge in the Cooperative Agreement was a broad one, including the development of improved methods, the demonstration of new techniques, the collection of primary data and the assessment of the usefulness of the resulting benefit estimates. The emphasis in this first year of work has been where it must be, on the first items in this list, although progress has been make on each task.

Nonmarket Benefit Evaluation and the Development of Methods

Despite the near consensus which currently exists in market-oriented welfare theory (i.e. welfare changes in private markets), economists are far from embracing a complete methodology for valuing public (often environmental), non-market goods. It hardly seems necessary to document this contention. One need only consider some of the many recent conferences which have attempted to resolve difficulties and increase consensus on these issues, (e.g. Southern Natural Resource Economics Committee, Stoll, Shulstad and Smathers, 1983; Cummings, Brookshire and Schulze, 1984; EPA Morkshop on the State of the Art in Contingent Valuation, and AERE Workshop on Valuation of Environmental Amenities, 1985.) In essence “Nonmarket valuation has a long way yet to go before all the problems will be solved and its acceptance by economists will be unequivocal (SNREC, p.4).”

The valuation exercise has been viewed by many economists as an attempt to bring nonmarket goods into policy considerations on a comparable footing with private marketed goods. However, to be accurate, some economists and many non-economists have questioned the relevancy of the market analogy for public good valuation. Arguments by philosophers include reference to a social ethic and contend that societies may have collective values independent of individual preferences. Not so well articulated are our own concerns about how people think about public goods and how they relate public goods to private expenditures. To what extent can a change in a public good be translated into an effect on an individual such that an individual’s willingness to pay is a meaningful concept?

The existence of rival theories and the lack of consensus we see in the non-market benefits literature is not unlike the early stages of the development of other fields of economics and of other sciences. In the early stages of a science or a subfield of a science, Thomas Kuhn has argued that competition exists among a number of distinct views all somewhat arbitrary in their formulation. Eventually a set of theories, Kuhn's now familiar "paradigm," emerges which provides focus to future work. The paradigm is the set of fundamental concepts and theories which all additional work takes as given. The eventual acceptance of a paradigm allows, and in fact encourages, research to become more focused, more refined, and more detailed. This body of accepted thought provides the necessary structure and standards of judgement without which research becomes confusion. Kuhn's essential point was that the science could only be advanced in the context of the paradigm.

Whether we wish to view it as a pre-paradigm stage or a crisis in the neoclassical paradigm, the development of what has become "traditional" welfare economics (i.e. welfare measurement in private markets) provides a case in point. Welfare economics has a long history of controversy, beginning with loosely defined and imprecisely measured concepts of rent and consumer surplus extending as far back as Ricardo and Dupuit. The establishment of these concepts as foundations of a theory of economic welfare was a long and uphill battle involving attacks by new welfare economists on the old welfare economics and the development of the compensation principle. For a very long period the state of welfare economics was one of crisis, with applied economists pursuing empirical studies which theoreticians condemned. Over time, and with theoretical developments by economists such as Willig, Hausman, Just et al., Hanemann, and others, a theoretical foundation for feasible empirical practices has emerged in the form of the "willingness to pay" paradigm.

With the recognition that public policies frequently produce benefits and losses outside of markets comes a new controversy and an attempt to stretch the existing "willingness to pay" paradigm to cover new ground. To many established economists, the problem seems straightforward: the valuation of nonmarket benefits through benefit-cost analysis, under ideal procedures for extracting value measures, is assumed to provide the same answer that the market mechanism would provide. The major difficulties lie in defining those ideal procedures. Some question whether these measures exist, or are meaningful, in the context in which we wish to use them - i.e.

can the willingness-to-pay paradigm really be stretched and modified to resolve the anomalies which public good valuation present?

This subfield of economics, the valuation of public goods, is in a period of crisis in its development, but it is not unlike periods of crisis which have arisen in other areas of economics or in the natural and physical sciences. Kuhn describes these periods as marked by debates over legitimate methods, over relevant experiments, and over standards by which results can be judged - a description which fits closely the current activities in non-market valuation. In these periods of crisis, Kuhn argues, many speculative and unarticulated theories develop which eventually point the way to discovery.

The implication of Kuhn's thesis is that more refined and precise analysis either establishes a closer match between theory and observation or provides more evidence that such a match does not exist. The only way to determine whether standard welfare economics can be stretched to resolve the public good valuation problem is to explore nonmarket valuation problems in a rigorous welfare theoretic framework. If the anomalies can not be resolved, even with increasingly careful modelling and precise measurement, then the balance will tip In favor of seeking a new paradigm. But it is only in the context of some carefully conceived theoretical structure that progress can be made. "Truth emerges more readily from error than from confusion (Kuhn, 1969)."

Making Benefit Measures More Defensible

An attempt to apply scientific methods to nonmarket benefit analysis immediately raises problems. Our approaches provide estimates of welfare for which we have no direct observations for comparison. The absence of direct observation on welfare changes directly only suggests that welfare measures should be defined on models of <u>behavior</u> which <u>can</u> be observed.

Starting, as they do, from models of economic behavior, one would think that welfare measures derived from models of observable behavior in markets related to environmental goods (e.g. recreational demand models) would be a popular approach. Certainly, the travel cost approach, a specific variant of more general models of economic behavior, has produced many benefit estimates in its long life. Yet this approach's credibility has been challenged on two counts.

First, policy makers argue that many amenities of interest can not be associated closely enough with a market or with observable behavior to allow for the use of related market methods. This criticism has some very important implications. On the pragmatic side, it is useful to note recent results in contingent valuation assessment. Contingent valuation, the principle alternative method, has been pronounced quite reliable as long as the good to be valued is closely related to a market experience. What is more germane to the argument here is that when valuation is unrelated to observable behavior, it is impossible to test the predictions of theories against observations - and as a consequence we can have no confidence in those predictions. In fact, it is unclear that economic valuation has any meaning in a context where there exists no related observable economic behavior. We are reminded of Kuhn's warning "measurements undertaken without a paradigm seldom lead to any conclusions at all."

The second criticism of market related valuation approaches is that the same valuation problem can generate a vast array of radically different benefit estimates. How can one trust a method which appears capable of generating a number of very different answers to the same question?

If we examine the literature or conduct experiments ourselves, we inevitably encounter this embarrassing problem: benefit estimates seem very sensitive to specification, estimation method, aggregation, etc. It is the contention of the current work, however, that valuation methods based on behavioral models allow the potential for resolving inconsistencies, since the apparent arbitrary choices we make about specification, etc. are really implicit but testable hypotheses about individual behavior. By being more precise about the behavioral assumptions of our models, more defensible benefit estimates can be defined.

The philosophy inherent in our research agenda is that if benefit measures are to be taken seriously by policy makers they must be based on defensible, realistic models of human behavior. Perfect measures can not be defined and will always be inaccessible. But arbitrariness in estimating human behavior can be reduced by careful model specification and estimation, so that we know ultimately what assumptions are implicit in the benefit estimates as well as the direction of possible biases in these estimates.

This philosophy requires that we first assess the state of benefit estimation using indirect market methods and then attempt to make improvements in those areas which seem either the most confused or the most vulnerable. A goal of the current research is to bring together the many recent advances in recreational demand estimation, specifically, and applied welfare economics, more generally, to further the development of defensible models of measuring water quality improvements.

One comment needs to be made with regard to alternative benefit measurement techniques. The arguments in this Chapter are not intended to champion the cause of recreational demand models over contingent valuation techniques. The purpose of this as well as other studies should be to improve the credibility of techniques for valuing environmental amenities. It is our opinion that the science will be advanced if contingent valuation and indirect market methods are considered as complements. To the extent that the two approaches can be made comparable, their conjunctive use can only strengthen benefit estimation. While many studies have compared estimates derived from the two approaches (e.g. Knetsch and Davis 1966; Bishop and Heberlein 1979; Thayer 1981), few have tried to relate the approaches conceptually and none have attempted to ensure that the underlying assumptions of the models are consistent. The two approaches applied to the same circumstances can potentially be made comparable since they are both the realization of individual’s preferences subject to constraints. Just as there are assumptions about behavior implicit in the way in which we specify and estimate recreational demand models, there are similar if less conspicuous assumptions implicit in the way contingent valuation experiments are framed and the way benefit estimates are derived from the hypothetical answers. While a means for making the two approaches comparable is beyond the scope of this year’s project, future efforts in this direction will be rewarding.

The Empirical Foundation of Recreation Demand Models: The Traditional Travel Cost Model

The recent research in environmental valuation has had a foundation upon which to build. The earliest work focused on the valuation of a single recreation site, using aggregate “zonal” data.

> “Let concentric zones be defined around each park so that the cost of travel to the park from all points in one of these zones is approximately constant. . ..If we assume that the benefits are the same no matter what the distance, we have, for those living near the park, consumer’s surplus consisting of the differences in transportation costs. The comparison of the cost of coming from a zone with the number of people who do come from it, together with a count of the population of the zone, enables us to plot one point for each zone on a demand curve for the service of the park (Hotelling 1948).”

In fact the development of methods of estimating the demand for recreation so closely paralleled the use of zonal models that the so-called travel cost method is often considered synonymous with the use of zones.

The concept of this original travel cost model took advantage of the fact that unlike other goods, recreational sites are immobile and users must incur specific costs to access a site. Thus, travel costs were proposed as a proxy for market price, with consumption of the recreational opportunity expected to decline as distance from the site and travel costs rose. Clawson, in 1959, and Clawson and Knetsch, in 1966, developed the travel cost idea into an operational model by estimating demand for a recreation site and measuring the total value or benefits of the site.

This basic model has been widely replicated and extended to account for various complexities of the recreation experience. The procedure is recommended for project benefit estimation in the 1979 revision of the Water Resources Council's "Principles and Standards." Thus a long evolutionary process has established a precedent for the use of travel cost models in valuing aspects of recreation activities.

The essence of the traditional travel cost approach to valuing benefits is shown in Figure 1.1. The sum of travel costs and entrance fees act as a surrogate for the price of the recreational trip. The demand curve of a "representative" individual is estimated by regressing trips per capita in each zone against average travel cost per trip and other average characteristics of each zone. An aggregate demand curve is then formed by combining the representative demand curve with zonal characteristics of the population. The shaded area between the aggregate demand curve and the actual entrance fee is viewed as a measure of the consumers' surplus from the site.

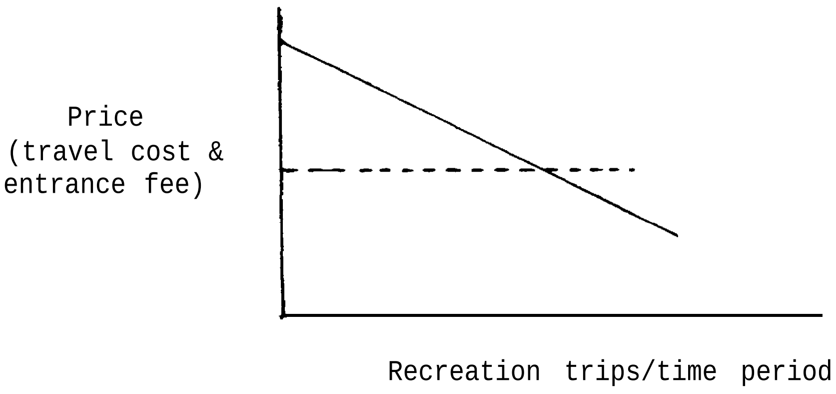


Figure 1.1: The Recreation Demand Curve

The fundamental problem with using the simple travel cost approach as shown above is that it is defensible only in certain rather restrictive circumstances. Much of the research since 1970 has expanded the travel cost model to a more general recreational demand model, making it more defensible in a wider variety of circumstances. In addition, because its role has been benefit estimation, a closer correspondence to axioms of welfare economics has been established. Development of increasingly sophisticated estimation techniques is present throughout this period.

The Theoretical Foundation of Recreation Demand Models: The Household Production Approach

While the travel cost method has been applied to empirical problems for decades, its connections with the theory of welfare economics have only recently been articulated. With the increased acceptance of benefit measurement by the economics discipline in the 1970's came the need to link travel cost valuations to welfare theory. The travel cost method had rested mainly on the presumed analogy between travel costs and market prices. In the 1970's more general models of individual behavior, such as the household production function, established the link between travel cost and individual utility maximizing behavior giving greater credibility to existing empirical practices.

The household production framework is not an approach to estimation but a general model of individual decision making. Its antecedent can be found in the economics literature on the allocation of household time among market and nonmarket employment (Becker, 1965; Becker and Lewis, 1973). The applacability of the household production framework for recreation decisions was first noted by Deyak and Smith (1978) and later explored by Brown, Charbonneau and Hay, (1978).

The household production function takes a broader view of household consumption than traditional market approaches. Commodities, for which individuals possess preferences and from which they derive utility, may not be directly purchasable in the marketplace. In fact some goods which can be purchased may not yield utility directly but may need to be combined with other purchased goods and time to generate utility. Rarely are goods combined by the household rather than by firms unless they require substantial time inputs. Thus, time is a critical feature of the model.

One can then view the household as a producer, purchasing inputs, supplying labor, and producing commodities which it then consumes. This makes for a perfectly defensible utility theoretic decision model which can be expressed as

$$\text{(1a)} \qquad \max\ u(z_1,\ldots,z_n)$$

$$\text{(1b)} \qquad \text{s.t.}\quad z = f(x_1,\ldots,\ x_m,\ t_x)$$

$$\text{(1c)} \qquad Y(t_w,\ w) + R - \sum_{i=1}^{m} p_i\ x_i = 0$$

$$\text{(1d)} \qquad T - t_w - t_x = 0$$

where z's are commodities, x's are market goods, and p their prices, t_x is time spent producing commodities, t_w is time spent working, w is the wage rate, Y is wage income, R is nonwage income, and T is total time endowment. Included in the above series of expressions is the usual utility function (1a), a budget constraint (1c), a production function for the z's (1b), and a time constraint (1d). If one of the z's represents recreational trips with inputs of time, transportation, lodging, equipment, etc., then we have the makings of a recreational demand model.

A major contribution of this framework is that it provides a justification for using the travel cost model in certain instances, as well as a way in which to generalize the traditional model to incorporate other elements. While the household production framework provides a general and flexible way of presenting the individual's (household's) decision problem, restrictions are required to make the model empirically tractable. One difficulty inherent in the general form is that the marginal cost of producing a z_i is likely to be nonlinear. The implications of this for estimation and welfare evaluation are explored in Bockstael and McConnell (1981, 1983) and an application can be found in Strong (1983). If the production technology is Leontief and there is no joint production, however, the marginal cost of producing a z_i (e.g. a recreation trip) is constant and thus functionally analogous to a market price. Interpreting "travel" as the principal input and ignoring the time dimension equates this model to the traditional travel cost model. Travel costs no longer depend for their credibility on being a "proxy" for market price. They are a legitimate component of the marginal cost of producing a trip.

It is important to note that this model, as well as all of welfare theory, is grounded in individual behavior. For this reason, and other more practical ones, researchers have tended to move toward using individual observations rather than zonal averages in more recent applications. The zonal-individual observation controversy will receive greater attention in Chapter 3.

The general model also offers a framework from which other aspects of recreational demand, such as the opportunity cost of time, can be introduced (Desvousges, Smith and McGivney, 1983). As far back as Clawson, researcher's knew time costs were an important determinant of recreational demand. However, these costs have often been ignored or treated in an ad hoc fashion. A treatment of time, which is theoretically consistent and empiri~cal tractable, is the subject of Chapter 4.

The Plan of Research for Part I

The conceptual problems which are addressed in Part I have been chosen because benefit estimates have turned out to be extremely sensitive to their arbitrary treatment. In each case attempts have been made to show the sensitivity by citation to existing literature, by use of existing data sets, or by simulating behavioral experiments. Also we demonstrate, by using existing data or simulation results, the application of each improvement which we develop.

Two criteria are used in the development of improved techniques: theoretical acceptability and empirical tractability. Improvements are proposed only if they can be implemented with accessible econometric techniques and with data which can reasonably be collected with manageable surveys.

Part I makes substantive contributions to the single site or activity recreation demand model. Several issues - such as the treatment of time, specification and functional form, aggregation and benefit estimation - are explored. This work forms the foundation for the multiple site modelling techniques discussed in Part II.

CHAPTER 2

SPECIFICATION OF THE RECREATIONAL DEMAND MODEL: FUNCTIONAL FORM AND WELFARE EVALUATION

In the period of only a few years, a number of theoretical papers concerning precision in welfare measurement and the relationship among welfare measures has emerged. Perhaps the most often cited of these is by Willig (1976), who has shown that the differences among ordinary consumer surplus, compensating variation, and equivalent variation are within bounds which are determined by the income elasticity of demand and the ratio of ordinary surplus to total income. The issue of the accuracy of the approximation has become less consequential since the work by Hanemann (1979, 1980b, 1982d), by Hausman (1981), and by Vartia (1983). The first two have shown how to recover _exact_ welfare measures from some common functional forms of demand functions. The latter has developed algorithms yielding numerical solutions which provide arbitrarily close approximations to true welfare measures for functional forms which have no closed form solutions. The first part of this chapter provides a review of this literature on integrability and exact welfare measures.

The second part of the chapter addresses the choice of functional form. While a particular functional form may be consistent with some underlying preference function, it may not be a preference structure consistent with actual behavior. That is, arbitrary choice of functional form may imply too specific a preference structure and one which is inappropriate for the sample of individuals.

The sensitivity of benefit estimates to functional form has frequently been cited in the literature and may be far greater than differences between Hicksian variation and ordinary surplus measures of benefits. This chapter suggests one means of addressing the choice of functional form. We show how close approximations to compensated welfare measures can be derived from flexible forms of the demand function. Emphasis is given to the choice of functional forms which are both consistent with utility theory and supported by the data.

There are two general ways to develop utility theoretic measures of consumer benefits. The first employs an assumed utility function from which demand functions are derived through the appropriate constrained utility maximization process. The other begins with a demand specification and integrates back to a utility function.

The preferable approach depends on whether the problem in question involves a single good or a vector of related goods. In general, it is desirable to begin with a demand function and integrate to derive welfare measures. As Hausman points out, the only observable information is the quantity-price data, data which can be used to fit demand curves not utility functions. Good econometric practice would suggest we choose the best fitting form of the demand function among theoretically acceptable candidates. The demand function approach is preferable because it allows the researcher to include as choice criteria how closely the functional form corresponds to observed behavior. For these reasons this approach will be used for single site models. Unfortunately, multiple good models pose severe integrability problems. As such we are forced in the latter half of this volume to employ the alternative approach of first choosing a preference structure and then deriving demand functions from that structure.

The conditions for integrating back to an indirect utility function from demand functions are now well known. Integrability depends on solving the system of partial differential equations:

(1) $$\partial m/\partial p_i = x_i(p,m)$$

where m is income, p is the price vector, and xi and pi are the quantity demanded and price of the i^{th} good. The solution is called the income compensation function m(p,c), where c is the constant of integration. This function is identical to our concept of the expenditure function, if c is taken as an index of utility. The indirect utility function can be derived by inverting m(p,u) to obtain u=v(p,m). Hurwicz (1971) has shown that partial differential equations of the type in (1) have solutions if a) the $x_i(.)$ are single valued, differentiable functions and b) the Slutsky symmetry conditions hold:

$$\partial x_i/\partial p_j + x_j \partial x_i/\partial m = \partial x_j/\partial p_i + x_i \partial x_j/\partial m.$$

If the problem of interest involves just one good, the convention is to assume that the prices of all other goods (those not of immediate interest) either are constant or move together so that these goods can be treated as a Hicksian composite commodity with a single price. This price can be represented by a price index, or set to one when price is unlikely to vary over the sample. The problem is now reduced to the two good case: x and a composite commodity. Since a system of N partial differential equations can always be replaced by a system of N - 1 such equations by normalizing on the price of one good, the two good case requires the solution of only one differential equation. There is only one element to the Slutsky matrix now, so there is no question of symmetry, and any function which meets regularity conditions is mathematically integrable (although a closed form solution for the expenditure function may not always exist).

Mathematical integrability does not necessarily imply economic integrability, i.e. that the implied utility function be quasi-concave. Economic integrability conditions require that a) the adding-up restrictions hold, i.e. p’x=m, and the functions are homogeneous of degree zero in prices and income and b) the Slutsky matrix is negative semi-definite, i.e.

$$\partial x_i/\partial p_i + x_i\ \partial x_i/\partial m \leq 0.$$

Hanemann (1982d) has shown that for the two good case the adding-up property implies the homogeneity property, so that for this case one need only check that the negative semi-definite condition holds. However, this latter condition is nontrivial; its violation may cause anomolies to arise in the calculation of welfare measures. Violation of negative semi-definiteness conditions implies upward sloping compensated demand functions and meaningless welfare measures.

Exact Surplus Measures for Common Functional Forms

Closed form solutions to (1) are possible for several commonly used functional forms. The procedure discussed above and outlined in the Appendix 2.1 to this chapter has been used to derive parametric bivariate utility models consistent with tractable ordinary demand functions. In what follows, the results of this procedure when applied to the linear, semi-log, and log-linear demand functions are presented (for reference see Hanemann, 1979, 1980b, 1982b; Hausman 1981).

Consider the three specifications

$$x_1 = \alpha + \beta p_1/p_2 + \gamma y/p_2 \tag{2}$$

$$x_1 = \exp(\alpha + \beta p_1/p_2 + \gamma y/p_2) \tag{3}$$

$$x_1 = \exp(\alpha)\ (p_1/p_2)^{\beta}\ (y/p_2)^{\gamma} \tag{4}$$

where α, β, and γ are parameters, p_1 is the price of the good in question, p_2 is the price of the Hicksian bundle, and y is income. Henceforth p will designate normalized price, p_1/p_2, and m normalized income, y/p_2.

The expenditure functions (denoted m(p,u)) which result from integrating back from each of the above forms are presented in Table 2.1. Inverting the expenditure functions yields indirect utility functions, v(p,m), also presented in Table 2.1. It is also possible in the simple two good case to retrieve the bivariate direct utility function, utility as a function of goods rather than prices and income. For the simple two good case, the Marshallian demand function for x_1 together with the budget constraint can be solved for y/p_2 and p_1/p_2 as functions of x_1 and x_2. Substitution into the indirect utility function yields the direct one. Knowledge of the direct utility function implied by an estimated demand function is particularly useful as it provides insight into the properties of the preference structure implicitly assumed.[1]

The compensating and equivalent variations for price changes from p_1^o to p_1' can be derived by calculating the change in the relevant expenditure function when price changes.[2] Thus

$$CV = m(p^o, U^o) - m(p\ , U^o)$$

and

$$EV = m(p^o, U\ ) - m(p\ , U\ ),$$

where U' takes the value of the indirect utility function evaluated at p' and m°. The expressions for CV and EV as well as that for ordinary surplus, i.e. the Marshallian consumer surplus, are also recorded in Table 2.1.

Not all estimated demand functions corresponding to the functional forms in (2), (3) and (4) can be integrated back to well behaved (i.e. quasi-concave) utility functions. The negative semi-definiteness condition for these functions translates into restrictions on the functions' coefficients. These restrictions are given in Table 2.1. While frequently ignored, the conditions are critical. If, in a given empirical problem, estimated coefficients violate these conditions, then one can presume that the model is misspecified in some way. That is, the estimated coefficients imply an upward sloping compensated demand function and are therefore inconsistent with utility maximizing behavior.

Evaluating the Elimination of a Resource

The formulas in Table 2.1 presume interior solutions, i.e. x_1 and x_2 strictly greater than zero. Frequently, however, we are interested in evaluating situations when $x_1 = 0$. For example, we may wish to calculate the lost benefits associated with elimination of access to a resource. Alternatively the conditions at the axis may be important in assessing a change in a quality aspect of a good (more on this in Part II.)

Typically, economists have evaluated the losses associated with the elimination of a resource in the same way that they have evaluated the gains or losses of a price change. The price is simply assumed to increase sufficiently to drive demand to zero. This practice can generate anomolies, since resource elimination really involves a restriction on quantity rather than a *de facto* change in price. For many functional forms, the price which drives the Marshallian demand to zero is different from the price which drives the corresponding compensated demand to zero. When the two cut-off prices do coincide, it is generally because the cut-off price is infinite. An infinite cut-off price frequently (although not always) implies that an infinite sum is necessary to compensate for elimination of the good.

Consider first the linear demand function, an example of a form for which a finite cut-off price exists. If the Marshallian function is expressed as $x_1 = \alpha + \beta p + \gamma m$ then its cut-off price is $\tilde{p}^m = -(\alpha + \gamma m)/\beta$. The corresponding Hicksian demand is $x_1 = \gamma \exp(\gamma p)u - \beta/\gamma$ with a cut-off price $\tilde{p}^h = \frac{\ln\beta - 2\ln\gamma - \ln u}{\gamma}$. For purposes of comparison with the Marshallian demand curve, it is useful to substitute $V(p^o,m^o)$ for u in the expression for $\tilde{p}^h$ so that we identify the *particular* compensated curve which intersects the Marshallian demand at the initial point $(x(p^o,m^o),p^o)$. This gives us

$$\tilde{p}^h = \frac{\ln\beta - \ln\gamma - \ln(x^o + \beta/\gamma)}{\gamma} + p^o .$$

The difference between $\tilde{p}^h$ and $\tilde{p}^m$ is $(\ln\beta - \ln\gamma - \ln(X^o + \beta/\gamma))/\gamma - X^o/\beta$.

There is some ambiguity as to which $\tilde{P}$ should be used in calculating the compensating variation associated with the elimination of the resource. Thinking about the problem as one of a quantity not a price change suggests the question "How much compensation would leave you as well off if your access to the good were denied?" This implies a movement along the compensated demand function to its intersection with the axis. It is this latter interpretation which is advocated by Just, Hueth and Schmitz (1982), and which seems the most convincing.

The implication of the choice of $\tilde{p}$ in calculating CV is a potentially important one. All usual comparisons of CV, EV and OS are made on identical effective price changes. When considering the elimination of a resource, the usual relationship between CV, EV, and OS is now distorted. OS is the area behind the ordinary demand function between p^o and $\tilde{p}^m$ (the price which drives ordinary demand to zero). CV is the area behind the compensated demand curve which passes through p^o, but <u>not</u> between the same bounds as the ordinary samples. Instead we must integrate between p^o and $\tilde{p}^h$ (the price which drives the Hicksian demand to zero). EV must logically be defined as the area between p^o and $\tilde{p}^m$, behind that compensated demand curve which passes through $\tilde{p}^m$.

Because the bounds of integration for CV are not the same as for OS and EV, the usual relationship between the latter and former is destroyed and Willig's bounds no longer hold. Whether or not the difference is of practical significance depends on the relative sizes of the parameters and can only be determined empirically. Unfortunately the greater the difference between ordinary surplus and compensating variation, the greater the difference in the two CV measures.

For some functional forms, there exists no finite price at which demand is zero. This does not, however, mean that the area behind the respective demand curve is necessarily unbounded. In some cases, the <u>limit</u> of the demand for x_1 is zero as $p' \to \infty$ and thus, the area behind the demand curve converges to some finite value. In other cases the limit of x_1 does not equal zero as $p' \to \infty$, and the area behind the demand curve is infinite.

To understand this phenomena, one needs to consider the concept of essentiality. Marty equivalent definitions of essentiality exist but perhaps the most intuitive and descriptive is the following:

> A good, x_1, is essential if, given an initial consumption vector $(x_1^0, x_2^0, \ldots, x_n^0)$ there exists no subvector $(x_2', x_3', \ldots, x_n')$ such that $u(x_1^0, x_2^0, \ldots, x_n^0) = u(0, x_2', x_3', \ldots, x_n')$.

An equivalent definition is that there exists no finite sum which can compensate for the elimination of x_1. These definitions are both equivalent to the condition that for x_1 to be essential

$$\lim_{p_1 \to \infty} x_1^h(p,u) \neq 0$$

and for x_1 to be nonessential

$$\lim_{p_1 \to \infty} x_1^h(p,u) = 0.$$

It should be noted that these definitions are in terms of the compensated not the ordinary demand function. In fact, there is not a perfect correspondence between the limiting conditions for the compensated demands and those for the ordinary demands. There exist preference structures which imply ordinary demand functions which do converge but compensated functions which do not.

An interesting example for illustration is the general CES form for the direct utility function, $u = (x_1^\rho + x_2^\rho)^{1/\rho}$, which generates the following functions:

$$v = (p_1^{1-\sigma} + p_2^{1-\sigma})^{\frac{1}{\sigma-1}} y$$

$$m = (p_1^{1-\sigma} + p_2^{1-\sigma})^{\frac{1}{1-\sigma}} u$$

$$x^m = \frac{y/p_2}{\frac{p_1}{p_2} + (\frac{p_1}{p_2})^{\sigma}}$$

$$x^h = \frac{u(p_1^{1-\sigma} + p_2^{1-\sigma})^{\frac{\sigma}{1-\sigma}}}{p_1^{\sigma}}$$

where $\sigma = 1/(1-\rho)$.

Note that $\lim_{p_1 \to \infty} x^m = 0$ for all values of σ. Correspondingly $\lim_{p_1 \to \infty} x^h = 0$ and $\lim_{p_1 \to \infty} m(\cdot)$ is finite if $\sigma > 1$ but $\lim_{p_1 \to \infty} x^h \neq 0$ and $\lim_{p_1 \to \infty} m(\cdot) = \infty$, otherwise.

All of this is of importance not only in calculating losses associated with elimination of a resource, but also in assessing the relative merits of different functional forms. Essentiality is a property of preferences which may not be very applicable when dealing with recreational goods. It is difficult to conceive of a recreational experience which is indeed essential, i.e. its elimination would reduce utility to zero. Thus, functional forms which imply essentiality are probably poor choices.

In this light, let us examine the last two "popular" functional forms to see what they imply about the essentiality property. For the semi-log demand function, $x_1 = \exp(\alpha + \beta p + \gamma m)$, there is no finite price at which the demand for x_1 is zero. However, the limit of compensating variation is finite as $p_1 \to \infty$, and thus x_1 is non-essential.

For the log-linear demand function, $x_1 = e^{\alpha} p^{\beta} m^{\gamma}$, the price that drives the demand for x_1 to zero is also infinite. For relatively elastic demands compensating variation converges to a finite quantity as $p_1 \to \infty$. However, when $0 > \beta > -1$ the compensating variation associated with elimination of the resource is infinite. This implies that x_1 is an essential good.

Functional Form Comparison

While there are no previous studies where compensating variation measures are compared across functional form, there are some which document the potential differences in ordinary surplus estimates which arise when different functional forms are estimated on the same data and others which simply address the issue of choice among functional form in recreational demand models. In a study of warm water fishing in Georgia, Ziemer, Musser and Hill (1980) assessed the importance of the functional form on the size of ordinary consumer surplus estimates. They chose to consider linear, semi-log and quadratic forms and found average surplus per trip estimates of $80, $26 and $20 respectively. The researchers estimated a Box-Cox transformation to discriminate among the three functional forms and determined that the semi-log was preferable.[3]

Two other papers of note identified the semi-log function as most appropriate. Both papers addressed functional form in the context of the heteroskedasticity issue (a more detailed discussion of these papers can be found in Chapter 3). Vaughan, Russell and Hazilla (1982) tested for appropriate functional form and heteroskedasticity, simultaneously. They used the Lahiri-Egy estimator which is based on the Box-Cox transformation, but also incorporates a test for nonconstant variance. They concluded that both the linear heteroskedastic and linear homoskedastic models were inappropriate. The semi-log form which did not exhibit heteroskedasticity was found to be preferable. In a second paper Strong (1983a) compared the semi-log model with the linear model based on the mean squared error in predicting trips. She also found that the semi-log function performed better.

Another consideration of the functional form issue can be found in Smith's (1975b) analysis of visits to the Desolation Wilderness area in northern California. He examined the linear, semi-log and double-log functional forms for wilderness demand using the zonal approach with 64 origin zones from California, Nevada and Oregon. While the R^2 is not an appropriate test to compare specifications with different dependent variables, the linear model exhibited such a low R^2 that it was not considered further.

To try to establish more conclusively which functional form was more appropriate, Smith chose to use a method suggested by Pearsan which discriminates between non-nested competing regression models. Smith found that in his sample of wilderness recreators he was able to reject both the semi-log and the double-log functional forms based on this criteria. His conclusion that the travel cost model may be inappropriate for wilderness recreation modelling may be correct but is too extreme a conclusion to be supported by this analysis. Even if the Desolation Wilderness area is representative of other wilderness recreation problems, the alternatives tested in this study are by no means exhaustive. The functional forms chosen are but three among a vast array of choices. Additionally, Smith's poor statistical results could well be a reflection of other specification problems inherent in his conventionally designed zonal travel cost model. (See discussions in Chapters 3 and 4.)

Estimating a Flexible Form and Calculating Exact Welfare Measures

Each of the above studies was concerned with calculating ordinary surplus measures from commonly estimated functional forms using zonal data. These studies either implicitly assumed or explicitly demonstrated that consumer surplus estimates would differ depending on the choice of functional form. Not surprisingly, compensating (or equivalent) variation measures derived from different functional forms may also exhibit vast differences.

In the previous literature, the focus seems to have been one of identifying a means of choosing which of the popular functional forms was preferable. If it were possible to select one, then the exact welfare results of the previous section could be directly applied. Many of the articles appear to point to the semi-log as a desirable form, yet the evidence is far from conclusive and there is no reason to believe that the same form would necessarily be appropriate for all situations.

It would be far preferable to consider a wider array of functional forms than the three discussed above and to allow the data to choose among them. One way to access a slightly broader range of functional forms is to estimate a flexible form such as the Box-Cox transformation. However, Box-Cox forms do not in general integrate back to closed form expressions for the expenditure or indirect utility functions. A solution to this problem can be found in the recent work by Vartia (1983), among others, who demonstrates a means of obtaining extremely close approximations to compensating variation when exact measures are not possible. The procedure uses a third order numerical integration technique to obtain an approximate solution to the differential equation.

The Vartia algorithm, and others like it, is based on an intuitively appealing proposition. The ordinary and compensated demand curves are very close in the neighborhood of their intersection. The difference in the curves which occurs with a movement away from that intersection reflects an adjustment in consumption in response to additional (compensations in) income. Therefore, it should be possible to trace out the compensated curve, approximately, by a) starting at the intersection point of the two demand curves, b) considering a very small (incremental) change in price, c) calculating the approximate money compensation associated with that change, d) awarding that amount of income to the individual and then shifting his ordinary demand function, e) designating this new consumption level at the first price increment to be a point on the compensated demand function and f) starting the process once again with a new price increment. This procedure is described graphically in Figure 2.1.

The only step of any difficulty in this procedure is (c), calculating the approximate money compensation leaving utility unchanged, which is associated with the small price change. Of course this is the very problem we set out to solve, since this is the definition of variation. We can not calculate this number directly but we have information on the bounds of this compensation. The compensation for any price change will presumably be greater than (or equal to) zero and less than (or equal to) the total market value of the lost (or gained) consumption, $(p^1 - p^0)x^0$. The latter number is an upper bound which would equal the compensation if, for example, the given quantity of x consumed were essential and x had no substitutes. Approximation algorithms employ iterative techniques to calculate income adjustments using an interpolation of this upper bound as a starting point. While Vartia's procedure will handle systems of demands and multiple price changes, we describe heuristically the one equation, one price change case here.

The Vartia procedure requires the following initial information: the specific form of the ordinary demand function(s), the income level and the initial and final values of the price(s). To implement the procedure one must also choose the number of steps, N, one wishes to make in moving from the initial to the final price. The approximation will, in general, improve with more steps (and thus smaller increments) but rising computer costs and rounding error will eventually take their toll.

As pointed out above, the difficult task in the procedure is the calculation of the appropriate income compensation to accompany each price step.

This is accomplished through an iterative interpolation scheme of the following sort. For any given price step (p_j to p_{j+1}) define the first guess at the income compensation for that step by

$$\Delta y_1 = \frac{x(p_{j+1}, y_j) + x(p_j, y_j)}{2} * (p_{j+1} - p_j)$$

where p_{j+1} and p_j are the upper and lower bounds of the price step and y_j is the income associated with the starting point for this step (i.e. the income associated with the ordinary and compensated demand intersection as at point A in Figure 2.1). If this income adjustment were awarded then compensated demand would be $x(p_{j+1}, y_j + \Delta y_1)$, but we know that this would be an overadjustment, i.e. utility will have increased rather than been held constant. The second guess at the income adjustment will be based on the average of the two new consumption levels $x(p_{j+1}, y_j)$ which implies no compensation (and thus is a lower bound) and $x(p_{j+1}, y_j + \Delta y_1)$ which is based on too much compensation. Thus

$$\Delta y_2 = \frac{x(p_{j+1}, y_j + \Delta y_1) + x(p_{j+1}, y_j)}{2} * (p_{j+1} - p_j).$$

The iterative procedure progresses with each new guess at the income adjustment for this price increment equalling

$$\Delta y_k = \frac{x(p_{j+1}, y_j + \Delta y_{k-1}) + x(p_{j+1}, y_j, + \Delta y_{k-2})}{2} * (p_{j+1} - p_j),$$

until the Δy_k converge, i.e. $\Delta y_k - \Delta y_{k-1}$ < convergence criteria. Once the convergence criteria has been met at Δy_k, we shift to a new ordinary demand curve and a new point has been identified on the old compensated demand function, $x(p_{j+1}, y_{j+1})$ where $y_{j+1} = y_j + \Delta y_k$. The compensating variation of the total price change is approximated by summing Δy_k over all N price steps. This will be equal to $y_N - y_0$ in the above notation. A computer algorithm developed by Terrence P. Smith to implement Vartia's procedure is presented in Appendix 2.2.

The Vartia approximation was tested for a functional form for which exact compensating variation expressions exist. The Vartia measure improved with the number of steps chosen in the algorithm, but quickly came within a half of one percent of the true measure. Thus the approximation would seem to meet an acceptable tolerance criteria at low computing costs.

In what follows, we will demonstrate how this approximation procedure can be used with the Box-Cox transformation. The approach is equally applicable to other forms (flexible or not) for a single equation or system of equations. It should be noted, however, that the Vartia approximation does not circumvent either mathematical or economic integrability conditions. These conditions must hold for the results of the procedure to have meaning. The Vartia technique provides a close approximation to compensating and equivalent variation measures when no closed form solution to the differential equation in (1) exists or can easily be found.

An Illustration

To illustrate the application of this method for choosing functional form and calculating welfare measures, the Box-Cox transformation was estimated for a set of sportfishing data. The Box-Cox approach was chosen because of its wide familiarity and ease of estimation. However, as noted above, the procedure for deriving welfare measures is equally applicable to other less restrictive functional forms.

All individuals in the group took at least one trip of greater than 24 hours on a party/charter boat. This is a subset of a sample of 1383 sportfishermen who responded to a mail questionnaire asking details of their 1983 sportfishing activities in Southern California. A complete description of the data can be found in National Coalition for Marine Conservation (1985).

For purposes here, an individual's demand for party/charter trips (x) is considered to be a function of costs of the trip (c), income (y) and catch of target species (b).

Three models were estimated using the same data set. The first constrained the functional form to be linear, the second employed a semi-log function and the third used the more flexible Box-Cox transformation on the dependent variable so that the regression took the form:

$$\frac{x^{\lambda} - 1}{A} = \beta' z,$$

where x is trips and z is the vector of explanatory variables. The parameters to be estimated included the usual coefficients (the β vector) and the Box-Cox parameter, y .

The linear model produced the following estimated equation (t-statistics in parentheses):

$$x = \underset{(6.38)}{6.57} - \underset{(-4.08)}{.0045}\, c + \underset{(1.67)}{.0000189}\, y + \underset{(1.54)}{.179}\, b \qquad R^2=.17.$$

In contrast, the estimated semi-log demand function looked like

$$\ln x = \underset{(10.66)}{1.66} - \underset{(-6.09)}{.00102}\, c + \underset{(1.96)}{.0000034}\, y + \underset{(1.85)}{.0325}\, b \qquad r^2=.29.$$

Finally, the Box-Cox estimation produced the following results

$$\frac{x^{.14} - 1}{.14} = \underset{(9.79)}{1.91} - \underset{(-5.83)}{.0012}\, C + \underset{(1.93)}{.0000042}\, y + \underset{(1.82)}{.04}\, b \quad R^2=27.$$

In this particular example, the Box-Cox produced a λ close to zero. This result is somewhat consistent with the fact that the result of the semi-log function appear superior to that of the linear equation. This should not be construed as a general endorsement of the semi-log demand function, since other applications of the Box-Cox transformation have provided a wide range of values for λ.

In Table 2.2, the results of this experiment are presented. The estimated coefficients from the linear and semi-log models have been used in conjunction with the expressions in Table 2.1 to calculate estimates of ordinary surplus, and compensating and equivalent variation. The computation process is explained in Appendix 2.1. The Vartia algorithm has been used to obtain "approximate" measures of compensating and equivalent variation and ordinary surplus for the Box-Cox model. The algorithm is presented in Appendix 2.2.

Some important points are worth noting. First, these welfare measures seem large. It should be remembered that the sample included only those who took longer than one day trips and are therefore likely to be rather wealthy individuals. In fact, the mean income of this group is $58,000. Additionally, there are reasons why welfare measures calculated from estimated coefficients may produce overestimates of the true values. These considerations will be discussed in Chapter 5.

The important point for consideration here is that if one were arbitrarily to choose between the linear and semi-log specification in estimating the demand function, widely divergent benefit estimates would emerge. In the case above there is only a 3 to 5% difference across welfare measures (CV, EV, 0S) for any one functional form, but a 16 to 19% difference between the two most commonly used functional forms. The Box-Cox transformation offers a means of choosing among a continuous range of functional forms. In the example above, it seems to support the semi-log function. In other cases we have tried, where neither the linear nor the semi-log results appear superior, the Box-Cox analysis often selects an i significantly different from either zero or one. Then the Varita routine is necessary to calculate compensating and equivalent variation approximations.

While definitional differences in welfare measures will be of greater concern in problems with larger income elasticities (Willig, 1976), bounds on these differences are well developed, at least for simple models. The potential differences from functional form, however, may not be so well appreciated.

Table 2.2

Welfare Estimates

Calculated from Different Functional Forms

(annual average estimates for a sample of Southern California sportfishermen)

	Functional Form		
	Linear	Box-Cox	Semi-log
Compensating Variation	8339	6950	6999
Ordinary Surplus	8042	6812	6877
Equivalent Variation	7899	6779	6763

FOOTNOTES TO CHAPTER 2

1 LaFrance and Hanemann (1985) describe the process of obtaining direct utility functions from estimated demand functions for systems of demand equations.

2 There is some disagreement in the literature as to the precise form of the compensating and equivalent variation expression. All agree that compensating and equivalent variation must be of the same sign. However, differences of opinion exist as to whether the variational measures have the same or the opposite sign as the utility change. Here we adhere to the convention used by Just, Hueth and Schmitz (1982) which seems most closely aligned with the original description of Hicks. Compensating and equivalent variation are positive (negative) for price changes which generate increases (decreases) in utility.

3 The Box-Cox functional form is a limited flexible functional form developed by Box and Cox (1962) using a transformation of the dependent variable. The transformation is defined as

$$y^{(\lambda)} = y\frac{\lambda_{-1}}{\lambda}$$

so that the regression equation can be written as

$$y^{(\lambda)} = x\beta + \varepsilon.$$

The interesting feature of the Box-Cox transformation is that when λ takes the value of 1, the above expression is just a linear function of y in x. When $\lambda = 0$, $y^{(\lambda)}$ is not strictly defined but $y^{(\lambda)}$ is continuous at

$$\lambda = 0 \text{ since } \lim_{\lambda \to 0} y\frac{\lambda-1}{\lambda} = \log y.$$

Box Cox models are estimated by maximizing the maximum liklihood function with respect to the β's and the λ. Thus the functional form is not strictly imposed and one can establish confidence intervals on λ which allows testing hypothesis about functional form.

4 The Lahiri-Egy estimation is an extension of the Box-Cox transformation. It introduces an additional parameter which allows one to test for the presence of heteroskedasticity jointly with functional form. The estimator assumes that the error in the model

$$y^{(\lambda)} = x\beta + \varepsilon$$

is distributed such that the expected value of ε_i is $Z_i^{\delta/2}\, u_i$ where u_i is normal with mean, 0 and variance, σ^2, and Z_i is some variable which varies over observations (and is likely related to one of the x's). The variance of ε_i is then $\sigma^2\, Z_i^{\delta}$. Consequently, if $\delta = 0$ then the variance of ε is homoskedastic; if $\delta \neq 0$ then there is heteroskedasticity in the model.

Thus the Lahiri-Egy estimator uses a maximum likelihood procedure to estimate the Box-Cox transformation under conditions of potential heteroskedasticity. The likelihood function is maximized with respect to β, λ, δ, and σ^2.

APPENDIX 2.1

DERIVATION OF SOME UTILITY THEORETIC MEASURES FROM TWO GOOD DEMAND SYSTEMS

As Hausman has so bluntly, and some what unkindly, suggested

> From an estimate of the demand curve, we can derive a measure of the exact consumer's surplus, whether it is the compensating variation, equivalent variation, or some measure of utility change. No approximation is involved. While this result has been known for a long time by economic theorists, applied economists have only a limited awareness of its application.

a) Following Hausman's example, we can begin with a demand function where quantity is a function of price and income both deflated by the price of the other good. Letting p and m stand for the "deflated" price and income, and using Roy's identity then

$$x_1 = f(p,m) = \frac{-\partial v/\partial p}{\partial v/\partial m} \, . \tag{A1}$$

Now, this partial differential equation must be solved. Hausman uses the method of "characteristic curves". Using the notion of compensating variation, one can consider paths (designated by t) of price changes and accompanying income changes, such that utility is left unchanged as in the following:

$$\frac{\partial v(p(t),m(t))}{\partial p(t)} \; \frac{dp}{dt} = \frac{-\partial v(p(t),m(t))}{\partial m(t)} \; \frac{dm}{dt} \, . \tag{A2}$$

Since $x_1 = -\frac{\partial v/\partial p}{\partial v/\partial m}$, then (A2) can be re-expressed as

$$x_1 = -\frac{\partial v/\partial p}{\partial v/\partial m} = \frac{dm/dt}{dp/dt} = \frac{dm}{dp} \, . \tag{A3}$$

This gives an ordinary differential equation which in many cases can be solved with fairly standard techniques. As Hausman shows, the solution to the differential equation

$$dm/dp = \alpha + \beta p + \gamma m \qquad \text{(linear case)}$$

is

$$m(p) = ce^{\gamma p} - \frac{1}{\gamma}(\beta p + \frac{\beta}{\gamma} + \alpha).$$

The only confusion is in dealing with, c, the constant of integration. Clearly c will not be a function of any of the parameters in the demand function but it will certainly be a function of the utility level. In a sense it doesn't matter what function as long as it is increasing and monotonic, since we have no way of measuring or interpreting absolute levels of utility. As a consequence Hausman simply substitutes U^0 for c which is fine as long as everyone uses u^0 only for ordinal comparisons and does not try to interpret the absolute level of u^0. In some circumstances interpretting c as equal to u^0 will lead to confusion because utility will appear to be negative. There is no fundamental problem, however, as long as $\partial c/\partial u > 0$, the scaling of u^0 is arbitrary.

b) Once the expenditure function is obtained from solving the differential equations the indirect utility function is usually easy to obtain by solving $m(p,u^0)$ for utility giving $u = v(p,m)$. For some demand functions, it is easier to integrate back to the indirect utility function first, in which case the expenditure function is obtained by solving $v(p,m)$ for income as a function of utility and price. The three examples below demonstrate how straightforward this can be when there are closed form expressions for both indirect utility and expenditure functions:

$$\text{(A4)}\quad m = \exp(\gamma p)u^0 - \frac{1}{\gamma}(\beta p + \frac{\beta}{\gamma} + \alpha) \Rightarrow u = \exp(-\gamma p)(m + \frac{1}{\gamma}(\beta p + \alpha + \beta/\gamma)) \qquad \text{(linear)}$$

$$\text{(A5)}\quad m = -\frac{1}{\gamma}\ln(-\gamma u^0 - \frac{\gamma}{\beta}\exp(\beta p + \alpha)) \Rightarrow u = \frac{-\exp(-\gamma m)}{\gamma} - \frac{\exp(\beta p + \alpha)}{\beta} \qquad \text{(semi-log)}$$

$$\text{(A6)}\quad m = [(1-\gamma)(u^0 + \frac{e^{\alpha}p^{1+\beta}}{1+\beta})]^{\frac{1}{1-\gamma}} \Rightarrow u = \frac{-e^{\alpha}p^{1+\beta}}{1+\beta} + \frac{m^{1-\gamma}}{1-\gamma} \qquad \text{(log-linear)}$$

c) Once the expenditure function is derived, the Hicksian demand function together with compensating and equivalent variation measures are of course quite accessible:

(A7) $$x_1^H = \frac{\partial m(p,u^0)}{\partial p}.$$

Compensating and equivalent variations are, by definition

(A8) $$C = m(p^0,u^0) - m(p^1,u^0) \quad = \quad m^0 - m(p^1,u^0)$$

(A9) $$E = m(p^0,u^1) - m(p^1,u^1) \quad = \quad m(p^0,u^1) - m^0.$$

Thus they can be solved for directly from the expenditure function. (Note that Hausman defines C and E with reversed sign. The above definition is more in keeping with the original Hicksian definitions and has the property that the sign of C = sign of the welfare change associated with the price change.) To simplify expressions and to obtain actual values for C and E, $u^0 = v(p^0,m^0)$ and u's $v^1(p^1,m^\circ)$ must be evaluated.

An example is presented for the linear demand, where

$$m = \exp(\gamma p)u^0 - \frac{1}{\gamma}(\alpha + \beta p + \beta/\gamma) ;$$

$$c = \exp(\gamma p^0)u^0 - \frac{1}{\gamma}(\alpha + \beta p^0 + \beta/\gamma) - \exp(\gamma p^1)u^0 + \frac{1}{\gamma}(\alpha + \beta p^1 + \frac{\beta}{\gamma})$$

$$= \exp(\gamma p^0)\ \exp(-\gamma p^0)(\frac{\alpha + \beta p^0 + \gamma m}{\gamma} + \frac{\beta}{\gamma 2}) - \frac{1}{\gamma}(\alpha + \beta p^0 + \frac{\beta}{\gamma})$$
$$- \exp(\gamma p^1)\exp(\gamma p^0)\ (\frac{\alpha + \beta p^0 + \gamma m}{\gamma} + \frac{\beta}{\gamma 2}) + \frac{1}{\gamma}(\alpha + \beta p^1 + \frac{\beta}{\gamma})$$

$$= \frac{x^0}{\gamma} + \frac{\beta}{\gamma 2} - \frac{x^0}{\gamma} + \frac{\beta}{\gamma 2} + m - \exp[\gamma(p^1 - p^0)](\frac{x^0}{\gamma} + \frac{\beta}{\gamma 2}) + (\frac{x^1}{\gamma} + \frac{\beta}{\gamma 2}) - m$$

$$= (\frac{x^1}{\gamma} + \frac{\beta}{\gamma 2}) - \exp[\gamma(p^1 - p^0)](\frac{x^0}{\gamma} + \frac{\beta}{\gamma 2}) .$$

d) The one remaining function of interest is the direct utility function, $u(x_1,x_2)$, which is of interest because it best portrays the properties of the preference function being assumed. The task is to convert a utility function in (normalized price) and income into a utility function in terms of x_1 and x_2. Since we have two functions which relate the x's with p and m, i.e. the Marshallian demand function for x_1 and the budget constraint, it is conceptually possible to make the transformation. One must first solve $x_1 = f(p,m)$ and $m = px_1 + x_2$ for $p = g(x_1,x_2)$ and $m = h(x_1,x_2)$, and then the substitution into the indirect utility function is straightforward.

As an example, consider the linear case where

$$x_1 = \alpha + \beta p + \gamma m$$

$$m = px_1 + x_2.$$

then

$$p = \frac{x_1 - \alpha - \gamma m}{\beta} = \frac{x_1 - \alpha - \gamma p x_1 - \gamma x_2}{\beta}$$

$$\Rightarrow \quad p = \frac{x_1 - \alpha - \gamma x_2}{\beta + \gamma x_1};$$

$$m = px_1 + x_2 = \frac{x_1^2 - \alpha x_1 - \gamma x_2 x_1 + \beta x_2 + \gamma x_1 x_2}{\beta + \gamma x_1}$$

$$= \frac{x_1^2 - \alpha x_1 + \beta x_2}{\beta + \gamma x_1}.$$

By substitution

$$u = \exp(-\gamma p)(m + \frac{1}{\gamma}(\beta p + \alpha + \frac{\beta}{\gamma}) = \exp(\frac{-\gamma x_1 + \gamma\alpha + \gamma^2 x_2}{\beta + \gamma x_1})\frac{(\beta + \gamma x_1)^2}{\gamma^2(\beta + \gamma x_1)}$$

$$= \frac{\gamma x_1 + \beta}{\gamma^2} \exp[\frac{\gamma(\alpha + \gamma x_2 - x_1)}{\beta + \gamma x_1}]$$

APPENDIX 2.2

COMPUTER ALGORITHM FOR OBTAINING COMPENSATING AND
EQUIVALENT VARIATION MEASURES FROM ESTIMATEO
MARSHALLIAN DEMAND FUNCTIONS*

```
***************************************************************************+
* A COMPUTER ALGORITHM FOR APPROXIMATING CV AND EV FROM ESTIMATED DEMAND
* FUNCTIONS. CALCULATES NUMERICAL SOLUTION FOR SYSTEM OF DIFFERENTIAL EQUATION:
*
* BASED ON ALGORITHM BY VARTIA (ECONOMETRICA, VOL 51, NO 1, 1983)
*  WRITTEN IN VS/FORTRAN (FORTRAN 77 - ANSI(1978))
*       T. P. SMITH, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD
***************************************************************************+
***************************************************************************+
* PROGRAM REQUIRES STATEMENT FUNCTIONS (IN LINES 10-200) WHICH CORRESPOND
*  TO MARSHALLIAN DEMAND SYSTEM. FOR EXAMPLE, IF X1=BO+BI*P+B2*Y AND BO=2,
* B1=-5, B2=6, THEN THE FOLLOWING SHOULD BE ENTERED
*     10        X1(P1,INCOME)=2-5*P1+6*INCOME
*  A SYSTEM OF UP TO 20 EQUATIONS CAN BE ENTERED IN THIS WAY. THE FUNCTION
* CALLS THROUGHOUT THE PROGRAM MUST BE MODIFIED TO REFLECT THE APPROPRIATE
* ARGUMENT LIST FOR THE FUNCTIONS BEING USED. THE # OF EQUATIONS AND THE
* # OF STEPS FOR THE PRICE PATH MUST BE SUPPLIED.  AVOID A LARGE # OF STEPS
+  (>500) AS ROUNDING ERRORS CAN BECOME SERIOUS.
+  SAMPLE PROGRAM BELOW DEMONSTRATES TWO GOOD, ONE PRICE CHANGE CASE.
***************************************************************************+
      DOUBLE  PRECISION  P(20,500),Y,XC(20),INCOME,P1,P2,P3,P4,P5,P6,
     *P7,P8,P9,P1O,P11,P12,P13,P14,P15,P16,P17,P18,P19,P20,X1,X2,X3
     *X4,X5,X6,X7,X8,X9,XIO,X11,X12,X13,X14,X15,X16,X17,X18,X19,X20,
     *PSTEP(20),XT(20),TERM(20),DIFF,EPS,SUM+NEWY,YO
********************            STATEMENT  FUNCTIONS          ******************"
 10  X1(P1,INCOME)=EXP(3.56-.019*P1-. 027*INCOME+.00026*PI*INCOME)
*2O  x2(Pl,P2,INCOME)=(P1/P2)*(INCOME/(Pl/P2))
*3O  ETC.
********************            CONVERGENCE  CRITERION        *****************
      EPS=0.000I
*********ii*********                 PROBLEM  SIZE           ******************
      WRITE (6,1)
```

* This algorithm was developed by Terrence P. Smith, Department of Agricultural and Resource Economics, University of Maryland, College Park, Maryland.

```
1     FORMAT (' ENTER THE # OF EQUATIONS IN THE SYSTEM',/,
     * ' AND THE # OF STEPS FOR THE PRICE PATH')
      READ (5,*) NEQ,N
      WRITE (6,2)
2     FORMAT (' SPECIFY THE INITIAL AND FINAL VALUES FOR EACH',/,
     * ' PRICE, IN ORDER.  IF A PRICE DOESNT CHANGE, SPECIFY',/,
     * ' SAME INITIAL AND FINAL PRICE.')
      READ (5,*) ((P(I,1),P(I,N)),I=1NEQ)
      WRITE (6,3) ((I,P(I,1),P(I,N)),I=1NEQ),
3     FORMAT (' INITIAL PRICE  FINAL PRICE ',/,20(1H,I2,2F10.4,/))
      WRITE (6,4)
4     FORMAT (' NOW ENTER THE INCONE LEVEL')
      READ (5,*) YO
******************* CALCULATE THE PRICE STEPS AND PATHS ***********
      DO 1000 I=1,NEQ
      PSTEP(I)=(P(I,N)-P(I,1)/N
      DO 1000 J=2,N-1
      P(I,J)=P(I,J-1)+PSTEP(I)
1000  CONTINUE
******************* CALCULATE THE INITIAL VALUES ************
      DO 2000 I=1,NEQ
      IF (I.EQ.1) XC(I)=X1(P(1,1),Y0)
*     IF (I.EQ.2) XC(I)=X2(P(1,1),P(2,1),Y0)
*     ETC.
2000  CONTINUE
******************* ALGORITHM ***********
      ITIMES=0
      Y=Y0
      DO 3000 J=2,N
500   ITIMES=ITIMES+1
      OLDY=Y
      DO 4000 I=1,NEQ
      IF (I.EQ.1) XT(I)=X1(P(1,J),Y)
*     IF (I.EQ.2) XT(I)=X2(P(1,J),P(2,J),Y)
*     ETC.
      TERM(I)=((XT(I)+XC(I))/2)*PSTEP(I)
4000  SUM=SUM+TERM(I)
      NEWY=SUM+YO
      SUM=O
      Y=NEWY
      IF (ITIMES.EQ.500) STOP 'ENDLESS LOOP - NOT CONVERGING'
      DO 5000 I=I,NEQ
5000  XC(I)=XT(I)
      IF (DABS(NEWY-OLDY).GT.EPS) GO TO 500
      ITIMES=O
      YO=NEWY
3000  CONTINUE
      WRITE (6,5)
      WRITE (6,6) (XC(I),I=1,NEQ),Y
5     FORMAT (IHO, 'COMPENSATED DEMANDS', 13X, 'COMPENSATED INCOME')
6     FORMAT (1H ,5X,FIO.4,17X,F1O.4)
      STOP
      END
```

CHAPTER 3

AGGREGATION ISSUES: THE CHOICE AMONG ESTIMATION APPROACHES*

Our ultimate use of the recreational demand model is to derive aggregate welfare measures of the effects of environmental changes. However, the means by which these aggregate measures should be devised depends upon the level of aggregation of observations and the treatment of users and nonusers in the estimation stage. Thus, the appropriate aggregation of welfare measures depends very much on the initial decisions as to the types of observations used and the general sampling strategy employed.

Problems of aggregation plague applications of macroeconomics. The theory is derived from postulates of individual behavior, yet data is often more readily accessible in an aggregate form. In many types of microeconomic problems, market data is so much easier to obtain that rarely are cross sectional, panel data used. However, in recreational demand studies, where markets do not usually exist, survey techniques are necessary to generate data. Even in such surveys, however, data are often collected in aggregated form (by zone of residence). To many, the travel cost method is, in fact, synonymous with the "zonal approach", which employs visit rates per zone of origin as the dependent variable and values for explanatory variables which represent averages for each zone.

In its current state, the travel cost approach to valuing nonmarket benefits is the product of two legacies. One dates back to Harold Hotelling's extraordinary suggestion for estimating recreational demand. It has become intimately linked to the zonal approach and dependent on the concept of average behavior. The other legacy is the axioms of applied welfare economics which provide defensible means of developing benefit

* This Chapter is the work of Kenneth E. McConnell, Agricultural and Resource Economics Development, U. of Maryland, and Catherine Kling, Economics, U. of Maryland.

measures based on individual behavior. The two come in conflict over this issue which we broadly define as aggregation.

This chapter explores the relationship between the traditional zonal approach and a model based on individual behavior. A central theme in this discussion is the treatment of both recreational participants and nonparticipants. The implications for estimation and benefit calculations are discussed.

A Review of Past Literature

Before addressing the issues anew, it is useful to put in perspective the various discussions of aggregation problems found in the existing literature. The term “aggregation” has been applied in what we shall call the “national benefits” literature. These types of studies attempt to value widespread improvements in water quality due to changes in national environmental regulations. In this literature, the “aggregation problem” involves estimating benefits over a vast number of widely divergent water bodies, geographical regions, and recreational users. Vaughan and Russell have developed methods to evaluate comprehensive policy changes in this context (see, for example, Vaughan and Russell, 1981 and 1982; Russell and Vaughan, 1982). Perfecting these methods for obtaining approximate “value per user day” figures is of considerable importance and is being pursued under another EPA Cooperative Agreement.

The research reported here, however, is not designed to address these issues. The aggregation issues in question in this study are those which arise in all studies which attempt to use travel cost (or its more general form - household production) models to evaluate benefits to all individuals affected by an environmental change. The following brief review offers a menu of the problems which have been raised concerning aggregation within the context of the zonal and individual observation approaches to the travel cost method.

1. The Zonal Approach

Travel cost models that employ the zonal approach generally regress visits per capita in each zone of residence on the travel cost from the associated zone to the resource site and on other explanatory variables. The literature on these zonal models has addressed two types of problems. The appropriate size and definition of the zones and heteroskedasticity problems in estimation.

Sutherland (1982b) questioned the degree to which the size of the zones affected demand and benefit estimates and whether it was more appropriate to use concentric zones or population centroids. He estimated demand curves for boating using ten and twenty mile wide concentric zones as well as twenty population centroids. The study revealed larger consumer surplus estimates when concentric zones were used as compared to population centroids, suggesting that benefit estimates obtained from a travel cost model will be sensitive to the zone definition. However, Sutherland lamented the absence of clear criteria for choosing either population centroids or concentric zones.

In a recent paper, Wetzstein and McNeely (1980) discussed a related issue of aggregating observations. They argued that if it is indeed necessary to use aggregate data (i.e. zonal rather than individual observations), it is more efficient to aggregate the observations by similar travel costs rather than by the more traditional method of similar travel distances to determine zones. Aggregating the zones by travel cost would provide "a more efficient estimate of the coefficient associated with cost and thus improve the confidence in the value of the coefficient" (p. 798).

Wetzstein and McNeely estimated demand equations for ski areas under the two alternative aggregation schemes. When the data were aggregated by costs, both the distance and cost coefficients were significantly different from zero. However when the data were aggregated by distance, only the distance coefficient was significant. The paper suggests that estimated coefficients, and thus benefit estimates, may be highly sensitive to variation in explanatory variables within zones.

The final issue that has arisen concerning the determination of zones has to do with the spatial limits of the travel cost model. Smith and Kopp (1980) pointed out that including zones far from the site being valued will likely violate some basic assumptions implicit in the travel cost model. As the distance between origin zone and site increases, it is less likely that the primary purpose of the trip is to visit the site in question. It is also less likely that the amount of time spent on site and the form of transportation will remain constant. Smith and Kopp proposed the use of a statistical test to determine which zones should be included in the model and which should not. This test was developed by Brown, Durbin and Evans (1975) and is based on the fact that observations inconsistent with the assumptions of the travel cost model will produce nonrandom errors.

Smith and Kopp used 1972 United States Forest Service data on visitors to the Ventana area to illustrate the impact that the spatial limits of the travel cost model can have on benefit estimates. They had information on visitors from 100 zones encompassing 38 states. Applications of the Brown, Durbin and Evans procedure suggested that a spatial limit to the model could be established at a distance of about 675 miles from the site. The estimated per trip consumer surplus lost if the area were destroyed was $14.80 when all observations were included, but only $5.28 when the apparent spatial limits of the model were respected. Thus the definition of zones and the limitation of the number of zones are important issues and can have a significant impact on the size of benefit measures.

Another issue that has arisen in applying the zonal travel cost model concerns possible heteroskedasticity in the error term. This issue has been integrally related to the assumed functional form of the demand equation. Bowes and Loomis (1980) were among the first to warn of the potential heteroskedasticity problem which zonal data may create. When the defined zones encompass different size populations, the variance of the dependent variable, average number of trips in each zone, will vary with zones. If the variance of each individual's visitation rate is the same, i.e. $Var(v_{ij}) = \sigma^2$ for all individuals i in all zones j, then the variance of the mean visits per capita from zone j wIl1 be $Var(\Sigma_i v_{ij}/N_j) = \sigma^2/\surd N_j$ where N_j is zone j's population. This is a classic heterosskedasticity problem for which the correction procedures are well understood. One simply needs to weight all variables by the square root of the zone's population.

To illustrate the potential importance of this correction, Bowes and Loomis estimated a linear demand equation for per capita trips down a section of the Colorado River in Utah. Using the unweighed OLS estimates, total benefits were calculated as $77,728. When weighted observations were used to correct for the apparent heteroskedasticlty, only $24,073 in benefits could be attributed to the users of the Westwater Canyon.

Another possible source of nonconstant variance is suggested by Christianson and Price (1982). They argue that the variance in individual visitation rates is not likely to be constant across zones. Individuals located at different distances from the site will exhibit different participation rates and can be expected to have different individual variances. The source of heteroskedasticity is the unequal visit rates across zones. If both types of heteroskedasticity exist, the authors suggest that the proper weighting scheme would be $(N_j/E(V_j))^{1/2}$ where N_j is again the population in zone j and V_j is mean visitation rate per capita in zone j. This procedure causes the dependent variable to appear on the right hand side of the equation and thus would seem to generate further statistical problems.

In her response to Bowes and Loomis, Strong (1983b) made the case for the use of a nonlinear function (specifically the semilog form) as an alternative to the Bowes and Loomis correction for heteroskedasticity. Linear and semilog demand equations for steelhead fishing were estimated using data from zones around twenty-one rivers in Oregon, and a Goldfeld-Quandt test was employed to test for the existence and size of heteroskedasticity. The semilog model did not require a heteroskedasticity correction, but the linear model did. After correcting the linear model for heteroskedasticity (applying the appropriate weights), this model was compared to the semilog model by the mean squared error in predicting trips. The semilog form performed better than the corrected linear model in this test.

Vaughan, Russell and Hazilla (1982), in another comment on the Bowes and Loomis article, argued that an alternative to assuming a linear demand equation and heteroskedasticity is to test for both in the data rather than impose them as assumptions. To do this, they tested the Bowes and Loomis data for appropriate functional form and heteroskedasticity simultaneously by applying the Lahiri-Egy estimator which utilizes a maximum likelihood procedure to estimate the appropriate functional form with a Box-Cox transformation under conditions of potential heteroskedasticity. As a result of this procedure, they were able to reject the linear homoskedastic and the linear heteroskedastic models. The appropriate functional form for the data appeared to be nonlinear and with a nonlinear form heteroskedasticity appeared not to be a concern. The benefit estimate obtained with a semilog functional form (and no heteroskedasticity correction, since none was warranted) was only $14,000 as compared to the Bowes and Loomis estimate of almost twice the size. Vaughan et al. concluded from their analysis that the heteroskedasticity issue can not be separated from the choice of appropriate functional form and that it is likely that a non-linear specification is superior to a linear one.

In their study of partyboat fishing in California, Huppert and Thomson (1984) suggested another cause of heteroskedasticity that can not be mitigated with the semilog functional form. They argued that, in practice, the sampling scheme used to collect data for a travel cost model may give rise to heteroskedasticity. The semilog transformation suggested by Vaughan et al. and Strong will not eliminate the problem, unless the number of visitors surveyed from each zone is the same.

In their view, heteroskedasticity arises from the construction of the dependent variable from sample data. The trips per capita variable is calculated as $t_j = n_j t/p_j n$ where n_j = number of respondents sampled at the site from zone j, n = total number of respondents sampled at the site, t = total number of trips made to the site in 1979, and p_j = population in zone j. They argued that it is only n_j, the number of sampled respondents from zone j, that is random and that n can be thought of as a binomial variate since it is equivalent to the number of "successes" in n drawings. The variance formula is then $S^2 = n(\Pi_j)(1 - \Pi_j)$ where Π_j is the probability that an angler sampled will be from zone j. The variance for t_j is then $(t/np_j)^2 S^2$ and thus varies with zone. On the basis of this variance formula, Huppert and Thomson concluded that "variance due to sampling error depends inversely upon both sample size and zonal population" (p. 8). The authors also showed that the use of the semilog transformation would not eliminate this heteroskedasticity.

The discussions of the zonal approach in the literature have focused attention on practical or, perhaps more correctly, statistical problems which zonal aggregation may generate. By using zonal data, researchers are more likely to encounter multicollinearity and heteroskedasticity problems. Additionally, they are likely to lose precision in estimates whenever zones lack homogeneity and explanatory variables exhibit large variability within zones.

2. The Individual Observations Approach

The initial argument to use individual observations instead of zonal averages in the travel cost model can be traced to Brown and Nawas (1973) who sought to combat multicollinearity difficulties arising from more aggregated data. They wished to include the opportunity cost of time in travel cost demand models but found that since zonal money and time costs were likely to be highly correlated, multicollinearity became a serious problem. Brown and Nawas suggested using observations on individuals rather than grouped or averaged data as a solution. The authors offered an illustration on a data set consisting of 248 big game hunters in the northeast area of Oregon. In a model including money cost and distance (as a surrogate for time), the coefficient on money costs was significantly different from zero only when the model was estimated on individual observations.

Some years later, Brown, Sorhus, Chou-Yang and Richards (1983) reversed this position on the zonal versus individual observation question with the following argument. "The problem with fitting a travel cost-based outdoor recreational demand function to unadjusted individual observations is that such a procedure does not properly account for cases in which a lower percentage of the more distant population zones participates in the recreational activity. In such cases, a biased estimate of the travel cost coefficient results" (p. 154). The fact that more individuals choose not to participate from more distant zones holds important information for the researcher, and if such information is ignored, bias is likely to result. Zonal data implicitly incorporates this information, in a way, by using trips per capita. Brown et al. suggested that one might use individual observations without losing important participation data by transforming the left hand side variable to individual visits per capita (i.e. the dependent variable would be defined as visits by individual i in zone j/population in zone j).

While detailed discussion awaits the subsequent section of this chapter, the underlying problem here is one of truncated or censored samples. A few authors have attempted to deal with the problem of participation rates (numbers of participants versus nonparticipants) using econometric techniques designed to handle this type of phenomenon. Wetzstein and Ziemer (1982) illustrated Olsen's method of correcting for the bias introduced by the use of a truncated sample with permit data for Dome Land and Yosemite wilderness areas in 1972-1975. The Olsen method is a diagnostic tool which can determine the relative importance of the bias associated with omitting non-participants from a sample. It also offers an approximate correction for this bias using OLS parameter estimates. The impact of the truncation on the parameter values is determined by comparing the unadulterated OLS parameter estimates with the "Olsen" estimates.

The OLS and Olsen regression models were estimated for Yosemite and Dome Land. The Olsen correction was found to have a smaller influence on the Yosemite data than on the Dome Land data based on similarity of the Olsen estimates to the standard OLS estimates. This result is consistent with the underlying theory, since more zero visitor days were observed from Dome Land than from Yosemite. The authors also compared the OLS to the Olsen estimates based on forecast performance through the use of root-mean-square-error, mean error, and mean absolute error determined from predicted and observed visits in 1975. Again, the Dome Land 0LSestimates fared less well than the Yosemite OLS estimates as compared to the Olsen estimates, and the authors concluded that the severity of the bias is dependent on the nature of the data set.

Desvousges, Smith and McGivney (19831, recognizing the problem inherent in a sample which only included observations on the behavior of participants, also employed Olsen's method to evaluate the importance of the bias introduced by the omission of nonparticipants. They found that for several of their sites this truncation greatly biased their results. To compensate for the bias in their final model, they chose to use two samples, one which included all of the sites and one which omitted those sites that exhibited large biases from the effects of nonparticipants.

Models of Individual Behavior and Their Implications for Estimation

The controversy in travel cost literature surrounding the use of zonal vs. individual data focuses principally on data oriented problems. The zonal approach may be particularly susceptible to multicollinearity and heteroskedasticity. However, individual observations are expensive to collect and may be more vulnerable to severe errors in measurement. Discussions of substantive conceptual differences in the two approaches have been less frequent and less well developed. Recent work leaves one with the vague impression that welfare measures may be more difficult to define in the zonal approach but that, in some way, this approach better handles the problem of nonparticipants.

It is useful at this point to sort out some of these issues. One of the difficult problems in calculating total welfare changes as William Brown has pointed out, is accounting for the participation rate in the population. It turns out that this consideration plays an important role in the estimation stage as well as in the welfare calculations. Nonetheless, the proper perspective is still to think of the problem in terms of the individual. Throughout this report we have argued that the assumptions implicit in the estimation of any recreational demand model must be consistent with logical models of individual behavior. To model individual recreational demand adequately, one must allow an individual to choose not to participate. That is, a model of behavior must accommodate both positive and zero levels of demand. In what follows, a standard model of individual behavior which allows for zero levels of demand is presented and its implications for estimation using individual observations are explored.

The problem can be described as follows. For any recreational site, groups of sites, or activities, there will be many nonusers in the population. While corner solutions of this sort ($x = 0$ for some individuals for some goods) can be handled deftly in abstract models, they present complications for econometric estimation. These complications, and the biases resulting from ignoring the problem, are proportional to the rate of nonparticipation in the problem. Unfortunately recreational demand studies - no matter how broadly defined - frequently encounter low rates of participation in the population at large.

1. A Simple Model of Individual Behavior

The following might be conceived as a general model of an individual's demand for recreation trips

$$x_i = h(z_i, \beta, \varepsilon_i)$$

where x_i = quantity demanded by individual i, z_i is a vector of explanatory variables, β is a vector of parameters and ε_i is a random disturbance term. Unless this model is modified, though, it implicitly suggests the possibility of negative trips. For many functional forms (e.g. linear), a z_i vector could be faced and a disturbance term drawn from the distribution, such that x_i is less than zero. For other functional forms (e.g. semi-log), it may be impossible to generate negative values for x_i but equally impossible to generate zero which is a very legitimate and frequently observed value for x_i. The model must incorporate assumptions about individual behavior such that both positive and zero, but not negative, values of x_i will be generated.

The most popular assumption (and the one attributable to Tobin) is the following:

$$\text{(1)} \qquad x_i = h(z_i,\beta) + \varepsilon_i \qquad \text{if } h_i(\cdot) + \varepsilon_i > 0$$

$$x_i = 0 \qquad \text{if } h_i(\cdot) + \varepsilon_i \leq 0,$$

where $\varepsilon \sim N(0,\sigma^2)$. Presuming that the demand function is generated by utility maximizing behavior, this model seems to imply that preferences are defined over both positive and negative values of x, but reality prevents the consumption of negative quantities. Thus when the demand function would imply a negative quantity, a zero quantity is consumed.

Assuming that model (1) generated the behavior which is observed, let us consider what happens when conventional methods of estimation are employed. When individual observations are available, the customary practice is simply to estimate a demand function on data gathered from users. There are two problems with this approach. The first is that nothing is learned either about nonusers or about the factors which affect the decision to participate. There is, as a consequence, no way to predict changes in numbers of participants when parameters in the system change.

The second problem is that if nonparticipation is due to the underlying decision structure of the sort described in (1), then estimating demand functions from only users will generate biased coefficients. If behavior is described by model (1) and the ε's in the population are distributed as $N(0,\sigma^2)$, then the ε's associated with the sample of users will not meet Gauss-Markov assumptions. They will, by definition, be those ε's such that $\varepsilon_i > -h(z_i,\beta)$.

When only users are observed, the sample is said to be truncated. When the entire population is sampled but the value of the dependent variable (in this case, trips) is bounded (as in model (1)), the sample is said to be censored. Methods are well developed for consistent estimation of models from either type of sample (see G. S. Maddala, 1983, for a recent and extensive treatment), and some of these will be discussed below.

Both Wetzstein and Zeimer and Desvousges, Smith and McGivney recognized the presence of this problem in their recreational demand models. These studies employed Olsen's technique to make an approximate correction for the bias when only user data were available. It is useful, however, to explore other econometric techniques for eradicating the problem, some of which handle more general models of nonparticipation. We shall see that consistent parameter estimates can be obtained whether the sample is composed solely of users or drawn from the population as a whole. The latter type of sample will generate more efficient estimates, however.

If behavior is described by model (1), then the standard Tobit can be used to estimate the parameters of the model. From (1), an individual i will participate if $\varepsilon_i > -h_i(\cdot)$. Providing ε_i is distributed normally with mean 0, the transformed variable, ε_i/σ, has a standard normal distribution, and

$$\Pr\{i \text{ recreates}\} = \Pr\{\varepsilon_i/\sigma > -h_i(\cdot)/\sigma\}.$$

This probability equals 1 $-F(-h_i(\cdot)/\sigma)$, or $F(h_i(\cdot)/\sigma)$ where $F(\cdot)$ is the cumulative distribution function of the standard normal. The probability that i does not recreate is $F(-h_i/\sigma)$.

To form the likelihood function for the sample, we need an expression for the probability that i chooses x days given that $x_i > 0$. This is given by

$$\Pr(x_i|x_i>0) = \frac{f(\varepsilon_i/\sigma)/\sigma}{F(h_i/\sigma)} \tag{2}$$

where $f(\cdot)$ is the density function of the standard normal. Thus the likelihood function for the sample is

$$\begin{aligned} L_1 &= \prod_{i\varepsilon s} \Pr\{x_i > 0\} \Pr\{x_i | x_i > 0\} \prod_{i\varepsilon s} \Pr\{x_i = 0\} \\ &= \prod_{i\varepsilon s} f(\varepsilon_i/\sigma)/\sigma \prod_{i\varepsilon s} F(-h_i/\sigma), \end{aligned} \tag{3}$$

where s is the set of individuals who participate. The parameters β and σ, can be estimated from (3) using maximum likelihood methods.

There is a second procedure (attributable to Heckman) which uses a two step method in addressing the non-participation problem. Considering the same model, one can express the expected value of individual i's trips, given that i is a user as

$$E(x_i|\ x_i> 0) = h(z_i,\beta) + E(\varepsilon_i|\ \varepsilon_i> -h_i(\cdot)).$$

From the previous derivations, it can be seen that the second term is

$$E(\varepsilon_i|\ \varepsilon_i>-h_i) = \sigma E(\varepsilon_i/\sigma\ |\ \varepsilon_i/\sigma >-h_i/\sigma) = \frac{\sigma f(-h_i/\sigma)}{F(h_i/\sigma)}.$$

The demand for recreational trips can then be rewritten as

$$x_1 = h(z_i,\beta) + \sigma\lambda_i + v_i, \tag{4}$$

where λ_i equals $f(-h_i/\sigma)/F(h_i/\sigma)$ and v_i is a normal error with zero mean.

From this expression it is easy to see why OLS estimates of a model such as (1) are unsatisfactory. The denominator of λ, $F(h_i/\sigma)$, is the probability that an individual participates at the site. If there is a very high rate of participation among the population, the λ's will be small and

OLS estimates not too bad. The sample selection problem is most severe when there is a very low participation rate and the λ's are large. The presence of λ allows for the possibility of considerable misspecification, and its omission will cause the estimates of β to be biased where λ is correlated with any dimension of z.

Equation (4) can be estimated with ordinary least squares if λ_i is known. One way of obtaining an estimate of λ_i is to estimate a discrete choice model of the participation decision. Such a model would simply explain the yes/no decision. The logical choice for the qualitative response model is probit with a likelihood function expressed as

$$L_2 = \prod_{i \varepsilon s} F(h_i/\sigma) \prod_{i \varepsilon s} F(-h_i/\sigma). \tag{5}$$

From the earlier discussion, we know that $F(h_i/\sigma)$ is the probability of participating and $F(-h_i/\sigma)$ is the probability of nonparticipation. Maximum likelihood estimates of the β's and σ will allow construction of estimates of the λ_i's to be used in the estimation of (4).

One characteristic of this approach is that two sets of β's and σ are produced; one from each stage of the estimation. This may at first appear to be an unfortunate feature of the approach. However, two sets of estimates may be appropriate if the demand function is discontinuous or kinked at zero (see Killingsworth, 1983).

2. A Model of Behavior When Different Variables Affect Participation and the Demand for Trips

A logical extension of the discontinuity of the function at zero is the idea that different variables may affect the dichotomous participation decision and the continuous demand for trips decision. This may occur if factors such as good health or the ownership of an automobile or recreational equipment are necessary for an individual to become a participant. Along these lines, a final model is offered which employs Heckman's estimation technique but begins with a model of behavior which is more general than model (1). Consider a latent variable π^* which is an indicator of participation

$$\pi_i^* = g_1(z_{1i},\beta_1) + \varepsilon_{1i} \tag{6}$$

where the individual participates ($\pi_i = 1$) if $\pi_i^* > 0$ and the individual does not participate ($\pi_i = 0$) if $\pi_i^* \leq 0$. The number of trips taken by individual i given that i participates is

$$x_i = g_2(z_{2i}, \beta_2) + \varepsilon_{2i}. \tag{7}$$

Because π^* is an index denoting participation, π_i^* is observed only when $\pi_i^* > 0$. The vectors z_{1i} and z_{2i} may or may not have elements in common, and the covariance matrix of the ε's may or may not be diagonal.

The Heckman estimation technique is particularly suitable for this model . If information on nonusers as well as users is available, one can first estimate a probit model of the form

$$L_3 = \prod_{i \varepsilon s} F(g_1/\sigma_{11}) \prod_{i \varepsilon s} F(-g_1/\sigma_{11}) \tag{8}$$

where s is the set of participants and σ_{11} is the variance of the ε_{1i}'s. Note that this likelihood function is based only on the participation decision and requires a sample of the entire population.

Using Heckman's results,

$$E(x_i \mid z_{2i}, \pi_i^* > 0) = g_2(z_{2i}, \beta_2) + E(\varepsilon_{2i} \mid \varepsilon_{1i} > -g_1(z_{1i}, \beta_1))$$

so that

$$x_i = g_2(z_{2i}, \beta_2) + \frac{\sigma_{12}}{\sigma_{22}} \lambda_i + v_2 \tag{9}$$

where $\lambda_i = f(-g_{1i}/\sigma_{11})/(-F(-g_{1i}/\sigma_{11}))$ and σ_{12} is the covariance between ε_{1i} and ε_{2i}. Again an estimate of λ_i can be obtained from the probit model in (8).

3. Estimation When the Sample Includes Only Participants - the Truncated Sample

The above models are all well and good, but what happens when the sample of observations includes only participants? This is a common occurrence in specific recreational demand studies where the incidence of participation in the population at large is exceedingly low. In such cases, extremely large, and thus expensive, household sampling procedures would be necessary to produce sufficient observations on users. As a result, researchers sample on site and collect data only on participants.

While samples which include only participants preclude the use of some of the methods described above, it is still possible to obtain consistent although not particularly efficient estimates of the parameters of the demand for trips equation. To do this, we must refer back to the model of behavior presented in equations (1). It should be obvious that any more general model, such as those estimated with the Heckman technique, require information about non-participants and thus can not be used on a truncated sample. Model (1) however assumes that the same function determines whether individuals participate and if so, how much they participate, If this is true it is straightforward to estimate the demand for trips conditional on participation.

Referring back to equation (2), the probability that individual i's demand equals some x_i conditioned on the fact that he participates is given by

$$Pr(x_i | x_i > 0) = \frac{f(-h_i/\sigma)/\sigma}{F(h_i/\sigma)} .$$

The appropriate likelihood function for the sample is then simply

$$L_4 = \prod_{i \varepsilon S} \frac{f(-h_i/\sigma)/\sigma}{F(h_i/\sigma)} . \tag{10}$$

Because the added information about nonparticipants is missing, the estimates produced by this conditional maximum likelihood will be less efficient. Nonetheless the method corrects for truncated sample bias without requiring very expensive data collection.

Perhaps the greatest cost of a truncated sample is the paucity of information about the participation choice which it offers. Although it is technically feasible to use the coefficients generated by (10) to predict whether an individual drawn randomly from the population would participate in the activity or not, such predictions are dangerous. They rely on considerable confidence both in the estimated coefficients and in the model of behavior postulated in (1). Thus if other variables which are all-or-nothing threshold sorts of factors (e.g. health, equipment, etc.) affect participation, we will never learn much about the participation decision from a truncated sample.

Ultimately, the participation decision may be more or less important to capture. If the sorts of policy changes being considered (access, environmental quality, entrance fees) are likely to alter participation rates, then it is crucial for welfare evaluation that good predictions of participation be possible. Fortunately, the situations in which other discrete conditions affect participation may be just the cases where policy changes (such as environmental quality changes) are less likely to affect the participation/ nonparticipation choice.

One final caveat is in order here. Throughout this discussion, there has been an implicit restriction on the form of the demand function. While we have not required the demand function to be linear, we have assumed errors to be additive. Forms such as the semi-log do not have this property, and as we noted they have the additional problem of not admitting zero values for the dependent variable. As such the semi-log form is logically inconsistent with the notion of nonparticipation and the models of behavior presented above. More general functional forms, such as the Box-Cox transformation, do allow for nonparticipation. However, the error structure may not always be additive. In these cases the above results will hold in spirit but not in detail.

Implications for the Estimation of the Zonal Approach

While researchers have recognized the advantages of using individual observations to estimate recreational demand models, there has been some suspicion that the zonal approach avoids the types of participation rate problems encountered above. In truth, the zonal approach is plagued with similar and sometimes additional problems which become apparent when a model of behavior such as (1) is postulated.

Assume that the simple model in expression (1) reflects the actual behavior of individuals, but that only zonal data is available. The zones in our discussion will be assumed to be distinct and well-defined, whether determined by political boundaries such as counties or by distance from site as originally conceived by Hotelling. Suppose that there are M such zones, and in each zone $j(j = 1, M)$ there are P_j people (the level of population), I_j of whom visit the site at least once. The individuals, $i = 1, I$, within the zone may differ with respect to explanatory variables, z_{ij}, error term, ε_{ij}, and chosen number of trips, x_{ij}. The model in (1) is rewritten using this notation

$$
(11)\qquad
\begin{aligned}
x_{ij} &= h(z_{ij},\beta) + \varepsilon_{ij} && \text{if } h_{ij} + \varepsilon_{ij} > 0 \\
x_{ij} &= 0 && \text{if } h_{ij} + \varepsilon_{ij} \le 0.
\end{aligned}
$$

From the above definitions, n_j/P_j is the proportion of the population who visit the site at least once. Both n_j and n_j/P_j are random because n_j is the realization through expression (11) of random drawings of the disturbance term. Denote the first n_j people as participants and the last $P_j - n_j$ as nonparticipants. Defining $\bar{x}_j$ as the zonal average for zone j, the $\bar{x}_j$ are as follows:

$$\bar{x}_j = \sum_{i=1}^{P_j} x_{ij}/P_j = \sum_{i=1}^{n_j} x_{ij}/P_j + \sum_{i=n_j+1}^{P_j} 0$$

$$= \sum_{i=1}^{n_j} (h(z_{ij},\beta) + \varepsilon_{ij})/P_j$$

(12)
$$= \sum_{i=1}^{n_j} h(z_{ij},\beta)/P_j + \sum_{i=1}^{n_j} \varepsilon_{ij}/P_j.$$

Let us employ two assumptions for the moment, one of which in fact favors the zonal approach. We assume that $h(\cdot)$ is linear in the explanatory variables and that each zone is sufficiently homogeneous such that the assumption that $z_{ij} = z_{kj}$ for all i,k is reasonable.

Then (12) becomes

(13)
$$\bar{x}_j = \frac{n_j}{P_j} \beta' z_j + \sum_{i=1}^{n_j} \varepsilon_{ij}/P_j.$$

This expression reflects the nature of the zonal data observed when expresslon (11) describes the individual's decision process.

Two problems are encountered when one attempts to apply OLS techniques to zonal data. The first problem is that the error term in (13) does not possess the prescribed properties. If it is assumed that ε_{ij} is distributed normally with mean zero and variance σ^2, then ε_{ij}/P_j is distributed normally with mean zero, variance σ^2/P_j^2 which implies a heteroskedastic problem of the sort discussed in the literature, but easily corrected. The error term in (13) is actually a sum of such terms. While the sum of independent normals is itself a normal, the error term here is not the sum of independent normals. The term

$$\sum_{i=1}^{n_j} \varepsilon_{ij}/P_j$$

reflects the same sort of selection bias encountered in the previous section, because it is the sum of errors conditional on $h(\cdot) + \varepsilon_{ij} > 0$, and its expected value will not be zero.

The second problem relates to the fact that, in general, n_j/P_j will vary over observations (i.e. over zones). The participation rate per zone, i.e. n_j/P_j , will not tend to be constant, since as distance between origin and site increases, participants take fewer trips and there are fewer participants.

Consequently a regression of $\bar{x}_j$ on the z_j will not yield estimates of β (even up to a proportionality constant). To assume however implicitly that n_j/P_j is constant violates the assumptions of the model. The participation rate cannot be constant and non-random, because it is determined in part by random errors and in part by systematic variation in factors such as travel cost.

If the n_j/P_j were known, however, it would be possible to estimate the β's in (13) by weighting the explanatory variables by the participation rate. This would not, however, resolve the problem with the error term and a technique such as Heckman's would be needed to estimate the zonal model.

Conclusions

In principle, models estimated on individual observations are preferable to those based on zonal aggregates. Inferences about parameters of the preference function are more directly revealed and thus welfare measures easier to define. Individual observations also provide more information and may help avoid multicollinearity and heteroskedasticity problems aggravated by the zonal approach. Perhaps the chief drawback to using individual observations is that they are more likely to embody severe errors in measurement. Also it may be more difficult to extrapolate welfare measures for the entire population from models based on individual data.

All of this abstracts from the overriding aggregation issue implicit in estimating recreation models - the treatment of nonparticipants. There is some indication in the literature that the zonal approach may be superior in dealing with this problem. As we show in this chapter, this supposition is incorrect. In fact the participation issue arises in the estimation of both individual and zonal based models. Both models will yield biased parameter estimates if the problem - one of truncated or censured samples - is ignored. The key point is that individual data-based models which take this problem into account are well developed. Methods exist for estimating a wide selection of models of individual behavior which allow for nonparticipation or which use truncated samples and are conditioned on participation. While more flexible models and more efficient estimates are possible when both users and nonusers are sampled, methods for obtaining consistent estimates exist for samples of users only. In contrast, zonal models actually confound the problem of participation. It is never quite clear what such models are estimating and how they can be adjusted to recover the parameters of interest to us.

In the next chapter, we provide an example of the application of some of the methods for taking account of the participation decision when individual data is available. This is pursued in conjunction with a development of the treatment of the value of time, so that a more complete model can be presented.

CHAPTER 4

SPECIFICATION OF THE INDIVIDUAL'S DEMAND FUNCTION: THE TREATMENT OF TIME

Economists, especially those working in the area of recreational demand, have long recognized that time spent in consuming a commodity may, in some cases, be an important determinant of the demand for that commodity. It remains true, however, that even though the potential importance of time has been discussed at some length in the literature it is only relatively recently, and in a fairly small set of papers, that the problem of explicitly incorporating time into the behavioral framework of the consumer has been addressed.

This chapter provides a discussion of the ways in which researchers have traditionally incorporated time costs into recreational demand models and attempts to develop a more complete and general model. Improvements in both specification and estimation of the model are achieved by integrating recent labor supply and recreational demand literature. The new model of individual decision making is characterized by two constraints. Insights into the dual constraint model are offered.

The treatment of time is one of the thorniest issues in the estimation of recreational benefits. A number of approaches (e.g. Smith et al., 1983; McConnell and Strand, 1981; Cesario and Knetsch, 1970) to valuing time are currently in vogue, but no method is dominant and researchers often improvise as they see fit. Unfortunately, the benefit estimates associated with changes in public recreation policy are extremely sensitive to these improvisations. Cesario (1976), for example, found that annual benefits from park visits nearly doubled depending on whether time was valued at some function of the wage rate or treated independently in a manner suggested in Cesario and Knetsch (1970). More recently, Bishop and Heberlein (1980) presented travel cost estimates of hunting permit values which differed four-fold when time was valued at one-half the median income and when time was omitted altogether from the model.

Recreational economists have understood the applicability of the classical labor-leisure trade-off to this problem. In his 1975 article, McConnell was the first to discuss the one vs. two constraint model. Recognizing that time remaining for recreation may be traded off for work time or it may be fixed, he shows how the nature of the decision problem is affected by the nature of the time constraint. This chapter begins within this context and develops a general framework for incorporating time. After discussing the wide range of complex labor constraints which the model can handle, we turn to making the model operational. The approach developed below not only incorporates a defensible method for treating the value of time but also permits sample selection bias (Chapter 3) to be addressed and exact measures of welfare (Chapter 2) to be derived.

Time in Recreational Decisions

Despite the general acceptance that time plays an important role in recreational decisions (e.g. Smith, et al., 1983), no universally accepted method for incorporating time into recreational demand analysis has emerged and methods for "valuing" time in recreational demand models are numerous. While many methods have been developed from assumptions based on utility maximizing behavior, there is no consensus as to which is the "correct" method. In actual applications, researchers have often been forced to take a relatively *ad hoc* view of the problem by incorporating travel time in an arbitrary fashion as an adjustment in a demand function or, alternatively, by asking people what they would be willing to pay to reduce travel time.

Ad hoc econometric specifications or general willingness-to-pay questions are particularly problematic with respect to time valuation because time is such a complex concept. Time, like money, is a scarce resource, for which there is a constraint. Anything which uses time as an input consumes a resource for which there are utility-generating alternatives. While time is an input into virtually every consumption experience, some commodities take especially large amounts of time. These have frequently been modeled in a household production framework to reflect the individual's need to combine input purchases with household time to "produce" a commodity for consumption. Because time is an essential input into the production of any commodity which we might call an "activity", time is frequently used as a measure of that activity as well. Thus, while time is formally an input into the production of the commodity, it may also serve as the unit of measure of the output.

The complexity of time's role in household decisions has implications for both travel and on-site recreational time. Both represent uses of a scarce resource and thus have positive opportunity costs. However, on-site time, and sometimes travel time, are used as units of measure of the utility generating activities themselves. Economists often measure the recreational good in terms of time, i.e. in hours or days spent at the site. Travel time may also be a measure of a utility generating activity, if the travel is through scenic areas or if it involves other activities such as visiting with traveling companions. Hence, direct questioning or poorly conceived econometric estimation may yield confusing results because the distinction between time as a scarce resource and time as a measure of the utility generating activity is not carefully made.

Both travel time and on-site time are uses of the scarce resource and must both appear in a time constraint to be properly accounted for in the model. The exclusion of either will bias results. But, does time belong in the utility function? Viewed as a scarce resource, time by itself does not belong in the utility function. What does enter the utility function is a properly conceived measure (perhaps in units of time) of the quantity and quality of the recreational activity. This does not present major problems when the commodity is defined in terms of fixed units of on-site time and when travel does not in itself influence utility levels. When time per trip is a decision variable, an appropriate and tractable measure is not easily conceived. This Chapter focuses solely on time as a scarce resource.

Time as a Component of Recreational Demand: A Review

The fact that time costs could influence the demand for recreation was recognized in the earliest travel cost literature (Clawson, 1959; Clawson and Knetsch, 1966), although no attempt was made to explicitly model the role of time in consumer behavior. The problems which arise when time is left out of the demand for recreation were first discussed by Clawson and Knetsch (1966). Cesario and Knetsch (1970) later argued that the estimation of a demand curve which ignored time costs would overstate the effect of price changes and thus understate the consumer surplus associated with a price increase.

In practical application, both travel cost and travel time variables have usually been calculated as functions of distance. As a result, including time as a separate variable in the demand function tended to lead to multicollinearity. Brown and Nawas (1973) and Gum and Martin (1975) attempted to deal with the multicollinearity issue by suggesting the use of individual trip observations rather than zonal averages. In contrast, Cesario and Knetsch (1976) proposed combining all time costs and travel costs into one cost variable to eliminate the problem of multicollinearity. These papers had a primarily empirical focus, with emphasis given to obtaining estimates. Demand functions were specified in an arbitrary way, with no particular utility theoretic underpinnings.

Johnson (1966) and McConnell (1975) were among the first to consider the role of time in the context of the recreationalists's utility maximization problem (although others had considered it in other consumer decision problems). McConnell specified the problem in the framework of the classical labor-leisure decision. The individual maximizes utility subject to a constraint on income and time. The income constraint is defined such that

(la) $$E + F(T_w) = px_N + c'x_R$$

where E is non wage income, T_w is work time, $F(T_w)$ is wage income, p is the price of a Hicksian good x_N, x_R is a column vector of recreational activities and c is the corresponding vector of money costs for each unit of x_R. His time constraint is

(lb) $$T = \Sigma a_j x_j + T_w$$

where a_j is the time cost of a unit of x_j. When work time is not fixed, (lb) can be solved for T_w and substituted into (la) yielding the maximization problem

$$\max U(x) - \lambda(px_N + \Sigma c_j x_j - F(T - \Sigma a_j x_j)),$$

so that the time cost is transformed into a money cost at the implicit wage rate.

McConnell (1975) also noted that if individuals were unable to choose the number of hours worked, the direct substitution of (la) into (1b) is not possible. tie suggested that in this case one should still value time in terms of money before incorporating it in the demand function. This is conceptually possible, since at any given solution there would be an amount of money which the individual would be just willing to exchange for an extra unit of time so as to keep his utility level constant. Unfortunately, this rate of trade-off between money and time, unlike the wage rate, is neither observable nor fixed. It is itself a product of the individual's utility maximizing decision.

Much of the recent recreation demand literature follows the line of reasoning which related the opportunity cost of time in some way to the wage rate. Of the many models of this sort, the one offered by McConnell and Strand (1981) is one of the most recent. (See also Cesario, 1976; Smith and Kavanaugh, Nichols et al., 1978). Their work demonstrates a methodology from which a factor of proportionality between the wage rate and the unit cost of time can be estimated within the traditional travel cost model.

More recently, Smith, Desvousges and McGivney (1983) attempted to modify the traditional recreational demand model so that more general constraints on individual use of time were imposed. They considered two time constraints, one for work/non-recreational goods and another for recreational goods. The available recreation time could not be traded for work time. The implications of their model suggest that when time and income constraints cannot be reduced to one constraint, the marginal effect of travel and on-site time on recreational demand is related to the wage rate only through the income effect and in the most indirect manner. Unfortunately, their model "does not suggest an empirically feasible approach for treating these time costs" (p. 264). For estimation, they confined themselves to a modification of a traditional demand specification.

Researchers are thus left with considerable confusion about the role of the wage rate in specifying an individual's value of time. But there is an important body of economics literature, somewhat better developed, which has attempted to deal with similar issues. Just as the early literature on the labor-leisure decision provided initial insights into the modeling of time in recreational demand, more recent literature on labor supply behavior provides further refinement.

Time in the Labor Supply Literature: A Review

The first generation of labor supply models resembled the traditional recreational demand literature in a number of ways. These models treated work time as a continuous choice variable. A budget constraint such as that depicted in Figure 4.1 was assumed for each individual, suggesting the potential for a continuous trade-off between money and leisure time at the wage rate, w. In this graph, E is non-wage income and T is total available hours Participants in the labor force were assumed to be at points in the open interval (BC) on the budget line, equating their marginal rates of substitution between leisure and goods to the wage rate. Those who did not participate were found at the corner solution B.

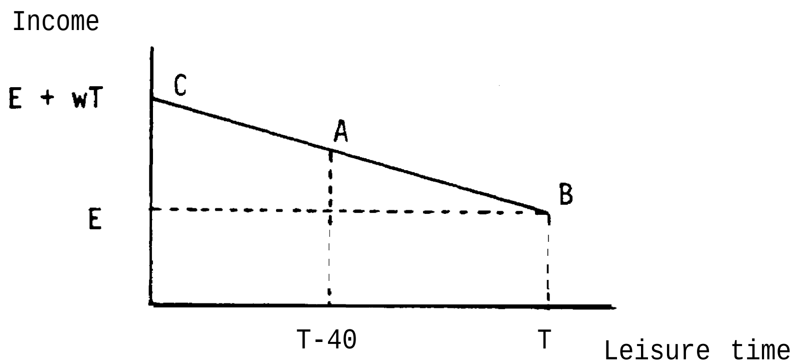


Figure 4.1: The First Generation Budget Constraint

Other researchers argued that work time may not be a choice variable. Individuals might be "rationed" with respect to labor supply in a "take-it-or-leave-it" fashion; that is they may be forced to choose between a given number of work hours (say 40 hours/week) or none at all (Perlman, 1966; Mossin and Bronfenbrenner, 1967). In this context, there is no opportunity for marginally adjusting work hours, and all individuals are found at one of two corner solutions (A or B in Figure 4.1).

While useful in characterizing the general nature of a time allocation problem, first generation labor supply models were criticized on both theoretical and econometric grounds. These concerns fostered a second generation of labor supply research which made improvements in modeling of constraints and in estimating parameters as well as making models more consistent with utility maximizing assumptions (see Killingsworth, 1983, p. 130-1). Each of these areas of development have implications for the recreation problem.

The second generation of labor supply literature (see for example Ashenfelter, 1980; Ham, 1982; Burtless and Hausman, 1978) generalized the budget line to reflect more realistic assumptions about employment opportunities. As Killingsworth states in his survey, "...the budget line may not be a straight line: Its slope may change (for example, the wage a moonlighter gets when he moonlights may differ from the wage he gets at his 'first' job), and it may also have 'holes' (for example, it may not be possible to work between zero and four hours)".

To appreciate this point, consider an example: an individual whose primary job requires T_p hours per week within a total time constraint of T hours per week. The relevant wage rate at this primary job is w_p and is depicted in Figure 4.2 as the slope of the implied line segment between A and B. This individual can earn more wage income only by moonlighting at a job with a lower wage rate (depicted by the slope of the segment between A and C). His relevant budget line is segment AC and point B. Depending on his preference for goods and leisure, he may choose not to work and be at B; he may work a fixed work week at A; or he may take a second job and be along the segment AC. Consideration of more realistic employment constraints such as these have implications for model specification. Only those individuals who choose to work jobs with flexible work hours (e.g. self employed professionals, and individuals working second jobs or part-time jobs) can adjust their marginal rates of substitution of goods for leisure to the wage rate. All others can be found at corner solutions where no such equi-marginal conditions hold.

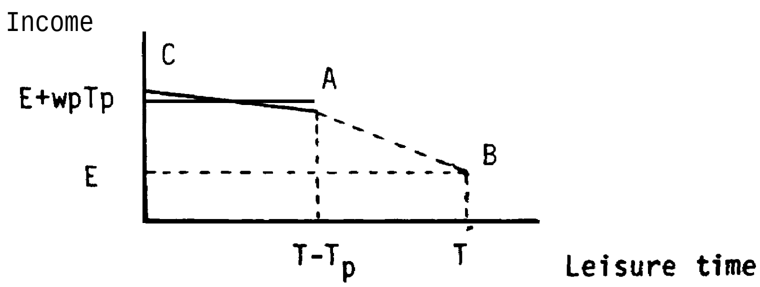


Figure 4.2: Second Generation Budget Constraints

Two other aspects of the second generation labor supply models are noteworthy. The first generation studies estimated functions which were specified in a relatively <u>ad hoc</u> manner. By contrast, second generation models have tended to be utility-theoretic. This has been accomplished by deriving specific labor supply functions from direct or indirect utility functions (Heckman, Killingsworth, and MacCurdy, 1981; Burtless and Hausman, 1978; Wales and Woodland, 1976, 1977). Such utility-theoretic models have particular appeal for recreational benefit estimation because they allow estimation of exact welfare measures. Additionally, first generation research was concerned either with the discrete work/non-work decision or with the continuous hours-of-work decision. Second generation empirical studies recognized the potential bias and inefficiency of estimating the two problems independently and employed estimation techniques to correct for this.

A Proposed Recreational Demand Model

It is clear that the nature of an individual's labor supply decision determines whether his wage rate will yield information about the marginal value of his time. In the recreational literature, researchers have conventionally viewed only two polar cases: either individuals are assumed to face perfect substitutability between work and leisure time or work time is assumed fixed. The choice between these two cases is less than appealing. Few people can be considered to have absolutely fixed work time, since part-time secondary jobs are always possible. On the other hand, only some professions allow free choice of work hours at a constant wage rate. Additionally no sample of individuals is likely to be homogeneous with respect to these labor market alternatives. A workable recreation demand model must reflect the implications which labor decisions have on time valuation and allow these decisions to vary over individuals.

In developing a behavioral model that includes time as an input it is useful to broaden the description of the nature of the decision problem beyond the simple travel cost framework. The more general household production model depicts the individual maximizing utility by choosing a flow of recreational services, x_R, and a vector of other commodities, x_N. A vector of goods, S_R, is combined with recreation time, T_R, to produce x_R. Both time, T_N, and purchased inputs, S_N, may be required to produce x_N.

The individual's constrained utility maximizing problem can be represented as

(2) $$\max_{S,T} U(x_R, x_N)$$

subject to $$x_R = f(S_R, T_R),$$
$$x_N = g(S_N, T_N),$$

$$E + F(T_w) = v_N'S_N + v_R'S_R,$$

and

$$T = T_w - T_R - T_N,$$

where U(...) is a quasi-concave, twice-differentiable utility function, f(...) and g(...) are vectors of quasi-convex, twice-differentiable production functions, $E + F(T_w)$ is the sum of the individual's non-wage and wage income, v_R and v_N are the price vectors associated with the vectors of recreational and non-recreational inputs respectively, T_w is labor time supplied, and T is the total time available.

We reduce the problem by assuming (as do Burt and Brewer, 1971, and others before us) a Leontief, fixed-proportions technology. This is equivalent to assuming that the commodities, i.e. the x's, have fixed time and money costs per unit given by t and p, respectively. For the recreation good, x_R, it implies that a unit of x_R (e.g. a visit) has a constant marginal cost (p_R) and fixed travel and on-site time requirements (t_R). All other commodities are subject to unit money or time costs and the general problem becomes

(3) $$\max_{x_R, x_N} \quad U(x_R, x_N)$$

subject to $$E + F(T_w) - p_R'x_R - p_N'x_N = 0,$$

and $$T - T_w - t_R'x_R - t_N'x_N = 0,$$

were p and t are the unit money and time prices of the x's.

In order to characterize an individual's solution to the problem posed in (3), it is necessary to know the nature of the labor market constraints. For any individual, it is possible that an interior solution is achieved, such as along line segment AC in Figure 4.2. The individual can adjust work time such that his marginal rate of substitution between leisure and goods equals his effective (marginal) wage rate. As Killingsworth points out, this is most likely to be true for individuals who work overtime or secondary jobs, but may also be true for those with part-time jobs and those (e.g. the self-employed) with discretion over their work time. An individual may, alternatively, be at a corner solution such as point A or B in Figure 4.2. Point B is associated with unemployment, while an individual at point A works some fixed work week at wage w_P and has the opportunity to work more hours only at a difficult wage. In neither case is there a relationship between the wage rate the individual faces and his valuation of time.

Strictly speaking, the problem in (3) requires the simultaneous choice of the x's and the individual's position in the labor market (i.e. interior or corner solution). It is, however, beyond the scope of most recreation demand studies to model the entire labor decision. Labor market decisions may well be affected by individuals' recreational preferences and the type of recreational opportunities available to them. However, the sort of day to day and seasonal recreational choices about which data is collected and models developed can reasonably be treated as short run decisions conditioned on longer run labor choices. Since there are high costs to changing jobs, adjustments in labor market situations are not made continually. Thus, recreational choices are considered to be conditioned on the type of employment which the individual has chosen. Of course if the individual chooses an employment situation with flexible work hours, then time spent working _is_ treated as endogenous to the model.

The problem as posed in (3) is restated and the first-order conditions provided, given alternative solutions to the labor supply problem. For individuals at corner solutions (such as B or A in Figure 4.2), the problem becomes

$$\max_{x} U(x) + \lambda(\bar{Y} - \Sigma p_i x_i) + \mu(\bar{T} - \Sigma t_i x_i) \tag{4}$$

where $\bar{Y}$ is effective income (including the individual's wage income if he works and nonwage income which may include the individual's share of the earnings of other household members). The variable $\bar{T}$ is time available (after job market activities) for household production of commodities, including recreation.

First order conditions are

$$\begin{aligned} &\partial u/\partial x_i - \lambda p_i - \mu t_i = 0 \\ &Y - \Sigma p_i x_i = 0, \\ &\bar{T} - \Sigma t_i x_i = 0. \end{aligned} \tag{4a}$$

Note that since work time cannot be adjusted marginally, the two constraints are not collapsible. Solving (4a) for the demand for xi yields a demand function of the general form

$$x_i = h^C(p_i, t_i, p^o, t^o, Y, T) + \varepsilon \tag{4b}$$

where p^o and to are the vectors of money and time costs of all other goods and ε is the random element in the model. (The properties of this demand function are detailed in the Appendix to this Chapter.)

For an interior solution in the labor market, however, at least some component of work time is discretionary and time can be traded for money at the margin. Thus, the time constraint in problem (3) can be substituted into the income constraint, yielding the one constraint

$$\bar{Y} + w_D\bar{T} - \Sigma(p_i + w_D t_i)x_i = 0$$

where w_D is the wage rate applicable to discretionary employment.

The maximization problem conditioned on an interior solution to the labor supply decision is

$$\max_{x} U(x) + \delta(\bar{Y} + w_D\bar{T} - \Sigma(p_i + w_D t_i)x_i). \tag{5}$$

First order conditions are

$$\partial u/\partial x_i - \delta(p_i + w_D t_i) = 0 \tag{5a}$$

$$\bar{Y} + w_D\bar{T} - \Sigma(p_i + w_D t_i)x_i = 0.$$

Solving for the general form of a recreational demand function for the interior solution yields

$$X_i = h^I(p_i + w_D t_i,\ p^o + w_D t^o, \bar{Y} + w_D\bar{T}) + \varepsilon. \tag{5b}$$

Note that, for empirical purposes, $\bar{Y} + w_D\bar{T}$ can be re-expressed in terms of variables easily elicited on a questionnaire. The term $\bar{Y} + w_D\bar{T}$ equals $Y + w_D t_D + w_D(T-t_D)$ where t_D is discretionary work time, Y is total income, and $T-t_D$ is the time available for household production (or total time minus all hours worked).

Consideration of demand functions (4b) and (5b) suggests that the data requirements of estimation are not overly burdensome. In addition to the usual questions about income, and the time and money costs of the recreational activity, one need only ask a) the individual's total work time and b) whether or not he has discretion over any part of his work time. If he does, his discretionary wage must be elicited.

In problem (5) the recreational demand function is conditioned on the individual having chosen an interior solution in the labor market. The wage rate (w_D) reflects the individual's value of time because work and leisure can be traded-off marginally. However, when this is not the case as in problem (4), the marginal value of the individual's time in other uses is not equal to the wage rate he faces. This does not imply that the opportunity cost of time is zero for such an individual. It is only that his opportunity cost is not equal to an observable parameter. The opportunity cost of an individual's time will be affected by the alternative uses of his time.

Considerations for Estimating Recreational Benefits

In order to estimate recreational demand functions and thus derive benefit estimates, it is necessary to define a specific form for the demand equation and to postulate an error structure.

This task is complicated by the fact that the individual's decision problem, as formulated in this Chapter, is not the classical one. The problem is now the maximization of utility subject to both an income and a time constraint. The comparative statics and general duality results of utility maximization in the context of two constraints are developed in the Appendix to this Chapter. There, it is demonstrated rigorously that maximization under two linear constraints yields a demand function with properties analogous to the one constraint case. The demand function is still homogeneous of degree zero, but in a larger list of arguments - money prices, time prices, income and time endowments. It also satisfies usual aggregation conditions. In addition, two duals are shown to exist - one which minimizes money costs subject to utility and time constraints and the other which minimizes time costs subject to utility and income constraints. Associated with each dual is an expenditure function and a compensated demand. Both income and time compensated demands are own price downward sloping and possess symmetric, negative semidefinite substitution matrices.

Despite the analogies which exist between the one and two constraint models, integrating a demand function back to an indirect utility function is not straightforward in the two constraint case. In addition, it is not altogether obvious how the Vartia numerical approximation techniques described in Chapter 2 can be applied when the demand function derives from utility maximization subject to two constraints. Consequently it is useful to begin with a direct utility function and solve for recreational demand functions by maximizing utility subject to the appropriate constraint set. The form of the demand functions and the indirect utility function will depend on which constraint set is relevant. Rather than deal with the general model, a specific case is shown here.

The utility function chosen for illustration is

$$(6) \qquad U(x) = \frac{(\gamma_1 + \gamma_2)x_1 + \beta}{(\gamma_1 + \gamma_2)^2} \exp \left[\frac{(\gamma_1 + \gamma_2)(\alpha + \gamma_1 x_2 + \gamma_2 x_3 - x_1 + \varepsilon)}{(\gamma_1 + \gamma_2)x_1 + \beta}\right].$$

In the above expression α, β, γ_1, and γ_2 are parameters common to all individuals and ε represents a random element reflecting the distribution of preferences over the population. The random variable, ε, is assumed to be distributed normally with mean zero and constant variance, σ^2.

The recreational good is designated as x_1. We partition the set of other goods such that x_2 is a bundle of goods with money but no significant time costs. The bundle, x_2, is a numeraire such that the money price of recreation is normalized with respect to p_2. Hicksian bundle x_3 is a bundle of goods with time but no significant money costs and serves as a numeraire such that time prices are normalized with respect to t_3. Thus the general constraint set is

$$Y - p_1 x_1 - p_2 x_2 = 0$$

and

$$\bar{T} - t_1 x_1 - t_3 x_3 = 0$$

where p_2 and t_3 are assumed to be equal to one, forthwith.

Solving the system for the optimum value of x_1, and denoting $\beta/(\gamma_1 + \gamma_2)$ as β', yields ordinary recreational demand functions conditioned on each labor supply decision of the form

(7) $$x_1 = \alpha + \gamma_1\bar{Y} + \gamma_2\bar{T} + \beta'\gamma_1 p_1 + \beta'\gamma_2 t_1 + \varepsilon$$

for individuals at corner solutions in the labor market, and

(8) $$x_1 = \alpha + \gamma_1(\bar{Y} + w_D\bar{T}) + \beta'\gamma_1(P_1 + w_D t_1) + \varepsilon$$

for individuals at interior solutions in the labor market.

A word about the particular utility function chosen is in order. Because there are potentially two constraints, the utility function must accommodate three goods. Only x_1 is of interest, however, and in order to avoid the more complex problem of integrating back from systems of demand equations, a bivariate direct utility function with useful properties was modified to include three goods. The modification involves the inclusion of x_2 and x_3 in such a way as to imply that they are perfect substitutes. This procedure has two unfortunate ramifications. For an interior solution, when the two constraints collapse into one, this form implies that either x_2 or x_3 is chosen (but not both). Which one is chosen depends on the relative sizes of the prices and parameters. It turns out that if x_3 is chosen, then the coefficient γ_1 must be replaced by γ_2 in (8). Another unsatisfactory feature is that for corner solutions, when the two constraints are not collapsible, the functional form implies a constant trade off between time and money, equal to $-\gamma_2/\gamma_1$. This is a direct result of the perfect substitutability between x_2 and x_3 which produces linear iso-utility curves in time and income space.

Despite the somewhat restrictive properties of the utility function in (6), its maximization subject to the two constraints allows us to make operational a demonstration of the suggested approach. It is interesting to note that equations (7) and (8), being linear in the respective variables, could easily have been specified as ad hoc demand functions, without reference to utility theory. This would not have altered the implicit restrictions on preferences implied - no one would have understood their implications. Additionally, one would have no way of properly interpreting the parameters or of calculating estimates of compensating and equivalent variation.

Since the two constraint problem possesses two duals and thus two expenditure functions, compensating variation can be measured in terms of either of two standards - time or money or a combination of both. The anomalies which this can cause are discussed elsewhere (see Bockstael and Strand, 1985). Here compensating variation measures of the price change which drives the demand for x to zero in terms of both time and money are presented. For the interior solution, the money compensating variation is given by

$$CV_I^Y = \exp[\gamma_1(\tilde{p}_1 - p_1^{\,o})]\,\left(\frac{x_1^{\,o} + \beta'}{\gamma_1}\right) - \frac{\beta'}{\gamma_1} \tag{9a}$$

for the interior solution, where $(p_1^{\,o}, x_1^{\,o})$ is the initial observed point. The time compensating variation for individuals at interior solutions is

$$CV_I^T = \exp\,[\gamma_1(\tilde{p}_1 - p_1^o)]\,\left(\frac{x_1^{\,o} + \beta'}{\gamma_1 w_D}\right) - \frac{\beta'}{\gamma_1 w_D}. \tag{9b}$$

Compensating variation for the two constraint case can be specified by first substituting demand functions into (7) to obtain the indirect utility function

$$V(p,t,Y,T) = \exp(-\gamma_1 p_1 - \gamma_2 t_1)\,\left(\frac{\alpha + \gamma_1\bar{Y} + \gamma_2\bar{T} + \beta'\gamma_1 p_1 + \beta'\gamma_2 t_1 + \beta'}{\gamma_1 + \gamma_2}\right)$$

and inverting to obtain the money expenditure function

$$m_y = \frac{\gamma_1 + \gamma_2}{\gamma_1} U^o \exp\,(\gamma_1 p_1 + \gamma_2 t_1) - \frac{1}{\gamma_1}\,(\alpha + \gamma_2\bar{T} + \beta'\gamma_1 p_1 + \beta'\gamma_2 t_1 + \beta')$$

where U^o is the initial level of utility. The money compensating variation for a loss of the recreation good conditioned on a corner solution in the labor market is then

$$CV_C^Y = \exp[\gamma_1(\tilde{p}_1 - p_1^o)]\,\left(\frac{x^o + \beta'}{\gamma_1}\right) - \frac{\beta'}{\gamma_1}. \tag{10a}$$

The time expenditure function for this group equals

$$m_Y = \frac{\gamma_1 + \gamma_2}{\gamma_2} U^o \exp(\gamma_1 p_1 + \gamma_2 t_1) - \frac{1}{\gamma_2}(\alpha + \gamma_2\bar{T} + \beta'\gamma_1 p_1 + \beta'\gamma_2 t_1 + \beta')$$

and the associated time compensating variation equals

$$CV_C^T = \exp[\gamma_1(\tilde{p}_1 - p_1^o)]\,\left(\frac{x^o + \beta'}{\gamma_2}\right) - \frac{\beta'}{\gamma_2}. \tag{10b}$$

Estimating the Model: The Likelihood Function

As discussed in Chapter 3, a random sample of the population will produce a significant portion of nonparticipants. To correct for the truncated sample problem which nonparticipation would generate, the Tobit model discussed in the previous chapter is employed. The j^{th} individual is observed to take some positive number of recreational trips, x, if and only if the cost of the trip, p, is less than his reservation price $\tilde{p}$, where the reservation price is a function of other factors influencing the individual. Thus

$$\begin{aligned} x_j &= h_j(\cdot) + \varepsilon_j && \text{if and only if } h_j(\cdot) + \varepsilon_j > 0 \\ x_j &= 0 && \text{otherwise} \end{aligned}$$

where $h_j(\cdot)$ is the systematic portion of the appropriate demand function evaluated for individual j (eq. 4b or 5b).

Referring back to the deviation of the likelihood function presented in equation 3 of Chapter 3, if the sample of persons is divided so that the first m individuals recreate and the last n - m do not, then the likelihood function for this sample is

$$(11) \qquad L_1 = \prod_{j=1}^{m} f(\varepsilon_j/\sigma)/\sigma \prod_{j=m+1}^{n} F(-h_j(\cdot)/\sigma).$$

This general form of the likelihood function will be true for each labor-market group. However, account must be given to the difference in the demand functions for each group. Thus, for our entire sample of persons with interior and corner solutions in the labor market, the likelihood function is

$$(12) \quad L^* = \prod_{j=1}^{m_C} f(\varepsilon_j^C/\sigma)/\sigma \prod_{j=m_C+1}^{n_C} F(-h_j^C(\cdot)/\sigma) \prod_{j=1}^{m_I} f(\varepsilon_j^I/\sigma)/\sigma \prod_{j=m_I+1}^{n_I} F(-h_j^I(\cdot)/\sigma)$$

where the subscripts c and I refer to numbers of individuals with corner and interior solutions respectively.

Should only observations on participants exist, one can still avoid sample selection bias by employing a form of the conditional likelihood function as presented in equation (10) of Chapter 3. The conditional probability of an individual j taking x_j visits given that x_j is positive is given by

$$(13) \qquad L^p = \prod_{j=1}^{m_C} \frac{f(\varepsilon_j^C/\sigma)/\sigma}{F(h_j^C(\cdot)/\sigma)} \prod_{j=1}^{m_I} \frac{f(\varepsilon_j^I/\sigma)/\sigma}{F(h_j^I(\cdot)/\sigma)}.$$

An Illustration

The purpose of this section is to demonstrate the application of the proposed approach for estimating recreational demand functions and for calculating recreational losses associated with elimination of the recreational site. In a Monte Carlo exercise, comparison of this model with those generated by traditional approaches is made. The exercise gives an example of how the traditional approaches can produce biased parameter estimates and inaccurate benefit measures. For an application to actual survey data see Bockstael, Strand, and Hanemann (1985).

To have a standard by which results can be compared, we begin with a direct utility function of the form in (6), choose parameter values (see Table 4.1, true model), and generate ten samples of individual observations. Each sample or replication is composed of 240 drawings, one third of which are consistent with each of the following situations: a) an interior solution in the labor market, b) a fixed work week solution, and c) unemployment. Two hundred forty values for wage income, non-wage income, secondary wage rate, travel cost and travel time are randomly drawn from five rectangular distributions R($0,$25,000), R($0,$1000), R($2.5, $5.0), R($0, $60) and R(0,4), respectively, and these values for the exogenous variables are repeated in each replication. The replications are different in that independent error terms are drawn from a normal distribution, N(0,25), for each of the 2400 individual observations.

Total recreational time is taken to be the sum of travel and on-site time. While it is assumed on-site time is exogenous, fixed at six hours per trip for all individuals, it is still necessary to include this fixed amount since in the collapsible time model it will be valued differently by individuals with different time values.

Table 4.1

Mean Estimates, Biases, Standard Deviations
and Mean Square Errors of Estimated Parameters

(10 replications of 240 random drawings)

			MODEL				
	True	OLS-I	OLS-C	ML-I	ML-C	CML*	ML*
Mean Estimates							
γ	-4.00	3.66	5.04	-6.45	-.56	-6.11	-4.72
β'	-120.48	-104.68	-196.03	-166.30	-204.28	-118.22	-113.65
γ_1^+	.50	.38	.22	.60	.53	.57	.52
γ_2^+	.33	...	2.05	...	.06	.77	.43
σ	5.00	3.88	3.78	5.38	5.15	4.95	4.65
Bias							
γ	...	7.66	9.04	-2.45	3.44	-2.11	-.72
β'	...	15.80	-75.55	-45.82	-83.80	2.26	6.83
γ_1	...	-.12	-.28	.10	.03	.07	.02
γ_2	...	...	1.72	...	-.27	.44	.10
σ	...	-1.12	-1.22	.38	.15	-.05	-.35
Standard Deviations							
γ	...	1.26	3.57	5.38	3.09	5.67	2.01
β'	...	44.66	110.76	60.34	96.86	54.72	30.87
γ_1	...	.06	.06	.15	.16	.15	.05
γ_2	...	...	2.05	...	1.31	.91	.74
σ	...	.21	1.77	.74	.73	.70	.33
Mean Square Errors							
γ	...	60.26	94.47	34.95	21.38	36.60	4.56
β'	...	2244.00	17975.00	5741.00	16404.00	2999.00	999.00
γ_1	...	.02	.08	.03	.03	.03	.00
γ_2	...	...	7.16	...	1.79	1.02	.56
σ	...	1.30	4.62	.69	.56	.49	.23

\+ Because of scaling differences, estimated values for γ_1 and γ_2 are one one-thousandth of the values shown in the table.

The true demand models have three forms, conditioned on the labor supply choice:

$$\text{(14a)} \quad x = \underset{(\alpha+\gamma_2 T)}{-3.22} - \underset{(\beta'\gamma_1)}{.06}\, p + \underset{(\gamma_1)}{.0005}\, Y - \underset{(\beta'\gamma_2)}{.04}\, t + \varepsilon \quad \text{(fixed work week)}$$

$$\text{(14b)} \quad x = \underset{(\alpha+\gamma_2 T)}{-2.56} - \underset{(\beta'\gamma_1)}{.06}\, p + \underset{(\gamma_1)}{.0005}\, Y - \underset{(\beta'\gamma_2)}{.04}\, t + \varepsilon \quad \text{(unemployment)}$$

$$\text{(14c)} \quad x = \underset{(\alpha)}{-4.00} - \underset{(\beta'\gamma_1)}{.06}\, (p + w_D t) + \underset{(\gamma_1)}{.0005}\, (Y + w_D T) + \varepsilon \quad \text{(discretionary work time)}$$

where the terms in parentheses under coefficients indicate how the coefficient is related to the utility model (equations 6, 7 and 8). The available time is assumed constant over all individuals in the sample. The Y denotes the relevant income level depending on the labor market choice.

The number of trips taken by each individual is generated by the demand functions (14a), (14b) or (14c). However, if the demand function together with randomly drawn values for p, t, Y, T, w_D and E produce a negative value for x, then the individual is assumed to be a non-participant and x is set to zero. Each replication of 240 observations generated between 100 and 120 participants (i.e. observations for which $x > 0$).

For comparison purposes, estimates for the parameters α, β', γ_1, and γ_2 are obtained using five different procedures. The first two procedures (OLS-I and OLS-C) approach the problem in the traditional manner: all individuals are treated identically with respect to time valuation and only participants are included in the sample. Ordinary least squares estimates of parameters are obtained for both models. The two models differ in the way time is incorporated in the model. In the OLS-I model, everyone is assumed to value time at his wage rate. In OLS-C, time and money costs are introduced as separate variables for all individuals. To distinguish the biases which may arise due to model misspecification from those attributable to sample selection bias, a second set of estimates are obtained from a maximum likelihood formulation (ML) which corrects for the truncated sample problem but not the misspecification. All individuals are incorrectly presumed to be at interior labor market solutions in ML-I, and all individuals are incorrectly presumed to be at corner solutions in ML-C. The final estimation represents the "correct" approach in that both the truncated sample problem and the specification problem are addressed.

CML* uses exactly the same data set as OLS-1, OLS-C, ML-I and ML-C; that is, only participants are included in the sample. Similar to ML-I and ML-C, the CML* approach corrects for the truncation problem by maximizing a conditional likelihood function, conditioned on participation (see eq. 13). Unlike ML-I and ML-C, this approach also conditions the recreational demand function on the labor market decision. Finally ML* is estimated by maximizing the likelihood function in (12). The difference-between CML* and ML* is that the ML* approach includes nonparticipants. This is the preferred approach when possible, but information on nonparticipants is often not available. It should be noted that ML*, by definition, is based on a slightly different sample since it includes nonparticipants. To facilitate some manner of comparison, the sample sizes upon which the parameter estimates are based are kept the same even though some of the observations differ across approaches.

In Table 4.1, statistics on the parameter estimates from the experiment are presented. The "true" parameters (denote these $\tilde{\theta}$), those used to generate the data, are recorded in the first row. These are followed by the average parameter estimates for each technique. ($\Sigma\theta_i/10$, where θ_i is the estimated value of a parameter on the ith repetition). The parameter estimates are averaged over the ten replications; consequently, these numbers represent the sample means of the estimators for each parameter and each approach. The second part of the table presents the estimated biases for each parameter and each approach. These are the differences between the "true" parameters and the sample means of the estimates (i.e. $\sqrt{\Sigma(\theta - \theta_i)}$). Finally, mean-square errors are provided for purposes of comparison (where mean-square error is defined as bias^2 + variance). A comparison of mean square errors shows the ML* approach to be superior to all others with respect to all parameters including the standard deviation of the disturbance term. On the basis of mean square errors, the CML* approach would appear to be second best. OLS-I provides estimates of β' and γ' with slightly smaller mean square errors (although the biases are larger), but the mean square error of the OLS-I estimate of σ is considerably larger than that of CML*. Both OLS approaches produce large MSE's for σ and both approaches which presume everyone is at a corner solution (OLS-C and ML-C) produce large MSE's for the preference parameters - particularly for β'. OLS-C is the poorest performing approach uniformly. This is the approach which ignores the truncated sample problem and includes time and money costs separately in the regression. It is important to note here that no correlation between these costs was introduced when generating the data. The correlation between money and time prices which is usually found in travel cost data would likely increase the variance in these estimates.

In addition to estimating parameters for each procedure, estimated welfare measures for hypothetical price increases sufficient to eliminate the recreational activity are provided. First the "true" compensating variation for each participant in each of the ten replications is calculated. These are calculated using the formulas in equations (9) and (10) from data on the individual's number of visits, costs, etc. together with the set of true parameters. These "true" compensating variations for the i^{th} individual in the j^{th} replication are denoted CV^*_{ij}. The CV^*_{ij} is the standard by which one can compare the results of the six estimation approaches.

In Table 4.2 are the results of compensating variation calculations. For each individual, six estimated compensating variations were calculated using the estimated parameters from each of the six estimation approaches. For comparison purposes the ML* parameter estimates are applied to exactly the same sample of individuals as the other parameter estimates. This is actually to the disadvantage of the ML* approach because the parameters for this approach were estimated from a slightly different sample.

The numbers in the table represent the averages of the CV calculations over all individuals in all replications. For each approach, the bias reported in this table is the average (over all individuals in all samples) of the difference between the "true" compensating variation for an individual and his estimated compensating variation. For the entire sample, including all participants from the ten replications, the average "true" compensating variation <u>per participant</u> is $428.85. This figure reflects the following calculation: $\Sigma_j \Sigma_i CV^*_{ij} / \Sigma_j N_j$, where N_j equals the number of participants in the jth replication. The average CV's can be transformed to "per capita" values by multiplying by .46.

In comparing the average CV's calculated from the estimated parameters, it is clear that the OLS estimates are by far the worst. These estimates are between two and three times as great as the "true" average CV. The results are consistent with the <u>a priori</u> reasoning that ignoring the truncated sample problem will bias welfare measures upward.

Interestingly, the ML estimates which take account of the truncation problem but which do not incorporate the individual's labor market decisions both appear to be biased downward. Also of interest is the fact that, at <u>least in this example,</u> if one misspecifies the demand by ignoring the labor market decision, it does not seem to matter very much which of the two constructs (corner or interior solution) is applied to the sample.

Table 4.2

Mean Estimates, Biases, and Standard Deviations
of Compensating Variation Estimates

Model	Average Compensating Variation	Average Deviation From True CV	Standard Deviation around Col. 1	Standard Deviation around CV*
True	$428.85	..	624.57	..
OLS-I	1169.00	740.13	2225.64	1137.37
OLS-C	972.31	543.46	1487.73	892.57
ML-I	311.03	-117.82	453.12	275.60
ML-C	306.26	-122.59	441.39	277.86
CML*	557.13	124.28	938.60	280.18
ML*	495.75	66.91	716.80	206.36

The ML* approach produces a CV estimate which, while larger than the true average CV, is by far the best. The CML* estimate is larger, but still is within 25% of the "true" value. It is of importance that both preferable approaches yield estimates larger than the "true" average compensating variation. In the next chapter the reasons why an upward bias may be expected are explored.

It would be helpful at this point to present measures of the variance of these compensating variation estimates. However useful measures of variability are difficult to define in this case. When examining parameter estimates from each approach, sample variances of the estimates were calculated. However in the case of the estimated compensating variation, sample variances might be misleading. In the parameter case the true parameters were fixed; increasing variation in estimates of these parameters was obviously undesirable. However the true values of compensating variation;

the CV^*_{ij}'s, vary over all observations; thus the CV*'s themselves have a nonzero sample variance. Expressed in another way, one no longer has the desirable circumstance of observing several estimates drawn from the same distribution as was true with the estimated parameters.

Also in Table 4.2 two statistics which reflect variability are presented. The first is the simple standard deviation, calculated for each replication and averaged over replications. The second statistic captures elements of both bias and variability. For each replication the square root of the sum of squared deviations of CV_{ij} from CV^*_{ij} is calculated. The number reported in the table is the average of these statistics over the ten replications. The number will increase with increasing bias and/or increasing variability around the true CV.

From Table 4.2 one can see that the 0LS estimates are once again quite dismal. The standard deviations around their own means are between two and four times as great as the variation in the "true" compensating variations in the sample. In contrast, the variation in ML* is only slightly greater than the variation in the CV*'s. Both ML-I and ML-C produce estimates with smaller variances than the actual variance in the sample. This is no doubt related to the fact that these estimators under-predict CV. Thus the same percentage variation around the mean will translate into a smaller standard deviation.

The second half of the table lists the standard deviations around the true values of CV. Note that the ML* approach is still superior to all others. The poor performance of the OLS models is once again apparent.

Observations

At this point it is useful to summarize the key aspects of this chapter and elaborate on some points not fully developed in the text. Perhaps the major contribution of the chapter is the integration of the labor supply and recreational demand literature. In so doing an attempt was made to provide a coherent and general approach to the treatment of time in the context of recreational demand models used to value natural resources and environmental improvements.

The essential property of the generalized demand model incorporating time is that it is derived from a utility maximization problem with two constraints. The details of the two constraint problem are explored in the

Appendix to this chapter. The presence of two constraints causes theoretical difficulties in moving from a demand function to a utility function to obtain exact welfare measures, and as such the results of Chapter 2 can not be applied directly. While models from Chapter 2 could be modified to serve our limited purposes here, an examination of the Vartia approximation method in the presence of two constraints would likely allow greater generality in the demand function, yet preserve the ability to obtain Hicksian measures.

The two constraint case also has interesting implications for welfare measurement. The utility maximization problem now admits of two duals, i.e. two expenditure functions and two compensated demand functions. This implies that the welfare effects of a policy change can now be measured in either (or a combination) of two standards - money or time. The implications of this dual standard are investigated elsewhere (see Bockstael and Strand, 1985).

The illustration in this chapter focuses on the traditional money compensating variation measures and explores the biases which can arise in the estimates of preference parameters and compensating variation by using a misspecified demand function. While Monte Carlo type examples are never completely conclusive, the experiments suggest wide disparities in CV estimates when different estimation approaches are used. Compared to the correctly specified approaches which also account for the truncated sample problem (the ML* and CML* approaches), the conventional OLS approaches produce upwardly biased estimates of CV with large variances around their own mean and around the true CV values. Maximum likelihood estimates which account for truncation but not misspecification of the time-price variable appear to be downwardly biased. The ML* estimate is much preferred with relatively small variance and deviations from the true value of CV.

Both ML* and CML*, although calculated from presumably consistent parameter estimates, produce CV estimates which on average exceed the true CV's. In the next chapter it is demonstrated why compensating variations, even when calculated from unbiased parameters, may themselves be upwardly biased

FOOTNOTES TO CHAPTER 4

[1] In fact, the wage rate may not even serve as an upper or lower bound on the individual's marginal valuation of time when labor time is institutionally restricted. That is, an individual who chooses to be unemployed may simply value his marginal leisure hour more than the wage rate, or he may value it less but not be better off accepting a job requiring 40 hours of work per week. If restricted to an all-or-nothing decision, 40 hours may be less desirable than O. An individual at a point such as A, however, may value the marginal leisure hour at more than w_p but choose 40 rather than 0 hours. Alternatively he may value leisure time at less than w_p but more than the wage he could earn for additional hours by working a secondary job.

APPENDIX 4.1

A COMPARATIVE STATICS ANALYSIS OF THE TWO CONSTRAINT CASE*

The subject of this Appendix is the consumer choice problem with two constraints. As we saw in Chapter 4, labor market restrictions and labor-leisure preferences cause individuals to be either at interior or corner solutions in the labor market. Classic comparative statics and welfare evaluation is directly applicable to interior solutions as the time and income constraints collapse into one. However the comparative statics and duality results associated with the corner solution case (i.e. utility maximization subject to time and income constraints) have received little attention.[1]

The first treatment of the problem was by A. C. DeSerpa (1971). Suzanne Holt's (1984) paper is the only other which explicitly deals with comparative statics of the time and income constraint. Both Holt's approach and that of DeSerpa's involves inversion of the Hessian, a tedious and difficult task for problems with large dimensionality. The Slutsky equation derived from this approach includes cofactors of the Hessian and, as such, is a complex function of the decision variables in the system. In what follows, a more modern approach is employed based on the saddle point theorem, as proposed by Akira Takayama (1977). Making use of the envelope theorem, this approach is simple to apply and far more revealing. From it can be derived Slutsky equations containing elements with clear economic interpretations.

This Appendix goes beyond the previous work by examining duality results and demand function properties in the context of the two constraints. Several new time analogs to the well known results in traditional demand theory are presented. Specifically, we derive a time analog to Roy's Identity and two generalized Slutsky equations. These Slutsky equations

* This appendix is the work of Terrence P. Smith, Agricultural and Resource Economics Department, University of Maryland.

which describe the effect of a change in a money price are similar to the traditional Slutsky equation but contain additional income (time) effect terms which describe how demand responds indirectly to income (time) changes through the trade-off between time and money in producing utility.

Utility Maximization with Two Linear Constraints

Consider the household who maximizes a utility function, $U(x)$, where x is a vector of activities that produce utility. These activities need not be actual market commodities. The link to the market is through a set of household production functions. Suppose that the household produces these activities, x, according to the non-joint production functions, $f_i(s_i,v_i)$ where s_i and v_i represent a vector of purchased goods and time inputs into the production of x_i. The problem, then is to

(A1) $\max U(x)$ subject to $x_i = f_i(s_i,v_i)$ for all i, and

$$Y = R + wT_w = \Sigma r_i s_i, \text{ and}$$

$$\bar{T} = T_w + T = T_w + \Sigma v_i,$$

where Y is total income, the sum of nonearned income R and wage income wT_w, and r_i is a vector of money prices corresponding to the vector s_i. To proceed to specific results, a fixed coefficients Leontief technology is assumed, that is, a technology with no substitution possibilities between the purchased inputs and time. This assumption implies that the activities, x_i, have fixed money and time costs, representable as the scalars, p_i and t_i.

As has been explained in the body of this chapter, the problem in (1) can take two forms. If work time is an endogenous variable, i.e. a decision variable of the individual who can choose T_w freely, then the two constraints in the problem collapse to one:

$$R + w\bar{T} = \Sigma(p_i + w_{ti})x_i.$$

In this case the problem is structurally similar to any other one constraint problem. If, as will be assumed in this appendix, work time is institutionally constrained, then T_w can be treated as fixed and two relevant and separate constraints remain. The problem then can be rewritten

(A2) $\max U(x)$ subject to $Y = p'x$ and $T = t'x$.

where $T = \bar{T} - T_w$.

$U(x)$ is a twice continuously differentiable concave utility function with x an n-dimensional vector of commodities. The consumer behaves so as to maximize this utility function. There is a commodity, say x_k, which represents savings such that the income constraint is always satisfied, and there is another commodity, say x_i, which is uncommitted leisure time such that the time constraint is effective.

Since the objective function is differentiable and concave in x, the constraints differentiable and linear in x and b, where $b=(p,t,Y,T)$, the constraint qualification and curvature conditions are met. This implies that, if a solution exists, then the quasi-saddle point (QSP) conditions of Takayama (1973) will be both necessary and sufficient. Also, note that, given the assumption of the existence of slack variables, savings and uncommitted leisure time, the constraints are effective, and if a solution exists it will be an interior one. Collectively, these conditions allow the application of the envelope theorem to our problem.

If a solution to (2) exists, it will be of the form $x(b)$, $\theta(b)$, $\phi(b)$. Hence we may substitute these solutions into the original Lagrangian to obtain

(A3) $$L(b) = U(x(b)) + \phi(b)\,[Y - px(b)] + \theta(b)\,[T - tx(b)].$$

Now $U(x(b))$ may be written as $V(p,t,Y,T)$ and interpreted in the usual way as the indirect utility function. Note that, in addition to the traditional parameters affecting indirect utility (prices, p, and income, Y), the time prices, t, and time endowment, T, are also relevant parameters. Applying the envelope theorem to the above we obtain

(A4a) $$\partial V(p,t,Y,T)/\partial Y = \phi(p,t,Y,T)$$

(A4b) $$\partial V(p,t,Y,T,)/\partial T = \theta(p,t,Y,T)$$

(A4c) $$\partial V(p,t,Y,T)/\partial p_i = -\phi(p,t,Y,T)\; x_i(p,t,Y,T)$$

(A4d) $$\partial V(p,t,Y,T)/\partial t_i = -\theta(p,t,Y,T)\; x_i(p,t,Y,T).$$

Combining (4a) and (4c) gives Roy's Identity, viz.,

(A5) $$\frac{\partial V(p,t,Y,T)/\partial p_i}{\partial V(p,t,Y,T)/\partial Y} = x_i \text{ for all } i.$$

Likewise, combining (4b) and (4d) gives analogous identity, viz.,

(A6) $$\frac{\partial V(p,t,Y,T)/\partial t_i}{\partial V(p,t,Y,T)/\partial T} = x_i \text{ for all } i.$$

Note that (6) gives an alternative way to recover the Marshallian demand from the indirect utility function. However, both differential equations may be required to be solved to recover the indirect utility function from the demand function, since it will be shown that there are two expenditure functions.

These envelope results can be manipulated in other ways to demonstrate time extensions to traditional demand analysis. For example, combining (4a) and (4b) with (4c) and (4d) we obtain

(A7) $$\frac{\partial V(p,t,Y,T)/\partial T}{\partial V(p,t,Y,T)/\partial Y} = \frac{\Theta(p,t,Y,T)}{\phi(p,t,Y,T)} = \frac{\partial V(p,t,Y,T)/\partial t_i}{\partial V(p,t,Y,T)/\partial p_i}$$

which is McConnell's m_T or "the opportunity cost of scarce time measured in dollars of income." Multiplying (4c) by p_i, (4d) by t_i and summing over all i yields

(A8) $$\sum p_i \partial V/\partial p_i + \sum t_i \partial V/\partial t_i = -\phi \sum p_i x_i - \Theta \sum t_i x_i$$

which by (4a) and (4b) implies

(A9) $$\sum p_i \partial V/\partial p_i + \sum t_i \partial V/\partial t_i + Y \partial V/\partial Y + T \partial V/\partial T = 0,$$

so that the indirect utility function V(p,t,Y,T) is homogeneous of degree 0 in money and time prices, income, and time.

The Two Duals and the Two Slutsky Equations

In this section the dual of the utility maximization problem is explored. Since there are two constraints, there are two duals to the problem. The first is (money) cost minimization subject to constraints on time and utility; the second is time cost minimization subject to constraints on income and utility. This exploration yields two expenditure functions, an income compensated function and a time compensated function. The existence of two expenditure functions allows one to compute welfare changes either in the traditional way as income compensation measures or, alternatively, as time compensation measures.

In addition, these expenditure functions are combined with the envelope theorem to reveal two generalized Slutsky equations. The first of these describes how Marshallian demand responds to money price changes and the second how the ordinary demand changes with a change in time prices. The manner of proof is in the style of the "instant Slutsky equation" as first introduced by Cook (1972).

The duals to the utility maximization problem (2) are

(A10) $$\min_{x} px \quad \text{subject to } T = tx \text{ and } U^{0} = U(x)$$

and

(A11) $$\min_{x} tx \quad \text{subject to } Y = px \text{ and } U^{0} = U(x)$$

where U^0 is some reference level of utility.

Notice that (10) and (11) can be cast in the notation of our original maximization problem, where the objective functions, px and tx, are linear and hence concave in x and p or t, and the constraint functions are quasi-concave since the first constraint is linear (either T - tx = O or Y - px = O) and the second, concave. It follows then, as in our earlier analysis of the primal problem, that if a solution exists, the QSP conditions will be both necessary and sufficient. Furthermore, maintaining the existence of the slack variables, savings and freely disposable time, and requiring that the reference level of utility be maintained ensures that the constraints are effective, that we have an interior solution, and hence, that the envelope theorem may be applied.

Consider, then, the two Lagrangians,

(A12a) $$\min L_Y(p,t,T,U^0) = px + \lambda(T - tx) + \gamma(U^0 - U(x))$$

and

(A12b) $$\min L_T(p,t,Y,U^0) = tx + \mu(Y - px) + \delta(U^0 - U(x)).$$

Solutions to these minimization problems, if they exist, are given by,

(A13a) $$x_Y(p,t,T,U^o)$$

and

(A13b) $$x_T(p,t,Y,U^o).$$

The first of these is the "usual" Hicksian income compensated demand, while (13b) is an analogous time compensated Hicksian demand. Of course, both depend (in general) on all money (p) and time (t) prices.

Solutions (13a) and (13b), when substituted back into the objective functions, imply the existence of two expenditure functions. The first of these,

(A14a) $$E_Y(p,t,T,U^o) = px_Y(p,t,T,U^o)$$

is the well known classical expenditure function with the exception that the time prices, t, and the time endowment, T, appear as arguments.

The second,

(A14b) $$E_T(p,t,Y,U^o) = tx_T(p,t,Y,U^o),$$

is a time compensated measure of the minimum expenditure level necessary to maintain U^o. Either (14a) or (14b) may be used to measure welfare effects of a change in money or time prices or both. The novelty of using (14b) for welfare analysis is that it measures the amount of time compensation, rather than income compensation, necessary to maintain a reference utility level in the face of, say, a money price change for one of the commodities.

Since the two expenditure functions, E_Y and E_T, are concave, then if these expenditure functions are twice differentiable the matrix of second derivatives is negative semidefinite. Therefore the slopes of the (compensated) own money price and own time price demands are necessarily non positive. Also, define $S_{ij} \equiv \partial x_i/\partial p_j = \partial^2 E_Y/\partial p_j \partial p_i$, as the money substitution effect for good i, given a price change for good j, and $T_{ij} \equiv \partial x_i/\partial t_j = \partial^2 E_T/\partial t_i \partial t_j$ as the time substitution effect. Then it follows that S and T are negative semidefinite and symmetric, and since x_Y and x_T are homogeneous of degree 0 in p and t, respectively, then

(A15a) $$\sum_i p_i \partial x_i/\partial p_j = 0 = \sum_i S_{ij}p_i$$

and

(A15b) $$\sum_i t_i \partial x_i/\partial t_j = 0 = \sum_i T_{ij}t_i.$$

That is, the aggregation conditions hold. Finally, note that by the envelope theorem

(A16a) $$\partial E_Y/\partial p_i = x_{Yi}(p,t,T,U^0) \text{ and}$$

(A16b) $$\partial E_T/\partial t_i = x_{Ti}\,(p,t,Y,U^0),$$

(Shepard's Lemma).

The above serves to formalize the equivalence of several of the well known properties of Hicksian demands in the classical and two constraint systems. The Slutsky relations that follow from the present problem are now derived. Although our results show structural similarity to the classical equations, our derivation results in two Slutsky equations, each of which has a time effect as well as an income effect.[2]

Consider the solution to the primal problem posed in the preceding section. This solution is the set of Marshallian demands which may be written,

(A17) $$x^m = m(p,t,Y,T).$$

Now recall that the solution to our money minimization problem, Y, is just $p'x_Y(p,t,T,U)=E_Y$, and likewise, the solution to the time minimization probem, T, is defined as $t'x_T(p,t,Y,U) = E_T$. Hence

(A18) $$x = m[p,t,E_y(p,t,T,U),\ E_T(p,t,Y,U)].$$

Note that (18) now defines the set of Hicksian demands. Differentiating (18) with respect to the j^{th} price, p_j, gives

(A19a) $$\partial x_i/\partial p_j = \partial m/\partial p_j + (\partial m/\partial E_Y)\ (\partial E_Y/\partial_{pj}) + (\partial m/\partial E_T)\ (\partial E_T/\partial_{pj})$$

using the chain rule. Consider also how the demand for x_i changes with a change in one of the time prices, say t_j. Differentiating (18) with respect to t_j yields,

(A19b) $\quad \partial x_i/\partial t_j = \partial m/\partial t_j + (\partial m/\partial E_Y)(\partial E_Y/\partial t_j) + (\partial m/\partial E_T)(\partial E_T/\partial t_j).$

These are the two generalized Slutsky equations that result from the dual constraint problem. To cast them in more familiar terms use the envelope theorem applied to equations (12) to obtain,

(A20a) $\quad \partial E_Y/\partial p_j = x_j$

(A20b) $\quad \partial E_T/\partial t_j = x_j$

(A20c) $\quad \partial E_Y/\partial T = \lambda$

(A20d) $\quad \partial E_T/\partial Y = \mu$

(A20e) $\quad \partial E_Y/\partial t_j = -\lambda x_j$

(A20f) $\quad \partial E_T/\partial p_j = -\mu x_j.$

Substituting (20a) and (20f) into (19a) and rearranging, obtains the money price Slutsky equation,

(A21a) $\quad \partial x_i^m/\partial p_j = \partial x_i^h/\partial p_j - x_j\,[\partial x_i^m/\partial Y \quad \mu\partial x_i^m/\partial T],$

where x_i^m denotes Marshallian functions and x_i^h denotes Hicksian functions. This Slutsky equation is identical to the classical version with the exception of the additional term $\mu x_j \partial x_i^m/\partial T$, which is the indirect effect of income through time. If x_i is an income normal, time normal good, then $\partial x_i^m/\partial Y$ and $\partial x_i^m/\partial T$ are both positive, and since μ, the Lagrangian multiplier on the income constraint in the time minimization problem, is necessarily non-positive, it follows that for a "normal-normal" good the "income" effect[3] is enhanced relative to the classic income effect.

Proceeding in exactly the same way, the time price Slutsky equation can be derived. Substituting (20b) and (20c) into the second Slutsky equation (19b) and rearranging, yields

(A21b) $\quad \partial x_i^m/\partial t_j = \partial x_i^h/\partial t_j - x_j\,[\partial x_i^m/\partial T - \lambda\,\partial x_i^m/\partial Y].$

Notice that in addition to the "pure" time effect, $x_j \partial x_i/\partial T$, there is an additional indirect effect, $\lambda x_j \partial x_i/\partial Y$, which, using the same argument as above, is an indirect time effect through income, converted to time by the marginal (time) cost of income (λ). Again the two terms will augment one another for a "normal-normal" good, and, of course, offset one another for a "normal-inferior" good, where "normal-inferior" is taken to represent a commodity which is income normal and time inferior or vice versa.

Utilizing the results that $\mu = \partial E_T/\partial Y = \partial T/\partial Y$ and $\lambda = \partial E_Y/\partial T = \partial Y/\partial T$, an equivalent way of writing (21) is

(A22a) $\partial x_i{}^m/\partial p_j = \partial x_i{}^h/\partial p_j - x_j[\partial x_i{}^m/\partial Y - (\partial x_i^m/\partial T)\ (\partial T/\partial Y)]$

(A22b) $\partial x_i{}^m/\partial t_j = \partial x_i{}^h/\partial t_j - x_j[\partial x_i^m/\partial T - (\partial x_i^m/\partial Y)\ (\partial Y/\partial T)]$.

This version makes clear the substitution between income and time in the two constraint model.

A Summary of Results

The "usual" properties of classical demand functions still hold when one solves the two constraint problem. The demand functions that solve our maximization problem are homogeneous of degree O in money and time prices, income and time, and satisfy the aggregation and integrability conditions. The compensated demands, be they income or time compensated, are own price (money or time) downward sloping. The "substitution" matrix is negative semidefinite, where the substitution matrix must be interpreted as the matrix which describes a response to a money (time) price change holding utility and the time (income) endowment constant. Finally, we can partition the ordinary demand response to a change in money (time) price as made up of two effects, a utility held constant effect, i.e. a movement along an indifference surface, and an income (time) effect, remembering the complication, however, that this income (time) effect is made up of a "pure" income (time) effect and an indirect effect of time (income) converted to money (time) terms.

These new demand functions contain additional arguments relative to the "classic" demand function. That is, the ordinary demands are functions of not only money prices and income, but also of time prices and of the time endowment. Likewise, the money and time expenditure functions depend not only on money prices and utility, but also upon time prices, and the time endowment (for the money expenditure function) or income endowment (for the time expenditure function). Therefore, welfare analysis may be done in a straightforward way using these expenditure functions provided we account not only for money and income changes but also for time price and time endowment changes.

One final result is of particular interest. The Slutsky equations (22a) and (22b) indicate a two term income effect for the money price version and a two term time effect for the time price equation. Restating the Slutsky equation for our own money price change,

$$\frac{\partial x_i^m}{\partial p_i} = \frac{\partial x_i}{\partial p_i} - x_i \frac{\partial x_i^m}{\partial Y} + x_i \frac{\partial x_i^m}{\partial T} \frac{\partial E_T}{\partial Y}.$$

The left hand side variable is the Marshallian price slope. The first term on the right is the Hicksian price slope. The total income effect is made up of the usual income effect term $-x_i \partial x^m/\partial Y$ and the effect of income through time effect $x_i(\partial x_i^m/\partial T)(\partial E_T/\partial Y)$. Both terms are negative if x_i is normal with respect to Y and T, because $\partial E_T/\partial Y$ is negative and represents the change in time costs necessary to achieve a given level of utility if the individual is given more income.

From this expression it is clear that the total (combined) income effect is greater in absolute value than the conventional (direct) effect. This has the interesting result of pushing compensated and ordinary demand functions farther away from each other.

This divergence between the Marshallian and Hicksian demands implies that the consumer surplus measure will be decreased and the compensting variation measure increased. Hence the use of the consumer's surplus welfare measure to approximate the theoretically correct compensating variation is made less defensible. It would seem useful to reexamine the Willig bounds on using the consumer's surplus as a welfare measure in light of these implications.

Whether a good is time normal or time inferior is not altogether obvious. One could develop examples which would suggest either case. It seems, that this is likely to be an important question for recreational goods along with the question of whether or not an individual's work time is fixed.

FOOTNOTES TO APPENDIX 4.1

1 The solution to, and sensitivity analysis of, a more general problem, i.e. maximization of an objective function subject to multiple, possibly nonlinear, constraints has appeared in the mathematical economics literature.

2 The similarity can also be seen in the approach of DeSerpa and Holt. Unfortunately, that approach, which relies on the inverted Hessian, tends to obscure the detail of the time and income effects.

3 The interpretation of μ is the marginal (money) cost of time, hence μ converts the time effect into income units, and therefore the second term in brackets may be interpreted as an additional income effect.

CHAPTER 5

THE CALCULATION OF CONSUMER BENEFITS

Until this point, emphasis has been placed on obtaining unbiased and consistent parameter estimates of the structural model of behavior. Developments have been made in the creation of models consistent with utility theory, in introducing realistic time constraints on recreational behavior, and in establishing appropriate estimation techniques. These efforts have all been directed to obtaining the relevant parameters of recreational preference functions. It has implicitly been presumed that consistent preference parameter estimates together with correct formulas for ordinary surplus and Hicksian variation measures will automatically produce unambiguous, consistent estimates of these welfare measures. In this chapter two aspects of the calculation of welfare measures from estimated preference parameters are examined.

Despite the scores of articles containing surplus estimates, only a few (e.g. Gum and Martin, 1975) have devoted even modest attention to the procedure for calculating benefits from estimated equations. Most studies presumably follow the process outlined by Gum and Martin, although Menz and Hilton (1983) indicate other ways of calculating benefits from a zonal approach. This “procedure” for calculating welfare efforts from estimated coefficients is the first aspect of consideration. The second is the explicit recognition of the fact that benefit estimates are computed from coefficients with a random component and therefore possess statistical properties in their own right. To our knowledge, no one in the recreational demand literature has been concerned with this.

The beginning of this chapter considers the common sources of regression error and the statistical properties of benefit estimates which arise because of that error. Three common sources are considered: omission of some explanatory variables, errors in measuring the dependent variable, and randomness of consumer behavior. For each, the procedures one would employ to obtain estimates of ordinary consumer surplus and examine the statistical properties of estimates derived following these procedures are outlined. Similar results would be true of CV and EV measures, but the derivations are considerably more difficult. The two familiar functional forms referred to frequently in the last few chapters, the linear and the semi-log specification, are used for illustration.

The general results are at first alarming. The expected value of consumer surplus seems to depend on the source of the error. Error from the common assumption of omitted variables leads to higher expected benefits than that from other error sources. Secondly, benefit estimates calculated in the conventional way are generally upwardly biased when they are based on small samples. The expected value of consumer surplus based on maximum likelihood estimates exceeds the true surplus values. All is not lost, however. The benefit estimates are, at least, consistent. Perhaps of greater importance, minimum expected loss (MELO) consumer surplus estimators with superior small sample properties are available.

The mathematical derivations are specific to the unbiased, maximum likelihood estimators and ordinary surplus calculations. Nonetheless, the specific results of this chapter 'are supported by more general theorems, and the message remains relevant whenever the welfare measures of interest are nonlinear functions of estimated parameters.

Sources of Error in the Recreation Demand Model

Discussions of the sources of error in recreation demand analysis are common in the existing literature. The most traditional line of thought (e.g. Gum and Martin, 1975) considers the error component in predicting the individual's recreation behavior to arise from unmeasured socio-economic factors. Others (e.g. Hanemann, 1983a) attribute at least some of this error to fundamental randomness in human behavior. Applied statisticians (e.g. Hiett and Worrall, 1977) on the other hand, suggest that recall of annual number of recreational trips (i.e. the quantity demanded) is subject to substantial error. Still others (e.g. Brown *et al*., 1983) have argued that recall of explanatory variables, such as travel expenses, contains error.

The several explanations for the stochastic term in econometric models which have been proffered by econometricians are made explicit below:

(1) Omitted variables: factors which influence recreational demand have not been introduced and, thus, error-free explanation of recreation demand is not possible.

(2) Human indeterminacy: behavior, even with all explanatory variables included and measured perfectly, cannot be predicted because of inherent randomness in preferences;
(3) Measurement error I: exact measurement of the dependent variable is not possible; and
(4) Measurement error II: exact measurement of the independent variable is not possible.

Each explanation has a particular relevance for welfare analysis. Yet only the first three sources of error conform to the Gauss-Markov assumptions, and then only if the omitted variables are assumed to be uncorrelated with included variables. Thus, the same estimation procedure (e.g. ordinary least-squares analysis) will be appropriate if the error is associated with (1) through (3) but not with (4). The fourth explanation violates the assumed independence between the error and explanatory variables. When such violations are expected, estimation techniques such as instrumental variables are frequently employed. However, these methods will generate different coefficient estimates from the other three. As such, meaningful comparisons between cases (1) through (3) on the one hand and (4) are nearly impossible to make. Discussion is thus restricted to consideration of (1) through (3) and throughout most of the chapter the error is assumed independent of included variables.

Two functional forms of individual demand are postulated here, each of which is consistent with utility maximizing behavior (see Hanemann, 1982d):

(1) $$x_i = \alpha + \beta p_i + \gamma y_i + u_i$$

and

(2) $$\ln x_i = \alpha + \beta p_i + \gamma y_i + u_i.$$

In each specification, x_i is the i^{th} individual's demand for the good in question, p_i is the price he faces for the good, and y_i is his income. Both p and y are normalized on the price of the numeraire good. The parameters α, β, and γ are preference function parameters. As is usual, all individuals are assumed to face different explanatory variables but to possess the same general form of preferences, except for random differences, so that preference function parameters are constant over the population.

The u_i in (1) and (2) are the disturbance terms which arise from the sources of error described above. Consistent with Gauss-Markov assumptions, u_i is assumed to be distributed normally with a mean of zero and constant variance which is denoted by σ^2, irrespective of the source of error. Additionally $E(u_i u_j) = 0$ for $i \neq j$, $E(p_i u_i) = 0$, and $E(y_i u_i) = 0$.

In most econometric applications, the source of the disturbance term or "error" is immaterial as long as Gauss-Markov assumptions hold. These conditions are sufficient to produce unbiased and efficient estimates of α, β, and γ. However, if the ultimate purpose of the estimation exercise is to compute consumer surplus estimates, then the story does not end here.

"True" Consumer Surplus

In this section, expressions for the value of consumer surplus are derived under the competing assumptions that the randomness is due to omitted variables, randomness in preferences, or errors in measurement. These expected values are determined on the premise that the coefficients of the demand equations are known with certainty, so that these coefficients do not embody any random element. In a later section, the discussion is extended to the case when consumer surplus is calculated from estimated coefficients.

Suppose that one knows with certainty the coefficients, α, β, and γ, which are common to all individuals and wishes to calculate the consumer surplus associated with a change from p_i^0 to $\tilde{p}_i$ (the price which drives individual i's demand to zero).[1] For each individual, consumer surplus will be determined by his relevant demand curve and his initial circumstances. Clearly an individual's observed price-quantity combination $(\bar{p}_i^0, \bar{x}_i^0)$ will not in general lie on the systematic portion of demand function $x^* = \alpha + \beta p + \gamma y$ because of the random component. The question is: does one calculate consumer surplus based on a demand curve drawn through the observed x and p combination (p_i^0, x_i^0) with slope β or do we base it on the systematic portion of the demand curve evaluated at (p_i^0, x_i^*)? It would seem these two methods have been used somewhat interchangeably in practice, without too much thought. Does the method make a difference in consumer surplus calculations? If so, what explanations of the error source are consistent with each usage?

1. Omitted Variables Case

Consider first the case in which the randomness across individuals derives from a relevant variable being omitted from the equation. (This variable is not correlated with the other explanatory variables). In this case it would make sense to use the demand curve drawn through (p_i^o, x_i^o), since the random term will represent an unknown component of the price intercept and thus will shift the systematic portion of the demand curve sufficiently to pass it through the observed price-quantity point. Gum and Martin's procedure seems consistent with this as it "utilizes the actual number of trips taken by a household and the actual average variable costs per trip to define the household's individual demand curve." An implicit assumption is that the omitted variables remain the same as price drives the individual from the market. Thus, the individual's true error, u_i, remains constant. The individual's "true" consumer surplus, if values of necessary variables and parameters are known with certainty, is

$$CS_{1i}^{*} = \int_{p_i^o}^{\tilde{p}_i} x_i(p_i)dp_i = \frac{(\alpha + \beta p_i^o + \gamma y_i^o + u_i)^2}{-2\beta} = \frac{(x_i^o)^2}{-2\beta} \tag{3a}$$

for a linear demand curve and

$$CS_{2i}^{*} = \int_{p_i^o}^{\tilde{p}_i} x_i(p_i)dp_i = \frac{(\alpha + \beta p_i^o + \gamma y_i^o + u_i)}{-\beta} = \frac{x_i^o}{-\beta} \tag{3b}$$

for a semi-log demand curve.

2. Random Preferences and Errors in Measurement

Two other explanations for error in regression analysis are considered: a) the individual's preferences vary randomly and b) the dependent variable (trips) is measured inaccurately. The first explanation has been used extensively in the literature (see, for example, Hausman 1981) and the latter has been studied by professional sample-gathering firms (e.g. Hiett and Morrall, 1977).

In both of these cases, it is the value of the systematic portion of the demand function $(x_i^* = \alpha + \beta p_i^o + \gamma y_i^o)$ instead of the observed value (x_i^o) which is relevant to the measurement of surplus. If the consumer has random preferences, then one cannot be certain that the observed value of x_i^o will be chosen by the i^{th} individual each time the same price-income situation arises. The "best guess" at the level of x_i consumed by the individual facing the price-income situation (p_i^o, y_i^o) is the systematic portion of demand x_i^*. When the error occurs because the individual cannot remember the exact number of trips or intentionally misrepresents his consumption level, then once again the "best guess" of the actual number of trips is the systematic demand x_i^*.

For the case in which these types of errors completely dominate, the individual's quantity demanded can be expected to be $x_i^* = \alpha + \beta p_i^o + \gamma y_i^o$. The consumer surplus for the linear demand is mathematically represented by

$$\text{(4a)} \qquad CS_{3i}^* = \int_{p_i^o}^{\tilde{p}_i} (\alpha + \beta p_i^o + \gamma y_i^o) dp_i = \frac{(\alpha + \beta p_i^o + \gamma y_i^o)^2}{-2\beta} = \frac{(x_i^*)^2}{-2\beta} .$$

For the semi-log demand function, the prediction of x_i is not an unambigious issue, but for the time being, let us use the systematic portion of demand given by $\exp(\alpha + \beta p_i^o + \gamma y_i^o)$, such that the consumer surplus is measured by

$$\text{(4b)} \qquad CS_{4i}^* = \int_{p_i^o}^{\tilde{p}_i} \exp(\alpha + \beta p_i^o + \gamma y_i^o) dp_i = \frac{\exp(\alpha + \beta p_i^o + \gamma y_i^o)}{-\beta} = \frac{x_i^*}{-\beta} .$$

Graphical Comparison of Surplus Computation and an Empirical Demonstration

Figure 5.1 is presented to recapitulate the argument and also to display visually the process of computing surplus with different sources of error. Although the disturbance term may include all three types of "error", the procedure chosen to calculate the consumer surplus implies a specific interpretation of the error term. When all error is implicitly assumed to be due to omitted variables (that is, when consumer surplus is calculated from a demand curve which is drawn through the observed price-quantity point (x_i^o, p^o)), the residual is treated as part of the constant term. In contrast, consumer surplus calculated from the demand curve which passes through (x^*, p^o) implies an error in measurement or random preferences interpretation. In this case, the error term represents the correction factor in the observed x_i value.

Two individuals facing the same price but with opposite and equal disturbance terms (u_i and u_k) are depicted in the graph. The points (x_i^o, p^o) and (x_k^o, p^o) represent their observed quantity-price points as well as their actual quantity-price combinations if omission of variables created the disturbance. To obtain the surplus, the price slope coefficient (β) is used to determine $\tilde{p}_i$ and $\tilde{p}_k$ and the surpluses $\Delta p^o B\tilde{p}_i$ and $\Delta p^o C\tilde{p}_k$. On the other hand, the point (x^*, p^o) represents the appropriate quantity-price for both individuals if the disturbance term is generated entirely by mismeasurement of x or random preferences. The appropriate surplus is then $\Delta p^o A\tilde{p}$ for both individuals. It may seem that these two alternative procedures will produce the same consumer surplus, on average. However, the graph illustrates, at least in the linear case, that they will not. The average of surpluses at x_i and x_k is larger than the surplus at x^*.

To demonstrate the different computation methods and to illustrate the degree to which the error assumptions actually cause differences in estimates of consumer surplus, consumer surplus for a sample of sportfishermen is estimated. The data set is the same one used in Chapter 2 to demonstrate differences due to functional form.

Because appropriate wage information for a more complex model incorporating treatment of time and nonparticipation such as the one in Chapter 4 is not contained in this data set, the same model and parameter estimates as shown in McConnell and Strand are presented. The individual is viewed in this model as being unaffected by institutional constraints in the labor market and therefore at the margin in labor-leisure decisions. Thus, fishermen are assumed to choose the hours they work and to make marginal tradeoffs between leisure and labor time.

The McConnell-Strand model yields the following estimated demand function (p. 154):

$$x_i = 9.77 - \underset{(3.89)}{.0206\, p_i} - \underset{(-2.00)}{.0126\, w_i t_i} + \underset{(2.50)}{1.90\, s_i} + \underset{(5.06)}{.157\, m_i}$$

where the numbers in parentheses are t-ratios, x_i is the number of annual sportfishing trips for the i^{th} angler, p_i is the i^{th} angler's trip expenses, t_i is his round trip travel time (computed as round-trip distance/45 mph), w_i is his hourly income (computed as annual personal income/2080 hours), s_i is a site dummy for the Ocean City resort, and m_i is the length of the angler's boat. The standard error of the estimate ($\hat{\sigma}$) is 6.00 trips/person, the F-statistic (4,411) is 12.8, and the $\bar{R}^2$ is .10.

The process depicted in Figure 5.1 is used to compute the competing surplus estimates. The first estimate, $\hat{C}S_3$ calculates the predicted surplus as the area behind the estimated demand function and above observed price. The line passes through predicted trips ($\hat{x}_i = x_i - \hat{u}_i$). The second estimate, $C\hat{S}_1$, represents the area behind the regression line after it is shifted to pass through the observed price and quantity (x_i^o, p_i^o).

For the entire sample, the omitted variable estimate ($C\hat{S}_1$) is calculated to be $801,274 or an average of $1,931 per fisherman. The error in measurement estimate ($C\hat{S}_3$) is calculated to be $450,086 for the sample or an average of $1084 per fisherman. Thus the assumption of omitted variable error increased the estimated average surplus by $847 (or 78%) relative to the measurement error assumption.

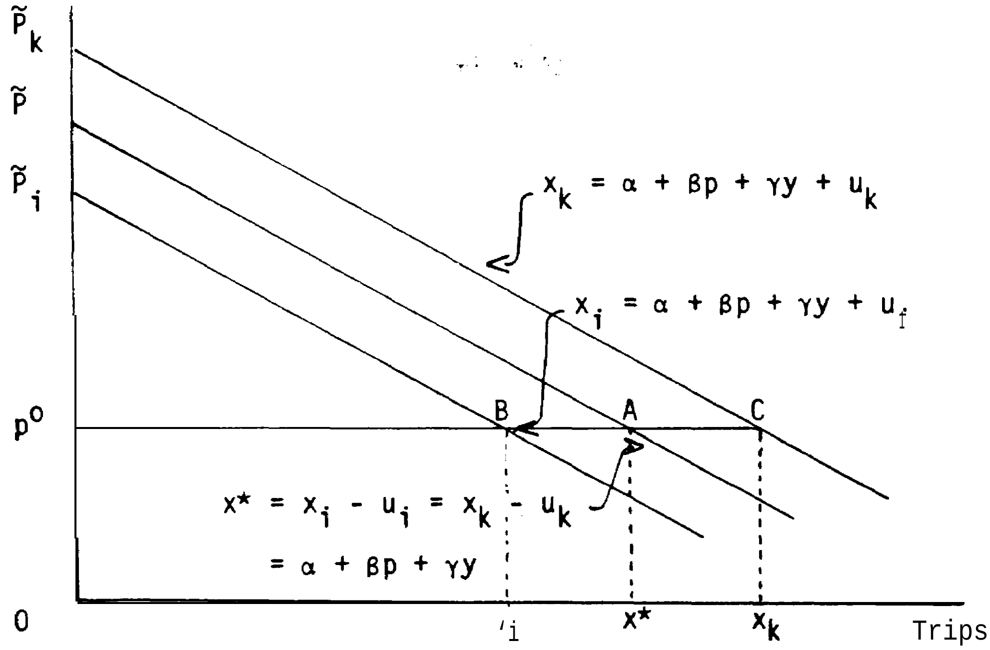


Figure 5.1

Two Different Procedures for Calculating Consumer Surplus

Calculating Expected Consumer Surplus

The graphical analysis and the empirical example demonstrate that consumer surplus calculations for an individual will differ depending on the error assumption. The analysis also suggests that these differences in consumer surplus calculations may not cancel out (as do the errors themselves) when aggregated over the sample. In order to determine the general conditions under which these differences in surplus arise it is necessary to consider expressions for expected consumer surplus (conditioned on explanatory variables), since the expected value is conceptually equivalent to the average over the sample.

Once again assume that the parameters (α, β, and γ) are known and that the expected surplus conditioned on values of p and y is to be calculated. It is obvious but nonetheless worth noting that the expected consumption level E[x] must be equal under the competing error source assumptions if the regression equation is linear. Consider first the expected consumption level if the error is assumed to arise from omitted variables:

(5a) $$E[x] = E[\alpha + \beta p + \gamma y + u] = \alpha + \beta p + \gamma y;$$

and if the error arises from measurement or random preference:

(5b) $$E[x] = E[\alpha + \beta p + \gamma y] = \alpha + \beta p + \gamma y.$$

Clearly the two are equal.

While expected consumption levels are equal, the expected value of consumer surplus will not be. Denote f(x) as the consumer surplus operator; then equality in expected consumer surplus requires that

(6) $$E[f(\alpha + \beta p + \gamma y + u)] = E[f(\alpha + \beta p + \gamma y)].$$

Note that $\alpha + \beta p + \gamma y$ does not include a stochastic term, so that the right hand side of (6) equals $f(\alpha + \beta p + \gamma y)$. Also since $\alpha + \beta p + \gamma y = E(\alpha + \beta p + \gamma y + u)$, the right hand side of (6) could be written as $f(E[\alpha + Bp + \gamma y + u])$ so that the condition in (6) can be rewritten as

(7) $$E[f(\alpha + \beta p + \gamma y + u)] = f(E[\alpha + \beta p + \gamma y + u]).$$

Jensen's inequality (Mood, Graybill and Bees, 1963) states that if q is a random variable and f(q) is a convex function, then $E[f(q)] \geq f(E[q])$. It is expected therefore that if the consumer surplus operator is a convex function then the omitted variable assumption will lead to an estimated surplus at least as great as the measurement error assumption.

This is borne out by the derivation of expected surplus in the linear case for the omitted variables explanation

$$\text{(8a)} \qquad \begin{aligned} E[CS_1] &= E[(\alpha + \beta p + \gamma y + u)^2/(-2\beta)] \\ &= (\alpha + \beta p + \gamma y)^2/(-2\beta) + \sigma^2/(-2\beta) \end{aligned}$$

and for the errors in measurement explanation

$$\text{(8b)} \qquad E[CS_3] = E[(\alpha + \beta p + \gamma y)^2/(-2\beta)] = (\alpha + \beta p + \gamma y)^2/(-2\beta).$$

The difference in the two expressions, $\sigma^2/(-2\beta)$, increases with the variance of the true error and decreases with price responsiveness.

For any consumer surplus function which is <u>convex</u> in x, the above discussion demonstrates that there will be a difference in calculated consumer surplus depending on the implicit assumption about the source of the error. One commonly used functional form for demand, the semi-log, generates a consumer surplus function which is <u>linear</u> in x. However, the semi-log has problems of its own, because the conditional expectation on x (the dependent variable) is now a convex function of the error. Unlike the linear case, the conditional mean of x for the semi-log function is not the systematic portion of the demand function. That is

$$\begin{aligned} E[x] &= E[\exp(\alpha + \beta p + \gamma y + u)] = \exp(\alpha + \beta p + \gamma y)\exp(\sigma^2/2) \\ &\neq \exp(\alpha + \beta p + \gamma y) = E(\hat{x}) \end{aligned}$$

because the mean of exp(u) is $(\sigma^2/2)$, if u is distributed $N(0,\sigma^2)$. It is solely because of this result that a difference arises in the semi-log's expected values of consumer surplus for the two error source interpretations.

$$\text{(9a)} \qquad E[CS_2] = \frac{E[\exp(\alpha + \beta p + \gamma y + u)]}{-\beta} = \frac{\exp(a + \beta p + \gamma y + \sigma^2/2)}{-\beta}$$

$$\text{(9b)} \qquad \neq E[CS_4] = \frac{E[\exp(\alpha + \beta p + \gamma y)]}{-\beta} = \frac{\exp(\alpha + \beta p + \gamma y)}{-\beta}.$$

Econometricians have suggested adjusting the constant term so that the expected value of predicted x's will be equal to the observed x's; that is, the adjustment would force the distribution of $\hat{x}_i$ to have mean xi. This adjustment would involve defining a new constant

$$\alpha' = \alpha + \sigma^2/2$$

and using α' to calculate $\hat{x}$.

There is a subtle inconsistency in the logic of the above adjustment however. If the researcher believes that the error (u) is due to errors in measurement, then there is no reason to desire $E(\hat{x}) = E(x)$. The errors in measurement explanation suggests no particular credence should be given the observed values of x. In fact, the semi-log specification implicitly assumes the errors in measurement of x are skewed. It may be this property of the semi-log which explains its frequent success at fitting recreational data. Surely errors in recall of x_i will be larger with larger x's.

Interestingly, calculations for consumer surplus under the two error sources would be identical if the constant were adjusted in calculating the consumer surplus from $\hat{x}$. However the calculation of consumer surplus from $\hat{x}$ implies the error source is errors in measurement and it is in just this case that adjustment of the constant term is ill advised. Without the adjustment a difference in consumer surplus of $\sigma^2/2$ will exist for the semi-log demand under competing error assumptions, even when the coefficients are known with certainty.

Consumer Surplus from Estimated Parameters

Seldom is the researcher blessed with knowledge of the true parameters of the demand function. Indeed, one is fortunate if the statistical analysis produces unbiased estimators of these parameters. Even if the estimators are unbiased, any set of parameter estimates will embody the inherent randomness of the sample and the parameter estimates will themselves be random variables.

In the previous section, the conditions under which the expected value of consumer surplus would differ with error source were explored. This analysis presumed known demand parameters. In the following, the analysis is generalized to the case when surplus is calculated from estimates of the true parameters.

Suppose the parameters of a linear demand function have been estimated on the basis of a sample of observations on x, p, and y. These parameter estimates are denoted $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$. Analogous to (3a) and (3b) the consumer surplus estimates for the individual, if the error is presumed to be due to omitted variables, are given by

(10a) $$\hat{CS}_{1_i} = \frac{x_i^2}{-2\hat{\beta}}$$

for the linear case and

(10b) $$\hat{CS}_{2_i} = \frac{x_i}{-\hat{\beta}}$$

for the semi-log. If one believes the errors in measurement or random preference explanation, the individual estimates analogous to (4a) and (4b) are

(11a) $$\hat{CS}_{3_i} = \frac{\hat{x}_i^2}{-2\hat{\beta}} = \frac{(\hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i)^2}{-2\hat{\beta}}$$

and

(11b) $$\hat{CS}_{4_i} = \frac{\hat{x}_i}{-\hat{\beta}} = \frac{\exp(\hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i)}{-\hat{\beta}}$$

respectively.

Comparing the estimates associated with the linear demand function under the two error source assumptions (i.e. (10a) and (11a)), the following difference arises for the individual

(12) $$\hat{CS}_{1_i} - \hat{CS}_{3_i} = \frac{\hat{x}_i^2}{-2\hat{\beta}} - \frac{(\hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i)^2}{-2\hat{\beta}}$$

$$= \frac{(\hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i + \hat{u})^2}{-2\hat{\beta}} - \frac{(\hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i)^2}{-2\hat{\beta}}$$

$$= \frac{\hat{u}_i^2 + 2\ (\hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i)\ \hat{u}_i}{-2\hat{\beta}}.$$

For any specific individual, this expression cannot be signed, but the average for the sample can be. Summing the difference in consumer surplus estimates over the sample and dividing by N yields

(13) $$\frac{\Sigma(\hat{CS}_{1i} - \hat{CS}_{3i})}{N} = \frac{\Sigma\hat{u}_i^2}{-2N\hat{\beta}} + \frac{2\Sigma(\hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i)\hat{u}_i}{-2N\hat{\beta}} = \frac{\Sigma\hat{u}_i^2}{-2N\hat{\beta}}$$

because by definition of the least squares estimators, $\Sigma\hat{x}_i\hat{u}_i = 0$.[2] Thus for any sample of data and linear model, the method for calculating consumer surplus which implicitly assumes omitted variables will produce a larger estimate of average consumer surplus than will the method which implicitly assumes all error is due to errors in measurement. The difference will be equal to

$$\frac{(N-k)s^2}{-2N\beta}$$

where s^2 = variance of the residual and k is the number of parameters in the equation.

Taking these results a bit further, it is useful to examine the properties of (13). Equation (13) is the expression for the difference between the two calculations of consumer surplus for a given sample. Its size will vary, of course, for different samples, since it is itself a random variable. The expression for the expected value of the difference suggests something about the problems in which this difference will likely be large.

Equation (13), which is the expected value of a ratio of random variables, does not have an exact representation. However, an approximation formula for such problems exists.[3] Applying the approximation to this case gives the following:

(14) $$E\left[\frac{\Sigma(\hat{CS}_{i1} - \hat{CS}_{3i})}{N}\right] = E\left[\frac{\Sigma\hat{u}_i^2/N}{-2\hat{\beta}}\right] \approx \frac{E(\Sigma\hat{u}_i^2/N)}{-2E(\hat{\beta})}\left(1 + \frac{\text{var }\hat{\beta}}{(E\hat{\beta})^2}\right).$$

If the model is correctly specified so that the coefficients are unbiased estimates of the true parameters, then (14) can be expressed as

$$(15) \qquad E\left[\frac{\Sigma(\hat{CS}_{1i} - \hat{CS}_{3i})}{N}\right] \approx \frac{\frac{N-k}{N}\sigma^2}{-2\beta}\left(1 + \frac{\mathrm{var}\,\hat{\beta}}{\beta^2}\right).$$

The first term of (15) is simply the ratio of the expected values of the numerator and denominator in (13). Since $\hat{\beta}$ is an unbiased estimator of β , the denominator is -2β. The numerator of this first term is simply the expected value of $\hat{u}^2$. The second term in (15) reflects the fact that the expected value of a ratio of two random variables is not the ratio of the expected values, but must be weighted by the population analog to the sample statistic

$$\left(1 + \frac{1}{(\text{t-ratio})^2}\right).$$

This weight will be greater than one since $1/(\text{t-ratio})^2$ is positive. The important point is that when one takes into account the fact that consumer surplus estimates are derived from estimates of the demand function parameters, a difference still remains between the omitted variables and errors in measurement consumer surplus estimates. The above demonstrates for the case of unbiased coefficients that the difference can be expected to be larger than if the coefficient were known with certainty.[3]

Returning briefly to the semi-log function, a comparison of expressions (10b) and (11b) depend on whether an adjustment in the constant term of the expression is employed. The econometric procedure of adjusting the constant term would now involve defining an estimate of $\sigma^2/2$, i.e. $s^2/2$ where

$$s^2 = \frac{\Sigma \hat{u}_i^2}{N-k}.$$

If this adjusted constant were used in calculating $\hat{x}$, then the expected value of the difference in consumer surplus estimates would disappear, since the adjustment is made such that $E(x_i) = E(x_i)$. However, consistent with the earlier arguments, the adjustment is considered to be inappropriate here. The nonsymmetrical pattern of errors around x values implicit in the semi-log specification may represent reality better and may be one reason why the semi-log often appears to provide a better fit. Thus for the individual

$$(16)\qquad \hat{CS}_{2i} - \hat{CS}_{4i} = \frac{x_i - \hat{x}_i}{-\hat{\beta}} = \frac{\exp(\hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i)}{-\hat{\beta}}(\exp \hat{u}_i - 1).$$

If the constant term is not adjusted, the difference in the individual's consumer surplus is given in (16). To evaluate the expected value of the difference, it is easier to evaluate the expected value of each expression first. The expected value of $\hat{CS}_{2i}$ (omitted variables interpretation) is

$$(17a)\qquad E[\hat{CS}_{2i}] = E[\frac{x_i}{-\hat{\beta}}] = E[\frac{\exp(\alpha + \beta p_i + \gamma y_i + u_i)}{-\hat{\beta}}]$$

$$= \exp(\alpha + \beta p_i + \gamma y_i)\ E[\frac{\exp(u_i)}{-\hat{\beta}}]$$

$$\approx \frac{\exp(\alpha + \beta p_i + \gamma y_i)\exp(\sigma^2/2)}{-\beta}\ (1 + \frac{\text{var}\ \hat{\beta}}{\beta^2}).$$

Calculating the expected consumer surplus under the errors in measurement assumption yields

$$(17b)\qquad E[\hat{CS}_{4i}] = E[\frac{\hat{x}_i}{-\hat{\beta}}] = E\ [\frac{\exp(\hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i)}{-\hat{\beta}}]$$

$$\approx \frac{\exp\ (\alpha + \beta p_i + \gamma y_i)\ \exp\ (k\sigma^2/2N)}{-\beta}\ (1 + \frac{\text{var}\ \hat{\beta}}{\beta^2}),$$

where the derivations can be found in the Appendix to this Chapter.

A comparison of equation (14) and (15) demonstrates the expected difference between the estimates obtained from the same data set with two different error source explanations when a semi-log function is fitted. Once again omitted variables will lead to a larger expected surplus estimate, because $\exp(\sigma^2/2) = \exp(N\sigma^2/2N) > \exp(k\sigma^2/2N)$ for any data set which will support estimation of the k parameters.

Properties of the Consumer Surplus Estimator

In the last sections consumer surplus estimates were shown to differ depending on the procedure used to calculate them which in turn implied assumptions about the source of the disturbance term. In this section it is demonstrated that, irrespective of the source of error, the conventional consumer surplus estimators (those presented above) will be biased.

The process by which surplus estimates are conventionally derived (i.e. the procedure employed in the previous section) is to replace the true parameters in expressions such as (3) and (4) by their regression estimates. Taking as an example the linear, omitted variables case, from (3a) we see that consumer surplus is given by

$$CS_1^* = x^2/(-2\beta)$$

and the conventional estimator (given in (10a)) is

$$\hat{CS}_1 = x^2/(-2\hat{\beta}).$$

If $\hat{\beta}$ is a maximum likelihood estimator of β, then $\hat{CS}_1$ will be the maximum likelihood estimator of CS_1^* (Zellner and Park, 1979). However, this maximum likelihood estimator has some undesirable properties. As can be seen from the derivations in the previous section (or derived from similar expressions in Zellner and Park), the expected value of $\hat{CS}_1$ is <u>not</u> equal to the expected value of CS_1^*.

That is

$$E(\hat{CS}_1) \approx \frac{(\alpha + \beta p + \gamma y)^2 + \sigma^2}{-2\beta} \left(1 + \frac{\operatorname{var} \hat{\beta}}{\beta^2}\right) \tag{18}$$

$$> E(CS_1^*) = \frac{(\alpha + \beta p + \gamma y)^2 + \sigma^2}{-2\beta} .$$

The conventional consumer surplus estimator is biased. It is biased upward by a factor of $(1 + \operatorname{var} \hat{\beta}/\beta^2)$.

Likewise in each case - linear or semi-log, omitted variables or errors in measurement - one finds that the conventional estimator is biased upward. In each case the bias is related to the term var $\hat{\beta}/\beta^2$.[4] This is because in each case the estimator for consumer surplus is a function of the reciprocal of $\hat{\beta}$.

Note that the bias decreases with the price slope and increases with the variance of the estimated price coefficient. The latter suggests that the bias will increase with a) increasing variance of u, b) decreasing dispersion in price across the sample, and c) increasing correlation between price and other explanatory variables in the equation. All of these bode ill for the travel cost method which depends on cross section data, frequently explaining only a small portion of the variation in trips, and is often plagued by multicollinearity problems particularly with respect to the treatment of the value of time.

While the conventional consumer surplus estimators can be shown to be biased, they appear to be consistent estimators. One can see this from the formula for var $\hat{\beta}$ which, in the general case is

$$\text{(19)} \qquad \operatorname{var} \hat{\beta} = \sigma^2 m^{\beta\beta}$$

where $m^{\beta\beta}$ is the element on the diagonal of the $(Z'Z)^{-1}$ matrix associated with the β coefficient (Z is defined as the vector of exogenous variables). For our particular case, this term can be written more intuitively as

$$\text{(20)} \qquad \operatorname{var} \hat{\beta} = \sigma^2 (\Sigma(p_i - p)^2 \; (1 - r_{py}^2))^{-1},$$

where r_{py} is the correlation between price and income. As sample size increases, the only term which changes is the dispersion in price. In the limit as $N \rightarrow \infty$, $\Sigma(p_i - p)^2 \rightarrow \infty$ and var $\hat{\beta} \rightarrow 0$.

There are more general principles upon which both the biasedness and consistency properties rest. Referring once again to Jensen's inequality helps establish the biasedness property for a broader range of cases. If the estimated consumer surplus, designated $g(\hat{\beta})$, can be shown to be a strictly convex function of $\hat{\beta}$ then, by Jensen's inequality, the expected value of the estimate $(E[g(\hat{\beta})])$ should be greater than $g(E[\hat{\beta}])$. This latter term equals the true consumer surplus, $g(\beta)$, if $\hat{\beta}$ is unbiased. Thus while $\hat{\beta}$ is an unbiased estimator of the true β, strict convexity in the estimated surplus implies upwardly biased consumer surplus estimates.

Zellner (1978) has shown that, indeed, when one calculates a function of the reciprocal of a maximum likelihood estimator, then the expected value of the function will be an upwardly biased estimate of the function of the expected value of the parameter. Additionally, the estimator of the function will not possess finite moments and, when using a quadratic loss function, has infinite risk.

However, the consumer surplus estimators are consistent. Mood, Graybill and Boes show that if $\hat{\theta}$ is a ML estimator for θ, then $f(\hat{\theta})$ is an ML estimator for $f(\theta)$, if there is a one to one mapping between θ and $f(\theta)$. Zehna has extended these results such that the property holds for any $f(\cdot)$ which is a function of θ. Maximum likelihood estimators may be biased but they generally can be shown to be consistent, except in unusual circumstances (Chandra). As a consequence, the consumer surplus estimators will be consistent estimators, if they are functions of maximum likelihood estimators of the parameters: α, β, and γ.

Minimum Expected Loss (MELO) Estimators

Consistency is certainly a desirable property for an estimator, but it is a large sample property. That is, it is not of great practical value if the estimates of interest are usually generated in the context of relatively small samples. Given the scarcity of large samples in recreational studies, it is the small sample properties of consumer surplus estimates which are of particular interest.

Zellner (1978) and Zellner and Park (1979) have proposed a procedure for correcting for the bias which arises when we are interested in a function which is the reciprocal of a maximum likelihood parameter. The core of their argument rests on providing an estimator that will minimize a loss function.

As an example of the technique, consider the function for consumer surplus in the linear-omitted variables case $CS_1^* = x^2/(-2\beta)$. Its ML estimator is $\hat{CS}_1 = x^2/(-2\hat{\beta})$. Zellner's loss function for the estimated surplus would be $[(CS_1^* - \hat{CS}_1)/CS_1^*]^2$. Minimizing this function implies a surplus estimator defined as:

$$(21) \qquad (x^2/-2\hat{\beta}) \left(\frac{1}{1 + \text{var}\,\hat{\beta}/\hat{\beta}^2}\right)$$

which is the ML estimator of CS_1^* times a "shrinking factor" (Zellner, p. 185). Interestingly, the shrinking factor is the ML estimator of the inverse of the multiplicative bias factor arising in (18).

Unfortunately, even (21) is of limited value to us because it presumes knowledge of var $\hat{\beta}$ (and hence σ^2) with certainty. For cases when σ^2 is not known, and specifically when $\hat{\beta}$ is a regression coefficient, Zellner gives the following MELO estimator

$$\hat{\hat{CS}}_1 = \frac{x^2}{(-2\hat{\beta})} \left(\frac{1}{1 + (n-k)s^2 m^{\beta\beta}/(n-k-2)\hat{\beta}^2} \right) \tag{22}$$

where s^2 is the variance of the estimate calculated as $\sum u_i^2/(n-k)$, and $m^{\beta\beta}$ is again the appropriate element of the $(Z'Z)^{-1}$ matrix. Note that $(n-k)s^2 m^{\beta\beta}/(n-k-2)$ is simply the usual estimate of the variance of the regression coefficient reported in all regression routines. Consequently, $(n-k)s^2 m^{\beta\beta}/(n-k-2)\hat{\beta}^2$ is actually the square of the reciprocal of the t-ratio for $\hat{\beta}$. The moments and risk associated with (22) exist and are finite; approximations are given by Zellner.

Consumer surplus for the semi-log function assuming omitted variables is also the reciprocal of a parameter. Consequently a similar MELO estimator can be derived:

$$\hat{\hat{CS}}_3 = (x_1/(-\hat{\beta})) \left(\frac{1}{1 + (n-k)s^2 m^{\beta\beta}/(n-k-2)\hat{\beta}^2} \right).$$

A similar procedure can be used to adjust errors in measurement formulas.

Conclusion

A potentially dramatic difference in benefit estimates can arise from alternative yet commonplace assumptions about the source of error in recreational demand analysis. Theoretical derivation shows that for three typical assumptions about the error - that it results from omitted variables, from random preference, or from inaccurate measurement of trips - computed consumer surpluses will differ. The omitted variables assumption, the one commonly used in travel cost analysis, will likely lead to larger values of consumer surplus than either the random preferences or measurement error in the independent variable. The difference can be expected to increase with the variance of the error, the variance of the estimated price coefficient, and price inelasticity of demand.

To give greater insight into how large these differences might be in practice, estimates of consumer surplus from a sample of sportfishermen are derived. The sample yielded relatively high t-statistics on independent variables although it did not predict very accurately ($\bar{R}^2 = .10$), implying a rather large variance of the error. These characteristics are fairly typical of cross-sectional data. The results show a substantially higher value (78%) for the omitted variable error assumption than for the measurement error/random preference explanation.

This is only half the problem, however. Surpluses computed as functions of regression parameters will likely be upwardly biased, even when these parameter estimates are themselves unbiased. When surplus estimates are non-linear in the parameters, their expected value is larger than the surplus when the true parameters are used. The degree of biasedness is positively related to the variance in the price parameter and the inelasticity of demand.

Large samples do, however, provide consistent measures for surplus. Thus, there are pay-offs from having large samples and confidence in parameter estimates. ML estimators of consumer surplus will have poor small sample properties (Zellner, 1978; and Zellner and Park, 1979). However, Zellner offers us MELO (minimum expected loss) estimators with far better properties. Since recreational surveys are costly, these MELO estimators are a valuable alternative to increased sample sizes.

What implications do the results of this chapter have for the researcher active in measuring benefits? There are a lot of forces at work to confound benefit estimates, and it is difficult to treat all of then? at once. This chapter shows that the source of error will make a difference in consumer surplus values.

If the researcher attributes all of the error to omitted variables (i.e. draws his demand curve through the observed (x^o_1, p^o_1)) when at least some of the error is due to measurement error, he may be substantially overestimating consumer surplus. If the researcher employs the alternative practice of calculating surplus behind the estimated regression line, then he will surely be underestimating surplus since omitted variables are always a source of some error.

In the past, the source of error has been considered of little consequence. Yet, it is shown that improved estimates of consumer surplus can result if one can a) reduce the variance of the error in the regression and b) provide information as to the source of the error. Survey designs which reduce measurement error, for example, by limiting recall information, will be helpful on both counts. Another approach is to collect more in the way of potential explanatory variables. The marginal cost of additional information may be low, but its pay-off may be great if it reduces the variance in the error of the regression. Thus, even though precision in travel cost coefficients is not gained, there is a decrease in the potential error arising from wrong assumptions concerning the error term.

A warning is offered against the usual practice of assuming all error is associated with omitted variables. The practice can lead to upward biases in benefits when either random preferences or measurement error are present. At a minimum, the researcher should explicitly acknowledge the likelihood of upwardly biased estimates. A bolder approach would be to offer estimates of benefits under competing assumptions about the source of error.

The second implication of the results is that the care and attention spent by researchers in obtaining statistically valid estimates of behavioral parameters must carry over to the derivation of benefits. Estimates of consumers surplus have, by construction, random components. Knowledge of how the randomness affects estimated benefits may be as important to policy makers as knowledge of the statistical properties of the estimated behavioral parameters. At a minimum, researchers should assess whether their consumer surplus estimates are likely to be badly biased. Since Zellner’s MELO estimators for the linear and semi-log (as well as other) functional forms are straightforward to calculate, MELO estimators of consumer surplus would be simple to provide.

FOOTNOTES TO CHAPTER 5

1 Since everything in this chapter is demonstrate in terms of the ordinary demand curve and ordinary consumer surplus, $\tilde{p}$ is the price which drives Marshallian demand to zero. Of course $\tilde{p}$ in the semi-log case depends on the limiting properties of the function.

2 The following approximation is necessary to derive expected values throughout the chapter:

$$E(x/y) \approx E(x)/E(y) - \operatorname{cov}(x,y)/(E(y))^2 + E(x)\operatorname{var}(y)/(E(y))^3.$$

The expected value of the ratio of two random variables does not have an exact equivalence.

3 Should the coefficients not be unbiased (that is, should the equation be at least slightly misspecified), then expression (14) will still be true but it will not simplify to (15). Given that the misspecification is due in some way to the correlation between included and omitted variables, it is not possible to determine a priori, whether the existence of such correlation will increase or decrease the difference in surplus estimates.

Suppose that Z_j and ε were correlated where Z_j is the j^{th} explanatory variable. The expected values of each of the terms in (14) would no longer be as simple, reflecting the fact that $E(Z_j\varepsilon)$ is no longer equal to zero.

Using matrix notation for efficiency and labelling the explanatory variable matrix, Z, the first term in (14) now becomes

$$E\left(\frac{\Sigma\hat{u}^2_1}{N}\right) = E\frac{(\hat{u}'\hat{u})}{N} = \frac{N-k}{N}\sigma^2 - E\left(\frac{u'Z(Z'Z)^{-1}Z'u}{N}\right) = \frac{(N-k)\sigma^2}{N} - \frac{E(u'Z)\ (Z'Z)^{-1}E(Z'u)}{N}$$

where the second term above no longer disappears but reflects whatever correlation exists between included and omitted variables.

The expected values of the estimated coefficient $\hat{\theta}$, now become

$$E(\hat{\theta}) = \theta + E((Z'Z)^{-1}Z'u) = \theta + (Z'Z)^{-1}E(Z'u)$$

where $E(\theta_j)$ will exceed θ_j if the correlation between Z_j and u is positive and vice versa. (Of course if there is also correlation with other explanatory variables everything becomes more complicated.)

Finally,

$$\text{var } \hat{\theta} = E(\hat{\theta} - E(\hat{\theta}))^2 = \sigma^2(Z'Z)^{-1} - (Z'Z)^{-1}E(Z'u)E(u'Z)(Z'Z)^{-1}.$$

The second term is positive, so correlation between Z and u will reduce the variance of θ.

As a consequence of the above three derivations, the presence of correlation can not be determined a priori either to increase or decrease the difference in the consumer surplus measures.

APPENDIX 5.1

DERIVATION OF DIFFERENCE IN ESTIMATED CONSUMER SURPLUS USING THE SEMI-LOG DEMAND FUNCTION

The following is the derivation for the expected value of the difference in consumer surplus estimates for the semi-log demand function. When omitted variables causes the error, the expected value of the individual's consumer surplus estimate is

$$(A1) \qquad E\left(\frac{x_i}{-\hat{\beta}}\right) = E\left(\frac{\exp(Z_i\theta + u_i)}{-\hat{\beta}}\right)$$

where Z_i is the i^{th} row of the matrix of explanatory variables $= [1 \; p_i \; y_i]$ and θ is the vector of coefficients $[\alpha \; \beta \; \gamma]'$.

Then, using the approximation formula for the ratio of two random variables yields

$$(A2) \qquad E\left[\frac{\exp(Z_i\theta + u_i)}{-\hat{\beta}}\right] = \frac{\exp(Z_i\theta)E[\exp(u_i)]}{-\hat{\beta}}\left[1 + \frac{\text{var}\ \hat{\beta}}{\beta^2}\right] + \frac{\text{cov}(\exp(Z_i\theta + u_i), -\hat{\beta})}{\beta^2}$$

Given that u_i is distributed as a normal with mean 0 and variance σ^2, then $\exp(u_i)$ is distributed as a lognormal with expected value $\exp(\sigma^2/2)$. Noting that the covariance term equals zero, expression (A2) can be rewritten as

$$(A3) \qquad E\left[\frac{x_i}{-\hat{\beta}}\right] = \frac{\exp(Z_i\theta)\exp(\sigma^2/2)}{-\hat{\beta}}\left(1 + \frac{\text{var}\ \hat{\beta}}{\beta^2}\right).$$

The expected value of the individual's consumer surplus estimate when errors in measurement is the principal cause of the disturbance term is

$$\text{(A4)} \qquad E\left[\frac{\hat{x}_i}{-\hat{\beta}}\right] = E\left[\frac{\exp(Z_i\hat{\theta})}{-\hat{\beta}}\right] = E\left[\frac{\exp(Z_i\theta + Z_i(Z'Z)^{-1}Z'u)}{-\hat{\beta}}\right].$$

Applying the approximation formula and noting the covariance term is zero gives

$$\text{(A5)} \qquad E\left[\frac{\hat{x}_i}{-\hat{\beta}}\right] = \frac{\exp(Z_i\theta)E[\exp(Z_i(Z'Z)^{-1}Z'u)]}{-\beta}\left(1 + \frac{\text{var}\ \hat{\beta}}{\beta^2}\right).$$

oting that $\exp[Z_i(Z'Z)^{-1}Z'u]$ is simply $\exp[Au]$ where A is a vector of on-random terms, we draw on the result that the expected vaue of exp(w) hen w is normally distributed is equal to exp((variance w)/2). The varince of $Z_i(Z'Z)^{-1}Z'u$ can be expressed as

$$\text{A6)} \qquad \text{var}(Z_i(Z'Z)^{-1}Z'u) = E[u'Z(Z'Z)^{-1}Z_i'Z_i(Z'Z)^{-1}Z'u].$$

he vector Z_i is simply the i^{th} individual's vector of explanatory ariables. So that the formula reflects the average values of the explanaory variables, the matrix $Z_i'Z_i$ can be rewritten as$(1/N)Z'Z$. Making this ubstitution gives us an idempotent matrix and allows the following simpliications:

$$\text{A7)} \qquad E[u'Z(Z'Z)^{-1}\ \frac{1}{N}\ Z'Z(Z'Z)^{-1}Z'u] = \frac{1}{N} E\ [u'Z(Z'Z)^{-1}Z'u].$$

loting that expression (A7) is a scalar and equal to its own trace,

$$\text{(A8)} \qquad \frac{1}{N} E[u'Z(Z'Z)^{-1}Z'u] = \frac{1}{N}\ \text{tr}(Z(Z'Z)^{-1}Z')\ E(uu')$$
$$= \frac{k}{N}\sigma^2$$

because the trace of an idempotent matrix equals its rank, which in this case is k (or 3 in our example).

Now since,

$$\text{var}\ (Z_i(Z'Z)^{-1}Z'u) = \frac{k\sigma^2}{N}$$

then

$$E\left[\frac{\hat{x}_i}{-\hat{\beta}}\right] = \frac{\exp(Z_i\theta)\exp\left(\frac{k\sigma^2}{2N}\right)}{-\beta}\left(1 + \frac{\text{var}\ \hat{\beta}}{\beta^2}\right). \tag{A9}$$