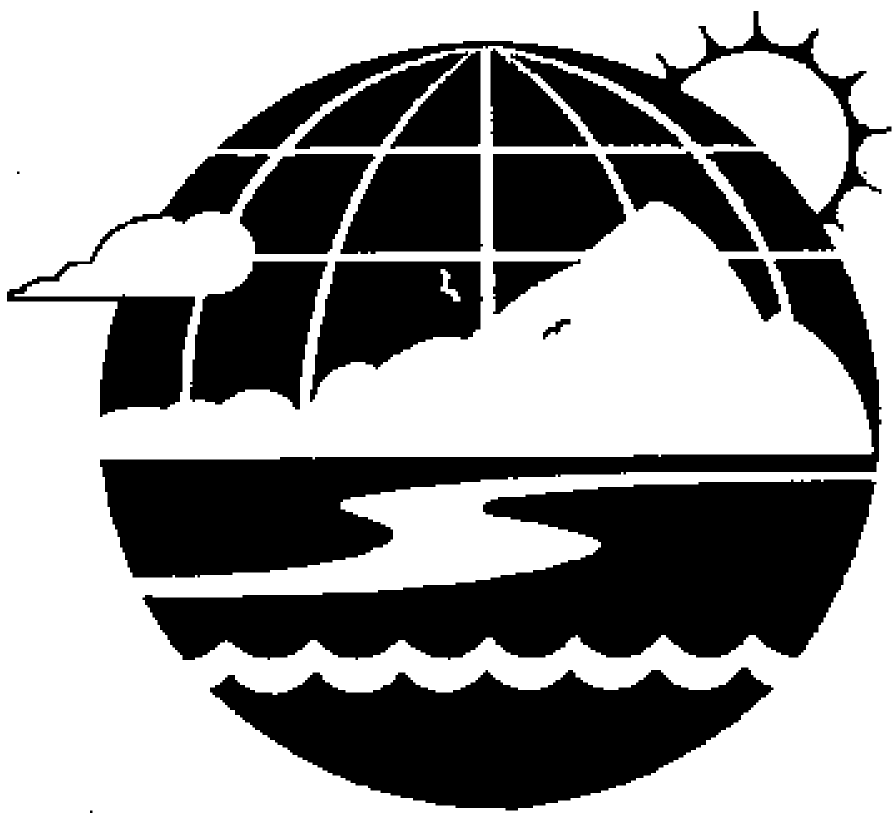
United States Environmental Protection Agency
Policy, Planning and Evaluation (2127)
EPA-230-10-89-069
October 1989

EPA

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 encompassed 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 behavior which can 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)."

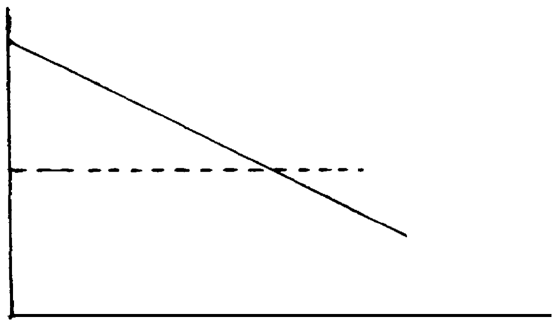
In fact the development of methods of estimating the demand for recreation so closely paralleled the use of zonal models that the so-called travel cost method is often considered synonymous with the use of zones.

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.

Price
(travel cost &
entrance fee)



Recreation trips/time period

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

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

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

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

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

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_1 (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 empirical 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 <u>exact</u> 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.

The Intergrability Problem and Demand Function Estimation

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^o. 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 $x_1 \doteq \gamma \exp(\gamma p)u - \beta/\gamma$ 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 $(x_1^o, x_2^o, \ldots, x_n^o)$ there exists no subvector $(x_2', x_3', \ldots, x_n')$ such that
> $u(x_1^o, x_2^o, \ldots, x_n^o) = u(0, x_2', x_3', \ldots, x_n')$.

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 over-adjustment, 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

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

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:

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

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

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

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:

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

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

(A6) $m = [(1-\gamma)(u^0 + \frac{e^{\alpha}p^{1+\beta}}{1+\beta})]^{\frac{1}{1-\gamma}} \Rightarrow u = \frac{-e^{\alpha}p^{1+\beta}}{1+\beta} + \frac{m^{1-\gamma}}{1-\gamma}$ (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),YO)
*     IF (I.EQ.2) XC(I)=X2(P(1,1),P(2,1),YO)
*     ETC.
 2000 CONTINUE
******************* ALGORITHM ***********
      ITIMES=0
      Y=YO
      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 wIll be $Var(\sum_i v_{ij}/N_j) = \sigma^2/\sqrt{N_j}$ 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:

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

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

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

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

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

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

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

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

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

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

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

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

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,

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

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

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

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

$$x_{ij} = h(z_{ij},\beta) + \varepsilon_{ij} \qquad \text{if } h_{ij} + \varepsilon_{ij} > 0$$

(11)

$$x_{ij} = 0 \qquad \text{if } h_{ij} + \varepsilon_{ij} \leq 0.$$

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, (1b) 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.

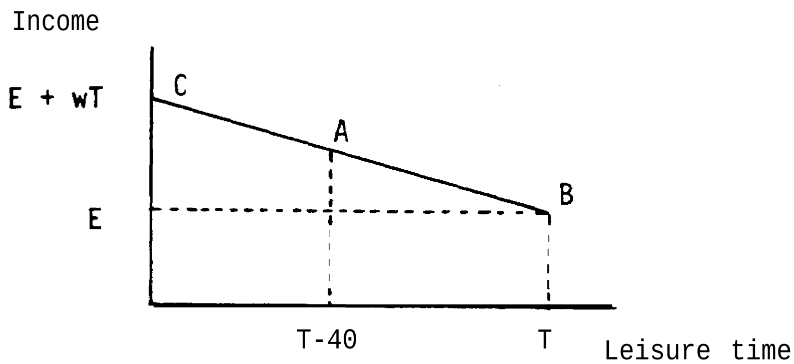


Figure 4.1: The First Generation Budget Constraint

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.

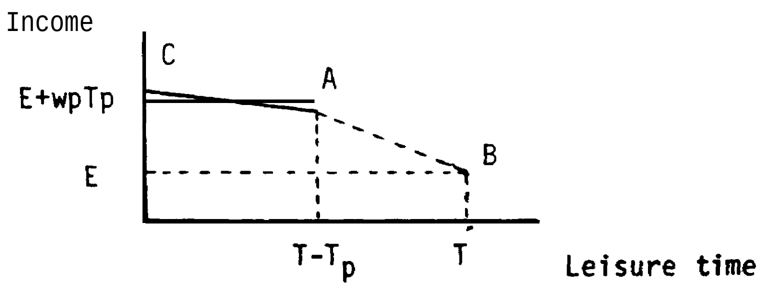


Figure 4.2: Second Generation Budget Constraints

Two other aspects of the second generation labor supply models are noteworthy. The first generation studies estimated functions which were specified in a relatively ad hoc 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

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

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

(4a)
$$\partial u/\partial x_i - \lambda p_i - \mu t_i = 0$$
$$Y - \Sigma p_i x_i = 0,$$
$$\bar{T} - \Sigma t_i x_i = 0.$$

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

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

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

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

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

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

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

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

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

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

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

$$x_j = h_j(\cdot) + \varepsilon_j \quad \text{if and only if } h_j(\cdot) + \varepsilon_j > 0$$
$$x_j = 0 \quad \text{otherwise}$$

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 *a priori* 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 *least in this example*, 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 OLS 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), θ(b), ϕ(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 = 0 or Y - px = 0) 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 0 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):

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

and

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

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

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

for a linear demand curve and

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

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):

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

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.

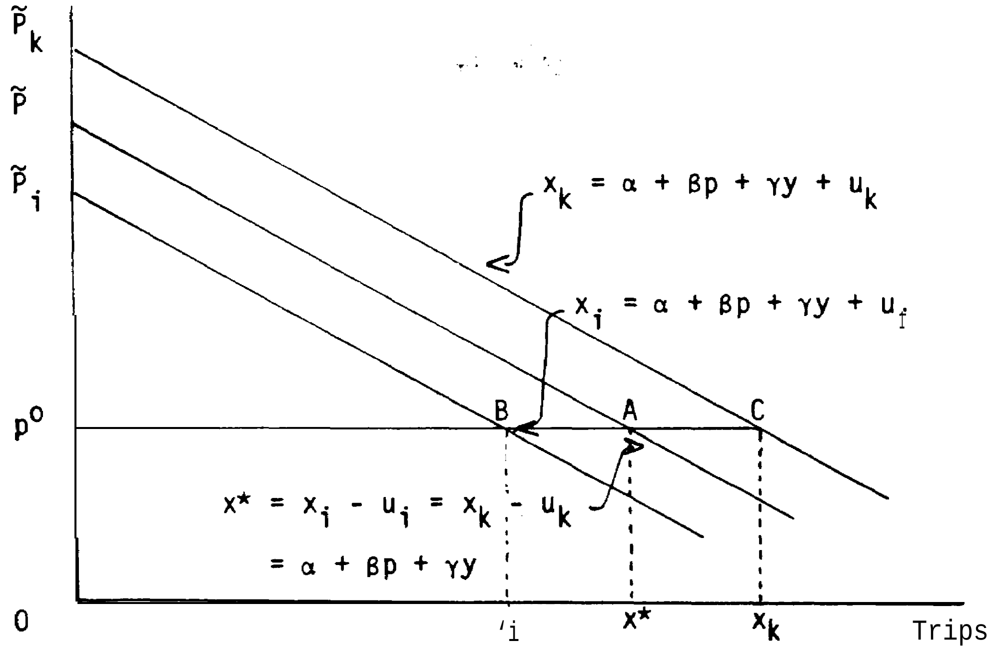


Figure 5.1

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

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

and for the errors in measurement explanation

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

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 convex 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 linear 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.

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

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

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

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

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

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

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

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

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

(17b) $$E[\hat{CS}_{4i}] = E[\frac{\hat{x}_i}{-\hat{\beta}}] = E \; [\frac{\exp(\hat{\alpha} + \hat{\beta} p_i + \hat{\gamma} y_i)}{-\hat{\beta}}]$$
$$\approx \frac{\exp\;(\alpha + \beta p_i + \gamma y_i)\; \exp\;(k\sigma^2/2N)}{-\beta} \; (1 + \frac{\text{var } \hat{\beta}}{\beta^2}),$$

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

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

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

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

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

$$\text{var } \hat{\beta} = \sigma^2 m^{\beta\beta} \tag{19}$$

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

$$\text{var } \hat{\beta} = \sigma^2(\Sigma(p_i-p)^2\ (1-r^2_{py}))^{-1}, \tag{20}$$

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 $N \to \infty$, $\Sigma(p_i-p)^2 \to \infty$ and var $\hat{\beta} \to 0$.

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 $(n-k)s^2m^{\beta\beta}/(n-k-2)$ is simply the usual estimate of the variance of the regression coefficient reported in all regression routines. Consequently, $(n-k)s^2m^{\beta\beta}/(n-k-2)\hat{\beta}^2$ 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:

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

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

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

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

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

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

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

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

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

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

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 vari- nce of $Z_i(Z'Z)^{-1}Z'u$ can be expressed as

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

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 explana- ory 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 simpli- ications:

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

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

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

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

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

PART II

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

CHAPTER 6

RECREATIONAL DEMAND MODELS AND THE BENEFITS FROM IMPROVEMENTS IN WATER QUALITY

In the past ten years, an increasing Federal interest in evaluation of benefits from water quality improvements has evolved. Emphasis in Executive Order 12291 on the comparison between benefits and costs of Federal actions has stimulated much of it. The initial research into benefit evaluation revealed both theoretical and practical problems with applying conventional methods to environmental valuation. Because a large portion of the benefits of water quality improvement are associated with recreational uses, much of the recent research has attempted to measure these benefits in the context of recreation demand models.

In their 1982 paper, Vaughan, Russell and Gianessi suggest five linkages that must be captured quantitatively in order to estimate recreational benefits from a water quality improvement program:

a. the effect of the program on levels of pollutant discharges;

b. the natural system's mechanism for transportation, dilution and transformations of pollutants which produce changes in ambient environmental conditions;

c. the translation of the ambient changes into terms readily perceived and acted upon by recreators;

d. the response pattern of recreators to changes in perceived ambient environmental quality, both by intensity and type of participation, and;

e. the valuation of recreationalists' responses.

The work presented in the following section of this report focuses on techniques for valuing site specific recreational experiences and is empirically tractable for regional water quality management (linkages d. and e. above). There is no question, however, that the total value of broad national policies cannot feasibly be estimated using site specific models. However, emphasis on d. and e. is important on two counts. First, site specific models obviously provide a means for valuing more regionally specific water quality changes - a topic of considerable importance when an estuary is specified (e.g. the Chesapeake Bay) as an ecologically vulnerable and socially valuable water body worthy of particular attention. The second use of site specific models is to provide the basis for reliable and defensible value estimates needed in national studies.

Valuing Quality Changes in Demand Models

Appealing to Maler's conditions (1974) of weak complementarily, Freeman (1979a) presented the now well-known theoretical justification for measuring benefits of quality changes from demand functions. The weak complementarily conditions discussed by Maler and others set out the requirements for valuing a change in quality as the change in the area behind the (compensated) demand for a market good conditioned on quality. The difference in these areas (as designated in Figure 6-1) for the market good was shown to be a complete measure only if the change in quality had no effect on the individual when the market good was not consumed. Many authors (e.g. Bouwes and Schneider, 1979; Norton, Smith and Strand, 1983) have used this justification in obtaining benefit estimates.

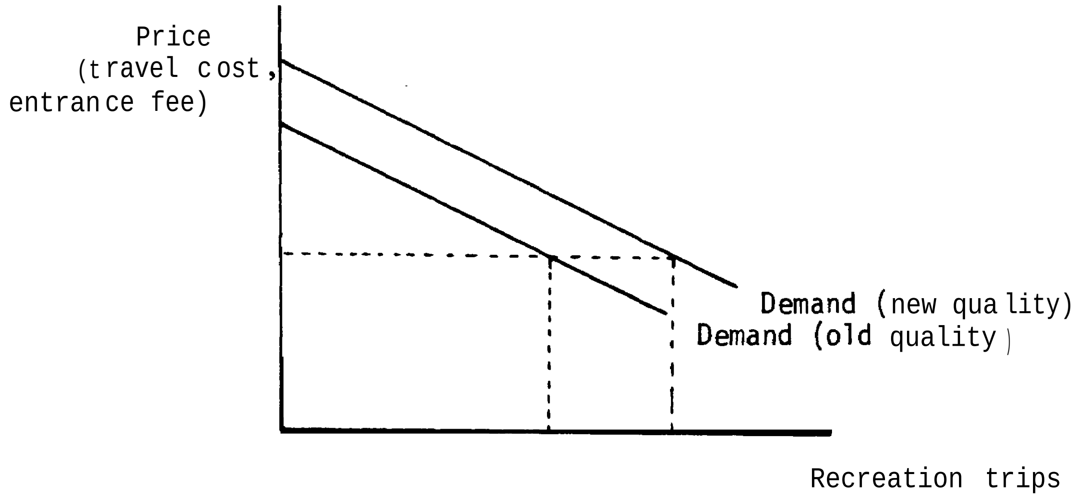


Figure 6.1: Benefits From Water Quality Changes

An extension of these arguments was offered in the context of the household production function by Bockstael and McConnell, (1983). They showed that the change in areas behind a derived demand function for an input into the household production process could be used as a measure of the benefits of a quality change if two conditions were met. The quality must be weakly complementary to a household produced commodity, and the input (whose demand function was being used for benefit measurement) must be essential to the production of that commodity. The simple travel cost model offers a special case of their model. Consider travel as an input (with a price or constant marginal cost) into the production of recreational experiences at site A. One wishes to value a change in the quality of that site. Then the area between the travel cost demand function conditioned on the two levels of quality will be a reasonable measure, if travel is an essential input into the recreational experience and if no utility is gained from an improvement in quality at site A when the individual does not recreate at site A.

At least two studies have attempted another approach to measuring the value of quality improvements in the context of a single site Hotelling-Clawson-Knetsch model, without explicitly incorporating quality. Both the paper by Davidson, Adams and Seneca (1966) and that of Sutherland (1982) postulate hypothetical water quality changes which would open a previously closed site. Thus, the entire recreational value of the site is attributed to water quality improvements in their studies.

Extending the Single Site Model to Value Quality Changes

The most important extension of the traditional recreational demand model has been to incorporate quality, thus facilitating the valuation of water quality changes as described above. Early applications of the travel cost model limited the set of explanatory demand variables to travel costs and income, with site quality omitted. Clawson recognized the importance of the site's quality in his 1959 study but did not incorporate it in his analysis because the single site, cross-sectional model precluded observation of any variation in the level of quality.

Persistent researchers nonetheless found ways to take account of quality aspects of the recreational trip in recreation demand models. One of the first was Stevens (1966) who used two methods to introduce quality into sportfishing demand at Yaquino Bay, Oregon. First, he examined aggregate data on salmon trips as influenced by average success rate of anglers. Short and long run elasticities were derived for success rate changes. He also used a mail survey and hypothetical questions about success rate change to estimate individual response.

The Stevens paper is one example of a class of approaches which incorporates quality into the model by some extension of the single site model. One such extension involves deducing the relationship between recreation and water quality by comparing use levels across a number of sites which were assumed to differ in quality only. The response to quality change was then imputed to an individual site of interest. Pooling data over a number of sites provided a means of capturing quality changes but remaining within the construct of a single equation model. The advantage of such an approach is that it is comparatively easy to estimate. The disadvantage lies in the implicit assumptions in the model which are extremely strong.

An example is the study of salmon and steel head trout fishing by Brown, Singh, and Castle (1965). They treated the sportfishing activity across Oregon as though it were all occuring at a single site and examined how the variation in catch rate (a quality change resulting from different fish runs) influenced the zonal trips per capita. In general, they found that there was a positive influence of catch rate on trips per capita. Another variation can be found in Reiling, Gibbs and Stoevener (1973) who also pooled data across sites, ignoring the possible substitution among sites and variation in other factors at sites.

Burt and Brewer (1971) and Cicchetti, Fisher, and Smith (1976) pioneered a generalization of the Hotelling-Clawson-Knetsch model to a multi-equation system. However they could not include quality variables explicitly in their models. In these studies, systems of demand equations were estimated with the prices of substitute sites entering all equations. Quality differences were not explicitly incorporated but were assumed intrinsic to the site. Differences in site demands over and above the effect of price were attributed to the inherent differences in the quality of the sites. While this approach represents a valuable extension to the more naive models which ignore substitutes, it does not provide a means for valuing changes in quality since quality is not introduced explicitly.

Desvousges, Smith and McGivney’s varying parameters model offered a means of explicitly incorporating the effect of quality in demand systems. After estimating a series of demand curves for single sites, they regressed the intercept and coefficients of the travel cost on quality aspects of the sites. Vaughan and Russell followed a similar route in their method of valuing fishing days.

The most recent approach to the problem of multiple, quality-differentiated demand is the use of the discrete choice or share models, in conjunction with the travel cost model. Whereas the early efforts treated the recreational decision as containing a single site with varying quality or as having many sites with one implicit quality difference, the recent literature explicitly models the choice among multiple, quality-differentiated sites. Multiple site models which incorporate quality directly were pioneered by Binkley and Hanemann (1978) and by Hanemann (1978). They specified a more elaborate model which included the estimation of a) total number of visits made to all sites, b) number of sites visited and c) the allocation of visits among chosen sites. This modelling effort required more sophisticated estimation tools than ordinary least squares and was the first to use discrete choice models for explaining recreational behavior. Of the studies which have since used discrete choice models, the work by Morey, (1981, 1984), Feenberg and Mills (1980), Caulkins (1982), Bouwes and Bishop (1982) and Rowe, Morey, and Shaw (1985) are good examples.

The appeal of these multiple site models is three-fold. As argued above, examination of multiple sites is critical to obtaining water quality variation sufficient to induce behavior change. Additionally, multiple sites may be of consequence because water quality changes resulting from EPA regulations may be wide-ranging (Vaughan and Russell, 1982). Ignoring them obliterates substitution possibilities and, some have argued, leads to an overstatement of the benefits associated with improving the quality of a site (Caulkins, Bishop, and Bouwes, 1982). Finally descrete choice and share models provide more or less satisfactory ways of handling the corner solution problem characteristic of multiple-alternative decisions. In most observed situations, individuals do not choose to take trips to all possible (quality-differentiated) sites. Thus they are at corner solutions with regard to the demand for trips to several sites. In Chapters 3 and 4, proper treatment of corner solutions was shown to matter critically in welfare estimation. The problem takes on added dimensions and importance in the context of multiple site models.

Plan of Part II Volume II

The purpose of the remaining portion of this volume is to explore the ways in which economists have modified the travel-cost type recreation demand models investigated in Part I to accommodate the assessment of quality changes. In practice, individuals' responses to changes in water quality can not be deduced unless choices among different levels of quality can be observed. The development of models of choices among multiple alternatives is an important contribution because it provides one means of capturing responses to changes in quality. Even when quality is not an issue, multiple site models provide a more realistic specification of the recreationalist’s choice, because the valuation of one site will depend on the existence of alternative, substitute sites. A model which ignores viable alternatives will be misspecified. Chapter 7 investigates the properties of systems of demands for resources of differing qualities and how quality variables can be introduced into such systems. Subsequent Chapters develop a utility theoretic model of the multiple alternative decision (Chapter 8) and provide a detailed examination of the models present in the literature which purport to value environmental quality changes (Chapter 9). Improvements in these models are developed and an example is supplied in Chapter 10. This example demonstrates the application of an empirically feasible multiple site model to actual data and shows how improvements in water quality can be valued in this context.

CHAPTER 7

EVALUATING ENVIRONMENTAL QUALITY IN THE CONTEXT OF RECREATION DEMAND MODELS: AN INTRODUCTION TO MULTIPLE SITE MODELS

Up to this point the report has addressed issues related to the estimation of demand equations for a single recreation activity. These single activity models are the foundation of the traditional "travel cost" model and are quite appropriate for valuing individual resources with no close substitutes. More and more, economists have been interested in modelling the demand for systems of recreation activities, sites or resources. Resources of interest are not always unique, individual sites. Sometimes we are interested in modelling the demand for activities which can be carried out at a host of competing sites - fishing sites along a river or across a system of rivers, beaches on an estuary or along a region's coastline, parks within a regional system, etc.

The modelling of demand for systems of activities or sites takes on added relevance when an environmental quality characteristic is in question. If we wish to value improvements in fish catch along those rivers, water quality at the beaches or visibility of scenic vistas in the parks, we need a mechanism for incorporating quality into our recreational demand models. We also need a means of observing quality variation within our data set, for without observations on behavior in the face of varying quality, there is no hope of estimating the value of improvements in quality. Given the fact that site specific recreation demand data is not systematically collected from year to year, it is unlikely that time series data capturing variation in environmental quality will be available. The cross-section data typically used for single site models will exhibit variation in variables such as costs and income over the sample but not in the quality characteristics of the site, unless data on individuals' perceptions of quality are used instead of objective measures. While perceptions may be the relevant variable for stimulating behavior, we would hope that perceptions are closely aligned with fact. Introducing varying perceptions of a constant objective quality characteristic to reflect individuals' responses to changing objective quality characteristics begs the important questions and introduces an additional vector of random error. When employing cross-section data, the only reliable means of incorporating quality is to model the demand for an array of sites of differing qualities. Because of this feature and the observation that recreation behavior is often defined over an array of discrete alternatives - be they differing activities or different sites at which to enjoy the same activity, environmental economists have turned to multiple site models of recreation demand.

In what follows we will explore the nature of recreation demand and what features of it make modelling so difficult. We will proceed by discussing theoretical issues related to the two aspects which make multiple site modelling different from what we have discussed to date. Multiple site models involve the treatment of demand for more than one good and require that quality characteristics be incorporated into demand estimation.

Chapter 8 presents a consistent theoretical treatment of the problem which serves as a foundation for a theoretically desirable, although practically difficult, estimation model. The theoretical treatment also provides a basis for discussing the alternative multiple site modelling approaches which can be found in the literature. This literature is critically reviewed in Chapter 9.

The Nature of Recreation Demand

At this point, it is important to develop more fully what we mean by the nature of recreational decisions and what makes recreational demand so difficult to model. The traditional zonal travel cost motivation for recreational demand is built on the frequency of visits to a resource site from different zones of origin. Behind this visitation pattern are individuals' decisions and, as we have argued throughout this report, it is the individual's behavior which is critical to welfare evaluation. As Chapter 3 makes clear, it is extremely important to model the individual's decision to participate or not as well as his frequency of participation.

For many recreation problems the simple travel cost model is inappropriate for another reason. There is not just one site and thus one travel cost which enters the decision to participate in the activity. Instead there are often many alternative sites which offer equivalent or similar experiences. If the sites are identical and travel costs approximately the same, it is costless to aggregate the problem into a single demand for the activity. If not, aggregation may be inappropriate and even infeasible.

Sites are rarely identical nor equi-distant from an individual's residence, but this fact can actually help researchers trying to value environmental quality characteristics. As noted earlier, the valuation of environmental improvements through demand modelling requires the observation of variations in environmental quality. If sites encompass different environmental quality levels and are subject to different costs of access, then observations on use may reveal individuals trade-offs between environmental quality and money. This requires a defensible model of individual choice in the context of a finite number of discrete, quality differentiated goods.

Introducing Quality Into The Demand Function

Welfare economics has historically been applied to the evaluation of price and income changes. In contrast, environmental and natural resource economists are most interested in using the results of welfare economics to evaluate exogenous changes in the quality characteristics of a commodity or to assess the value associated with the existence of a resource. The latter is the relevant concept when a new facility is proposed or when access to an existing resource is considered for elimination. The former includes attributes of resource related activities and commodities, such as fish catch, visibility, congestion and water quality.

In this section we address the evaluation of quality changes in the context of recreational demand models. While definitions of quality related welfare changes analogous to Hicksian measures of price-evoked welfare effects are easily drawn, problems arise in moving from the abstract to the operational level. Once again we face the dilemma of choosing a functional form for estimation among a number of seemingly arbitrary choices, each with its own implications and behavior.

Before examining these implications, let us briefly consider our ultimate end. We wish to evaluate the benefits to an individual of a change in the level of an exogenous quality characteristic, denoted b. (Often b will be a vector, but here we treat it as a scalar.) Ignoring for the moment any additional conditions we may wish to place on b, we presume that utility is a function of b so that $u(x,b,z)$ indicates that utility is a function of the quantity of the recreation good consumed (x), its quality (b), and the amount of the Hicksian good (z). Define $v(p,b,y)$ as the indirect utility function associated with $u(x,b,z)$. Then the measures of the value of a change in quality from b^0 to b^1, analogous to the Hicksian compensating and equivalent variation measures, are C_q and E_q in the following:

(1) $$v(p,b^1,y-C_q) = v(p,b^0,y)$$

and

(2) $$v(p,b^1,y) = v(p,b^0,y+E_q).$$

Written in terms of the expenditure function

(3) $$C_q = m(p,b^0,u^0) - m(p,b^1,u^0)$$

and

(4) $$E_q = m(p,b^0,u^1) - m(p,b^1,u^1)$$

where u^0 is the value of utility evaluated at $v(p,b^0,y)$ and u^1 is utility evaluated at $v(p,b^1,y)$.

These definitions are quite intuitive. However, it should be noted that as they stand, in general form, they imply nothing particular about either the ordinary or the compensated demand functions. In fact it is precisely these relationships which will be of interest to us in this section. As we have already noted, it is the demand function which we typically estimate since it is based on observable behavior. Yet it is the preference function about which we often have a priori hypotheses. Where do we start in incorporating quality - at the demand function or utility function level? What do properties of one imply about the form of the other?

Once again we are faced with the dilemma of choosing between specifying the form of demand or the form of preferences. We could begin by building quality in the ordinary demand function directly and immediately discerning the behavioral implications for consumer choices. Alternatively, we could incorporate the characteristic into the utility function to discern the implications for consumer preferences. Unfortunately, if we start at the demand function level with what would appear to be a desirable property, the implications for preferences are not always appealing.

An example should help demonstrate this. Consider the linear demand function, $h(p,y) = \alpha + \beta p + \gamma y$, and add quality (denoted b) in an arbitrary manner

(5) $$x = h(p,b,y) = \alpha + \beta p + \gamma y + \delta b \qquad \alpha,\delta > 0;\ \beta < 0.$$

The implication of this form is that an increase in quality shifts the graph of this demand curve in price-quantity space outward, in a parallel manner. Demand increases by the same quantity regardless of price or the individual's income. A second implication is that even though quality is introduced into the demand function in a way such that $\partial h/\partial b > 0$, this does not necessarily imply that the consumer's welfare is increased by an improvement in quality (i.e. $\partial u(x,b,z)/\partial b > 0$). If $\gamma < 0$ (i.e. the commodity is an inferior good) while $\delta > 0$, it turns out that $\partial u/\partial b < 0$. The consumer's welfare is <u>reduced</u> by an increase in quality.

To some extent, these results flow from the special structure of preferences associated with the linear demand function (5). Thus, for example, if one employs a semilog demand function of the form $\ln x = \alpha + \beta p + \gamma y$ or $\ln x = \alpha + \beta p + \gamma \ln y$, where $\beta < 0$, and makes the constant term, α, a function of b in such a way that $\partial h/\partial b > 0$, it can be shown that $\partial u/\partial b > 0$. However, if one uses the log-log demand function $x = \alpha \bar{p}^{\beta} y^{\gamma}$ or a demand function of the form $x = \alpha \bar{p}^{\beta} e^{\gamma y}$ and makes α a function of quality in such a way that $\partial h/\partial b > 0$, it turns out that sign $(\partial u/\partial b)$ = - sign $(1+\beta)$; thus if $\beta > -1$ (i.e., the good is essential), $\partial u/\partial b < 0$. (See Hanemann, 1982a and 1982b).

The point here is that mere inspection of the demand function h(p,b,y) does not always provide a reliable indication of how quality affects the consumer's preferences. A procedure which avoids this uncertainty starts with a utility function and incorporates quality directly to obtain some formula u(x,b,z) or v(p,b,y), from which the ordinary demand function h(p,b,y) can be obtained in the standard manner.

A useful way in which to incorporate quality might be called the "transformation method." In this method, one replaces the argument x in a two good utility function $\bar{u}(x,z)$ with some function f(x,b) to obtain $u(x,b,z) = \bar{u}[f(x,b),z]$. The advantage of this method is that, for various transformations f(x,b), the resulting demand function h(p,b,y) can be directly related to the demand function associated with $\bar{u}(x,z)$. Also, the implications for preferences will be clearly understood. Several examples of transformations are discussed in the Appendix to this chapter, including the "scaling" or "repackaging" transformation of Fisher and Shell (1971), Willig's "cross-product repackaging" transformation, and a common transformation known as "translation".

One final issue should be addressed before concluding this section - the role of the structural property of consumer preferences known as "weak complementarity". This concept, as introduced by Maler (1974), imposes the following condition on our utility function:

(6) $$\text{if } x = 0 \text{ then } \partial u/\partial b = 0.$$

Maler employed this property to derive a relation between compensating and equivalent variation measures and areas under compensated demand functions, (g(p,b,u)). If u(x,b,z) does not satisfy this property, the area

(7) $$\int_{p^0}^{\tilde{p}} [g(p,b^1,u^0) - g(p,b^0,u^0)dp]$$

understates the compensating variation, where $u^0 = v(p,b^0,y)$ and $\tilde{p}$ is the cutoff price (possibly infinite) at which max $[g(p,b^1,u^0),g(p,b^0,u^0)] = 0$. If u(x,b,z) does satisfy the property, the area in (7) measures compensating variation exactly.

Regardless of whether areas under compensated demand functions are exact measures, the property (6) has an important implication (which may or may not be appealing) as an axiom of consumer behavior. It implies that the consumer does not care about a change in a good's quality when he is not consuming the good. Both the scaling and the cross-product repackaging transformations discussed in the appendix possess this property, while the two "translative" transformations presented there do not.

Maler's analysis was based on the implicit assumption of a "smooth" relation between quality and welfare in the utility function u(x,b,z). If this smoothness assumption is dropped, the link between weak complementarily and the equivalence of welfare measures with areas under compensated demand functions changes. Suppose that, instead of

(8) $$u(x,b,z) = \bar{u}[x + \psi(b),z],$$

one writes

(9) $$u(x,b,z) = \bar{u}[x + \xi(x)\cdot\psi(h), z]$$

where $\xi(x)$ is a type of switching function:

$$\xi(x) = 1 \text{ if } x > 0 \text{ and } \xi(x) = 0 \text{ if } x = 0.$$

Then (9), unlike (8), satisfies weak complementarily. The only difference between (8) and (9) occurs at the boundary of the non-negative orthant, where $x = 0$; in the interior, where $x > 0$ the two indifference maps coincide. Thus over the space for which $h(p,b,y) > 0$ the behavioral implications of the two functions are identical.

In short, by means of the simple device used in (9), any demand function which on its face appears to violate weak complementarity can, in fact, be reconciled with this property. Of course, the reasonableness of the weak complementarily assumption is an empirical question and could presumably be tested using data which included cases where x was not consumed at all. However if one only has data for cases where a positive quantity of x is consumed, as is often the case, then it is impossible in practice to determine whether weak complementarily holds: one cannot discriminate between (8) and (9) as the true preference structure. For this type of data set, weak complementarily is a costless assumption.

The Specification of Demand Models for Systems of Alternatives

Arguments for using multiple site demand models have been stated in Freeman (1979a) and empirically addressed by Caulkins, Bishop and Bouwes (1982). The treatment of quality improvements in the single site context does not take into consideration all of the substitute or complementary effects among sites. For example, if sites were actually substitutes and a number of single site models were used to assess benefits from a regional water quality improvement, benefit estimates would likely be biased because substitution possibilities among sites would not be completely considered. To avoid upwardly or downwardly biasing estimates, more comprehensive systems of demand must be developed. When one moves to the more complete system, however, there are unique theoretical issues which arise in addition to the practical problem of how to make the model empirically tractable.

A principal theoretical issue is the question of how one goes about generating specifications for entire demand systems. Once again, the same two alternatives exist. One approach consists of specifying a direct or indirect utility function explicitly and then deriving the ordinary demand functions either by maximizing the direct utility function or by applying Roy’s Identity to the indirect utility function. The second approach consists of specifying the ordinary demand functions directly. Here, however, there is an important distinction between what is possible when dealing with a demand equation for a single good and demand systems for multiple goods.

Suppose that we wish to estimate the demand for a single good, x, and we are willing to treat all other goods as a Hicksian composite commodity, z, so that the direct utility function is u(x,z). Recognizing the homogeneity of demand functions in prices and income, we write the demand function of interest as:

(lo) $$x = h\left(\frac{p_x}{p_z}, \frac{y}{p_z}\right),$$

where it is understood that the implied demand function for z is

(11) $$z = g(p_x,p_z,y) = \left[y - h\left(\frac{p_x}{p_z}, \frac{y}{p_z}\right)\right]/p_z.$$

We could generate the formula for $\bar{h}(\cdot)$ by specifying some quasiconcave, increasing direct utility function or some quasiconvex indirect utility function increasing in y and decreasing in (p_x, p_z). Or we could simply write down an arbitrary formula for the function h(*). But, if we do the latter, we must ensure that our function satisfies the integrability conditions:

(12) $$\frac{\partial h}{\partial p_x} + x\frac{\partial h}{\partial y} \le 0$$

(13) $$\frac{\partial h}{\partial p_z} + z\frac{\partial h}{\partial y} = \frac{\partial g}{\partial p_x} + x\frac{\partial g}{\partial y}.$$

Now in the two good case, (12) is a substantive restriction but is not too difficult to satisfy. By contrast, (13) is a trivial restriction which is always satisfied. As long as it is understood that g(·) is given by (11), <u>any</u> bivariate function <u>automatically</u> satisfies (13) (for a demonstration, see Katzner, 1970, p. 68).

We are not so fortunate when dealing with demand functions for two or more goods in addition to the numeraire. In the general case with $u(x_1,\ldots,x_N)$, and ordinary demand functions $x_i = h^i(p_1,\ldots,p_N,y)$, i = 1,...N, satisfying $\sum p_i h^i(p,y) = y$, the integrability conditions are

(14) $$\frac{\partial h^i}{\partial p_i} + x_i\frac{\partial h^i}{\partial y} \le 0$$

(15) $$\frac{\partial h^i}{\partial p_j} + x_j\frac{\partial h^i}{\partial y} = \frac{\partial h^j}{\partial p_i} + x_i\frac{\partial h^j}{\partial y} \quad \text{all } i, j.$$

The symmetry conditions (15) are now a non-trivial restriction, since they are obviously <u>not</u> satisfied by an arbitrary set of homogeneous functions $h^1(\cdot),\ldots,h^N(\cdot)$. For this reason it is likely to be considerably more convenient to specify multiple site demand systems by starting from some explicit direct or indirect utility function. (See LaFrance and Hanemann, 1985, for the preference structure implications of imposing symmetry on ad hoc demand systems.)

The discussion above raises another issue which is often of practical importance when modelling demands for multiple goods, namely what to do about the demand functions for the goods that are <u>not</u> of interest. In the recreation context, for example, one typically has <u>no</u> data on the households' consumption and/or prices of other, nonrecreation goods. Writing on the subject of demand functions for multiple recreation sites Cicchetti, Fisher and Smith (1976, fn. 12) address the very question and conclude that, if one had data on only a subset of consumption activities, it is not appropriate to employ a system of demand equations that is consistent with the hypothesis of utility maximization.

The Cicchetti, Fisher and Smith conclusion does not always seem to be an acceptable one.[1] It is true that if one wishes to employ the fitted demand system merely for the empirical prediction of demand responses to changes in prices, a system of demand functions that violates homogeneity or the integrability conditions may be satisfactory. But if, as is often the case, one intends to derive welfare evaluations from the fitted demand equations (e.g. estimates of the value of a particular site or the benefits from some quality enhancement program), it is difficult to justify the use of demand functions which violate the postulates that are the foundation of welfare analysis. Moreover, the conclusion reached by Cicchetti et al. seems unduly pessimistic since several strategies exist for handling, albeit imperfectly, data on a subset of commodities in a manner more consistent with utility theory.

Let us begin by considering the general problem: The utility function can be written as $u(x_1,\ldots,x_N,z_1,\ldots,z_M)$ where the x_i's are the goods of concern, and the z_i's are all the other goods consumed by the individual. The respective price vectors are $p_1,\ldots,p_N$ and $q_1,\ldots,q_M$. The individual maximizes $u(x,z)$ subject to the budget constraint $\sum p_i x_i + \sum q_i z_i = y$, which generates two sets of ordinary demand functions

$$\text{(15a)} \qquad x_i = h_i^x \left(\frac{p_1}{q_1},\ldots,\frac{p_N}{q_1},\frac{q_2}{q_1},\ldots,\frac{q_M}{q_1},\frac{y}{q_1}\right) \quad i = 1,\ldots,N$$

(15b) $$z_j = h_j^z\left(\frac{p_1}{q_1},\ldots,\frac{p_N}{q_1},\frac{q_2}{q_1},\ldots,\frac{q_M}{q_1},\frac{y}{q_1}\right) \quad j = 2,\ldots, M$$

(15c) $$z_1 = [y - \sum_{i=1}^{N} p_i\, h_i^x(\cdot) - \sum_{j=2}^{M} q_j\, h_j^z(\cdot)]/q_1$$

and an indirect utility function $v(p,q,y)$, where z_1 has been taken as the numeraire. The demand functions in (15a) and (15b) taken together are known as a "complete" demand system. However, we only care about (15a) which is known as an "incomplete" demand system. What can be done with this?

Suppose for a moment that we had estimates of all the coefficients in the incomplete demand system (15a)[4] Suppose, too, that these demand functions satisfy the local integrability conditions for incomplete demand systems which, involve the symmetry of the Slutsky terms with respect to the p_i's,

$$\frac{\partial h_i^x}{\partial p_k} + x_k \frac{\partial h_i^x}{\partial y} = \frac{\partial h_k^x}{\partial p_i} + x_i \frac{\partial h_k^x}{\partial y} \quad \text{all } i,\, k = 1,\ldots,N$$

and the negative definiteness (not semi-definiteness) of the Slutsky matrix,

$$\frac{\partial h_i^x}{\partial p_i} + x_i \frac{\partial h_i^x}{\partial y} < 0 \quad i = 1,\ldots, N.$$

Given these assumptions, LaFrance and Hanemann (1985) show that the incomplete demand system (15a) contains sufficient information to allow us to calculate correctly money welfare measures for changes in the p_i's. Specifically, they show that the system (15a) can be integrated, treating the q_j's as fixed parameters, to obtain what is called a "quasi-indirect utility function", with the property that, if one uses it to calculate measures of compensating or equivalent variation for some price change from p' to p'' one obtains exactly the same results as if one knew the true indirect utility function $v(p,q,y)$ obtained by integrating the complete system (15a) and (15b). Thus, having access only to the incomplete demand system for the x's, does not cause problems for welfare evaluations with respect to changes in the prices of the x's.

The problems arising when one works with incomplete demand systems have to do with estimation rather than welfare evaluation. If one knows the q_j's but not the z_j's, there is in principle no problem in estimating the incomplete demand system (15a). But if one has no data on either the q_j's or the z_j's then there _is_ a problem in estimating (15a). An exception is when one can assume that the q_j's do not vary across the sample so that they can be taken as being subsumed in the coefficients of the demand functions. Only in this latter case can welfare measures safely be constructed. Otherwise, the estimation of (15a) omitting $q_2,\ldots,q_M$ is a form of specification error which may produce biased estimates of the coefficients pertaining to the p_i's, the nature of the bias depending on the functional form of the demand equations and the degree of correlation between the included p_i's and the omitted q_j's.

An alternative approach is the strategy of assuming a separable utility function. This avoids many of the estimation problems, but it raises some conceptual problems for welfare evaluation. In this case one assumes that $u(x_1,\ldots,x_N,z_1,\ldots,z_M) = f[\bar{u}(x_1,\ldots,x_N),z_1,\ldots,z_M]$, where f is an increasing function of M + 1 arguments and quasiconcave in z, and $\bar{u}$ is increasing and quasiconcave in x. Thus the marginal rate of substitution between any pair of x_i's is independent of each z. Ignoring the selection of a numeraire, the ordinary demand functions for the x_i's and z_j's resulting from the maximization of u(x,z) subject to $\Sigma\, p_i x_i + \Sigma\, q_j z_j = y$ are of the form

(16a) $$x_i = h_i^x(p,q,y) = \bar{h}_i^x[p,H^x(p,q,y)] \quad i = 1,\ldots,N$$

(16b) $$z_j = h_j^z(p,q,y) \quad j = 1,\ldots,M.$$

where $H^x(p,q,y) = y_x$ which equals the income spent on the purchase of the x's.

Additionally the functions

(17) $$x_i = \bar{h}_i^x(p,y_x) \quad i = 1,\ldots,N$$

may be obtained by solving the maximization problem

(18) $$\text{maximize } \bar{u}(x_1,\ldots,x_N) \text{ s.t. } \Sigma p_i x_i = y_x.$$

The function

$$(19) \qquad y_x = H^x(p,q,y),$$

as well as the functions $h_j^z(\cdot)$, may be obtained by solving the maximization problem

$$\underset{y_x,z}{\text{maximize}}\ f[\bar{v}(p,y_x),z] \quad \text{s.t.} \quad y_x + \sum q_j z_j = y$$

where $\bar{v}(p,y_x)$ is the indirect utility function arising from the maximization problem (18).

Equations (17) were termed "partial demand functions" by Pollak (1971). We will extend that term to refer to equations both (17) and (19). To distinguish between (14) and (15), we denote equations (17) as "micro-allocation" functions and equations (19) as "macro-allocation" functions. The reason for our terminology is that the functions in (17) indicate how a fixed total recreation budget, y_x, should optimally be allocated in terms of visits to individual recreation sites as a function of their costs, while the function in (19) indicates how the overall recreation budget should optimally be determined. Since they are derived from conventional (sub-) utility functions, the functions in (15) possess the standard properties of ordinary demand functions - homogeneity, summability, and symmetry and negative semi-definiteness of the Slutsky terms. Note that while some of them may exhibit zero income effects, it is <u>not</u> possible for <u>all</u> of the demand functions in (17) to have zero income effects; nor is it possible for all of them to have negative income effects. This is one of the ways in which they differ from incomplete demand systems, since it is entirely possible for all of the demand functions in (15a) to have zero or negative derivatives with respect to total income.

Although partial demand functions for the x_i's can be estimated under circumstances in which it is impossible to estimate incomplete demand functions, this advantage is not obtained without some cost. First, the partial demand functions can only be used to predict site variation patterns conditional on a given total recreation budget. If prices change, say from p' to p'', y_x also will change from $y_x' = H^x(p',q,y)$ to $y_x'' = H^x(p'',q,y)$, and this change cannot be ascertained from the partial demand functions for the x's. In practice one might try to circumvent this problem by estimating some sort of <u>ad hoc</u> macro-allocation function relating y_x to p and y and perhaps some general price index as a crude approximation to (19); this approximation would be used to predict the change from y_x' to y_x''.

The second problem, the construction of welfare measures, is more stubborn. Knowledge of the partial demand functions provides insufficient information about the individual's preferences to permit exact calculations of the welfare effects of the price change from p' to p''. By integrating (17) one can obtain the partial indirect utility function $\bar{v}(p,y_x)$, but not the full indirect utility function $v(p,q,y)$. The true welfare measures for the price change should be based on the latter. Suppose that, in addition to knowing $\bar{v}(p,y_x)$, one knows y_x' and y_x'', or can estimate them from some crude, non-utility-theoretic macro-allocation function. The best that one can hope to do is calculate the quantities $\bar{C}$ and $\bar{E}$, where

$$\bar{v}(p'', y_x' - \bar{C}) = \bar{v}(p', y_x')$$

$$\bar{v}(p'', y_x'') = \bar{v}(p', y_x'' + \bar{E}).$$

Hanemann (1983b) showed that these are in general different from the true welfare measures, but they at least provide bounds on them:

$$\bar{C} \leq C \text{ and } E \leq \bar{E}.$$

The empirical adequacy of these bounds, however, remains an open question.

Introducing Quality Into Multiple Site Demand Models

At this point, the next logical step is to combine what we have learned from the discussion of quality and the discussion of modelling systems of demands. There is an abundance of empirical evidence that individuals who participate in water based recreation visit more than one site and those who do not, generally have more than one effective alternative from which to choose. Moreover, in most regions there is some variation in the quality of recreation experience afforded by different sites, and casual evidence suggests that recreationists care about at least some dimensions of site quality and trade off price and quality in making their recreation decisions. There is, therefore, a strong case for introducing site quality into multiple site models of recreation behavior.

Following the previous discussions, one way of doing this is to introduce some measures of site quality - ordinal or cardinal measures - explicitly into the utility function and the demand functions. Let b_i represent the level of quality associated with a visit to site i, $i = 1,\ldots,N$ where b_i is likely to be a vector of characteristics but we will treat it here as a scalar. The utility function is assumed to be $u(x_1,\ldots,x_N,b_1,\ldots,b_N,z)$ and the resulting ordinary demand functions are[3]

$$x_i = h^i(p_1,\ldots,p_N,b_1,\ldots,b_N,q,y), \quad i=1,\ldots,N.$$

One implication should immediately be noted: the demand for any site depends in principle not only on its own quality characteristics but also on those of all other sites. This may cause some problems where one employs subjective rather than objective measures of site quality and when individuals do not visit all available sites, because it is often difficult in practice to elicit subjective ratings of site quality for sites that people do not visit (Hanemann, 1984b; Hanemann, 1978, Ch 6; Caulkins, 1982).

As discussed earlier, the utility function $u(x,b,z)$ may be generated either by adding the quality variables, b, to some known utility function $u(x,z)$ in an <u>ad hoc</u> manner, or by employing the transformation method. For example, with scaling, the utility function is

$$u(x,b,z) = \bar{u}[\psi_1(b_1)x_1,\cdots,\psi_N(b_N)x_N,z] \tag{20a}$$

where $\psi_i(b_i)$ may be interpreted as an overall index of site i's quality, and the resulting demand functions are

$$x_i = h^i(p,b,q,y) = \frac{1}{\psi_i(b_i)}\bar{h}^i[\frac{p_1}{\psi_1(b_1)}, \ldots,\frac{p_N}{\psi_N(b_N)},q,y] \quad i = 1,\ldots,N. \tag{20b}$$

The unfortunate implication of scaling mentioned earlier carries over to (20a, b). That is if the demand for a site is price inelastic, an increase in quality reduces its demand. With the cross product repackaging transformation, the utility function is

$$u(x,b,z) = \bar{u}\,[x_1,\ldots,x_N,\ z + \textstyle\sum\psi_i(b_i)x_i] \tag{21a}$$

and the demand functions are

$$x_i = h^i(p,b,q,y) = \bar{h}^i[p_1-q\ \psi_1(b_1),\ldots,p_N-q\ \psi_N(b_N),q,y], \quad i = 1,\ldots,N. \tag{21b}$$

However it is generated, a utility function incorporating quality characteristics of the general form $u(x,b,z)$ may appear from one point of view, to be too much of a good thing. This is because, even after quality differences have been accounted for, it treats each of the goods as different commodities. Even if all the sites have exactly the same characteristics, the general formulation $u(x,b,z)$ implies that they would have different demand functions, which may be implausible. This can be remedied, for example, by specializing (20a) or (21a) to

$$u(x,b,z) = \bar{u}[\sum \psi_i(b_i)x_i, z] \tag{22}$$

$$u(x,b,z) = \bar{u}\,[\sum x_i, z + \sum \psi_i(b_i)x_i]. \tag{23}$$

In these formulations, if all sites have exactly the same characteristics, they will have exactly the same demands. If they have different characteristics, they will have different demands. However, (22) and (23) imply that, allowing for quality differences, the sites are all perfect substitutes and an individual would generally visit only one site, the selection of this site involving a trade off between price and quality.[4] A less extreme approach would be to assume that recreation sites can be grouped into several classes, each class representing a different type of recreation experience (freshwater versus saltwater sites, isolated versus heavily urban sites, etc.) and, therefore, having a different demand function.

The above discussion raises an issue which will hold a prominent place in subsequent discussions. The models in (22) and (23) illustrate how "corner solutions" in which an individual has zero consumption of some goods (i.e. some of the inequality constraints that $x_i \geq 0$, $i = 1,...,N$ are binding) can arise from purely theoretical considerations. Corner solutions are, however, more than a theoretical phenomenon. In practice, whenever one works with data on individuals' consumption behavior and a fairly disaggregate commodity classification, he is likely to observe instances of corner solutions. In the recreation context, although individuals may visit several sites over the course of the recreation season, it is unusual to find that they visit <u>all</u> possible sites.

The ramifications of corner solutions, both statistical and utility-theoretic, have only recently begun to receive attention. From a statistical point of view, perhaps the most important implication is that there is a probability mass at $x_i = 0$ which needs to be incorporated into the estimation procedure, as in Tobit models. From the point of view of economic model formulation, an implication is that the ordinary demand functions must satisfy an additional restriction besides homogeneity, summability, and the symmetry and negative-semidefiniteness of the Slutsky matrix, namely that they assume only non-negative values; thus a function like

$$x_i = h^i(p,y) = \alpha_i + \sum_{j=1}^{N} \beta_{ij} p_j + \gamma_i y \quad i=1,\ldots,N$$

cannot in fact be a valid formula for an ordinary demand system without some further modification because its range extends to the negative orthant.

There is a more subtle problem in dealing with corner solutions in a manner consistent with the hypothesis of utility maximization. Suppose that, at the current prices and income, an individual is consuming some positive amounts of goods 3 through N but nothing of goods 1 and 2. Then, small changes in the prices p_1 or p_2 will have no effect on his demands for goods 3,...,N. Within some region of (p,y) space his demand functions $h^i(\cdot)$, $i = 3,\ldots,N$, will be independent of p_1 and p_2 and will depend only on $p_3,\ldots,p_N$ and y. Thus, as one moves from corner to corner the arguments of the demand functions change. Only at an interior solution (i.e., the individual visits every site) do the demand functions depend on the full set of prices, $p_1,\ldots,p_N$.

A further consequence of corner solutions concerns the manner in which a stochastic element is combined with an economic model in order to generate a statistical model suitable for estimation. By far the most common practice in demand analysis is to postulate some utility model devoid of stochastic elements, u(x,z), derive the corresponding demand system, also devoid of stochastic elements, $h^i(p,q,y)$, $i = 1,\ldots,N$, and then, in the last minute as it were, add some error terms $\varepsilon_1,\ldots,\varepsilon_N$ to justify the application of statistical procedures to the stochastic equations, $x_i = h^i(p,q,y) + \varepsilon_i$. An alternative procedure would be to introduce the random elements into the utility function at the very beginning, $u(x,z,\varepsilon)$, and then derive the demand system in the conventional manner allowing the random elements to carry over to the ordinary demand functions, $x_i=h^i(p,q,y,\varepsilon)$. Although this has occasionally been considered in the context of demand models corresponding to interior solutions, notably by Pollak and Wales (1969), it turns out to be of crucial importance in demand models for corner solutions, as will be shown in subsequent chapters.

FOOTNOTES TO CHAPTER 7

1. We are assuming a demand system which applies to the behavior of individual consumers. The question of modelling aggregate demand functions for recreation sites is considerably more complex and will not be addressed here.

2. This should not be construed as an assault on the use of Willig's (1976) or Vartia's (1982) approximations. These are procedures for approximating the indirect utility function $v(p,y)$ underlying the demand functions $h^1(\cdot),\ldots,h^N(\cdot)$, and are clearly needed when $v(p,y)$ does not have a closed-form expression. But they produce nonsensical results if the demand functions violate the integrability conditions, homogeneity, or summability.

3. This formulation is called a "Generalized Lancaster" model in Hanemann (1982a) which explains its relation to both Lancaster's (1966, 1971) quality model and the model employed by Houthakker (1951-52) and Theil (1951-52). The Houthakker-Theil approach can be shown to underlie the recreation demand model of Brown and Mendelsohn (1984).

4. In the case of (22), the individual selects that site for which $p_i/\psi_i(\beta_i)$ is lowest; in the case of (23) he selects that site for which $p_i-\psi_i(\beta_i)$ is the lowest.

APPENDIX 7.1

SOME TRANSFORMATION MODELS FOR INCLUDING QUALITY IN DEMAND FUNCTIONS

The simplest example of the transformation method is the transformation $f(x,b) = \psi(b)\cdot x$ where $\psi(\cdot)$ is some increasing function. This scaling or repackaging transformation was first introduced by Fisher and Shell (1971) and has been widely employed in the literature on quality and demand analysis. However, the utility function $u(x,b,z) = \bar{u}[\psi(b)\cdot x,z]$ has somewhat unusual implications which have not always been recognized by those who employ it. These stem from the fact that quality is a direct substitute for quantity in this formulation: a doubling of the quality index, ψ, has exactly the same impact on the consumer's welfare as a doubling of quantity, x. The consequences of this assumption may be observed in the ordinary demand function for x and the indirect utility function, which take the forms

(A1) $$h(p,b,y) = \frac{1}{\psi(b)}\ \bar{h}\left(\frac{p}{\psi(b)},\ y\right)$$

(A2) $$v(p,b,y) = \bar{v}\left(\frac{p}{\psi(b)},\ y\right).$$

Defining ε_b as the elasticity of demand for x with respect to its quality $(\partial h/\partial b)(b/x)$, ε_p as the price elasticity of demand $(-\partial h/\partial p)(p/x)$, and η as the elasticity of the ψ function with respect to b, $(\psi'(b)\cdot b/\psi)$, it follows from (A1) that

(A3) $$\varepsilon_b = \eta(1-\varepsilon_p).$$

Thus the consumer's response to a change in quality is linked to his response to a price change, and in a somewhat peculiar manner. Even though we explicitly assume that utility is increasing in quality (i.e. $\eta > 0$), inelastic demand will cause $h(\cdot)$ to be a decreasing function of b. That is if $\varepsilon_p < 1$, then $\partial h/\partial b < 0$. If demand is inelastic, an increase in quality reduces consumer's demand. This follows intuitively from the property that quality and quantity are substitutes.

An alternative transformation with somewhat more plausible behavioral implications is Willig's (1978) "cross-product repackaging" transformation $u(x,b,z) = \bar{u}[x,z + \psi(b)\cdot x]$. In this case

(A4) $$h(p,b,y) = \bar{h}[h - \psi(b),y]$$

(A5) $$v(p,b,y) = \bar{v}[p - \psi(b),y].$$

Hence,

(A6) $$\partial h/\partial b = -\psi'(b)\partial h/\partial p.$$

Thus, as long as x is not a Giffen good (i.e. as long as $\partial h/\partial p < 0$), an improvement in quality raises the consumer's demand for the good, but the response is proportional to the size of the price elasticity of demand. That is if demand is fairly inelastic (e.g. x is a necessity), a change in quality has a small effect. While if demand is very elastic, it is very responsive also to quality changes.

Another common transformation is $f(x,b) = x + \psi(b)$, which is known as "translation". In this case

$$u(x,b,z) = \bar{u}(x + \psi(b),z) \text{ and}$$

(A7) $$h(p,b,y) = \bar{h}[p,y + p\psi(b)] - \psi(b)$$

(A8) $$v(p,b,y) = \bar{v}[p,y + p\psi(b)]\cdot$$

It follows from (A7) that the effect of a change in quality on the demand for x is tied to the effect of a change in income:

(A9) $$\partial h/\partial b = \psi'(b)\ [\sigma\varepsilon_y - 1]$$

where ε_y is the income elasticity of demand for x, $(\partial h/\partial y)\ (y/x)$, and σ is the budget share of x, (px/y). Thus while it is still true that $\partial u/\partial b > 0$, the demand function is increasing or decreasing in quality according to

(A10) $$\partial h/\partial b \gtrless 0 \quad \text{as } \varepsilon_y \gtrless 1/\sigma\cdot$$

In this context, it might be more convenient to apply the translation transformation to the numeraire good and write $u(x,b,z) = \bar{u}(x,z + \psi(b))$. The simple linear demand function $h = \alpha + \beta p + \gamma y + \delta b$ is an example of this, where the translation is $\psi(b) = \delta b/\gamma$.

When the translation transformation is applied to the numeraire, the following demand function for x and indirect utility function are generated:

$$h(p,b,y) = \bar{h}\,[p,y + \psi(b)] \tag{A11}$$

$$v(p,b,y) = \bar{v}\,[p,y + \psi(b)]. \tag{A12}$$

Hence, $\partial u/\partial b > 0$ and

$$\partial h/\partial b = \psi'(b)\,[\partial h/\partial y] \gtrless 0 \quad \text{as } \partial h/\partial y \gtrless 0. \tag{A13}$$

If x is a normal good, an increase in quality raises its demand, the increase being proportional to the income responsiveness of demand. Note that the last two transformations have the property that C = E. However, (A12) has the additional implication that the compensated demand function for x is independent of quality. As German (1976) pointed out, the scaling and translation transformations can be combined to generate more complex transformations in which the sign of $\partial h/\partial b$ depends on the signs or magnitudes of both price and income elasticities.

CHAPTER 8

THE PROPERTIES OF THE MULTIPLE SITE RECREATION DECISION

The previous chapter highlighted the two characteristics of recreation demand which have been recognized as the most critical and most difficult aspects to model. The first aspect is the discrete/continuous nature of the recreational choice. The discrete components involve the choice of whether or not to participate and at which of a finite number of discrete sites to recreate. The continuous choice involves frequency of use - both in total and at each chosen site. The second characteristic of the recreation demand problem is quality. The finite set of discrete sites are often quality differentiated. Additionally, it is often the value of a change in the quality of a site or set of sites which is of interest to the researcher.

It is an empirical fact that individuals who participate in water based recreation often have a choice among sites - and often choose to visit more than one site in a season. Even those who visit only one site rarely visit their cheapest site, but instead trade-off price for site quality. Unfortunately, conventional neoclassical behavioral models do not take account of quality or of discrete/continuous decisions. The standard calculus is ineffectual in the face of the corner solutions which arise in discrete choice problems. These corner solutions are often of a special sort. Not only are zeros encountered in the data set when individuals do not participate in the activity but, for any individual, there-are zero visits made to a number of the alternative sites. While individuals rarely visit only one site, they are almost never observed visiting all available sites. In this chapter we present a consistent, utility theoretic model of multiple site recreation demand incorporating site quality and allowing for the discrete/continuous nature of the decision problem. We draw on material from Chapter 7 as well as the literature on discrete choice models and quality differentiated goods. It should be noted that this type of decision can be found in many economic problems. Progress made here will be useful, not only for recreational demand modelling, but also for the study of transportation demand, local public goods, the demand for quality differentiated (branded or graded) consumer goods, etc.

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 them 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:

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

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.

in order to illuminate some of the problems which arise when one attempts to model corner phenomena in a manner fully consistent with utility theory, it is convenient to begin by describing how one models a special type of corner solution which we shall call an "extreme" corner solution. An extreme corner solution problem is one in which the individual chooses to consume only one of a set of discrete alternatives. All other alternatives have zero levels of consumption. The utility maximization problem that concerns us in this section is:

$$(1) \quad \underset{x,z}{\text{maximize}}\ u(x,b,z;\varepsilon) \quad \text{s.t.} \quad p'x + qz = y, \quad x_i \geq 0,\ z \geq 0.$$

For simplicity we treat z as a scalar and set its price, q, equal to unity. We are now principally concerned with the non-negativity constraints in (1) and the circumstances in which they are binding. Extreme corner solutions arise when something in the structure of (1) forces a corner solution in which all but one of the xi's is zero - i.e. the consumer buys only one of the quality-differentiated goods. This can occur either because the utility function $u(\cdot)$ has a special structure which treats the x_i's as perfect substitutes or because there is a set of additional constraints in (1) of the form

$$(1a) \quad x_i x_j = 0 \quad \text{all } i \neq j.$$

That is, for some logical or institutional reason, the xi's are mutually exclusive in consumption.

By contrast, a "general" corner solution arises when some, but not necessarily N-1, of the xi's are zero at the optimum. For most recreation choices one finds evidence of a general rather than an extreme corner solution. However, the analysis of extreme corner solutions is more straightforward and will set the stage for more general models.

Suppose, for the moment, that the consumer has decided to consume only good i (visit site i). Invoking the assumption of weak complementarily, his utility, conditional on this decision, is

$$u_i = u(0,\ldots,0,x_i,0,\ldots,0,b,z;\varepsilon) \equiv u_i^*(x_i,b_i,z;\varepsilon).$$

Given his selection of this site, he still must make a decision as to the number of times he should visit it over the recreation season. This decision is made by maximizing the conditional utility function (conditioned on the choice of i and only i) subject to a budget constraint:

(2)
$$\begin{aligned} &\underset{x,\, z}{\text{maximize}} \quad u_i^*(x_i, b_i z; \varepsilon) \\ &\text{s.t. } p_i x_i + z = y \\ &\qquad x_i \geq 0,\ z \geq 0. \end{aligned}$$

The solution may involve setting $x_i = 0$, i.e. he will not participate in any recreation over the season. It may involve setting $x_i = y/p_i$, i.e. he will spend all of his income on recreation. Either of these corners may be handled by the methods to be described below.

For the moment we ignore these corner solutions and simply write the interior solutions to (2) as

$$x_i = h_i^*(p_i, b_i, y; \varepsilon)$$

and

$$z = z(p_i, b_i, y; \varepsilon) \equiv y - p_i h_i(p_i^*, b_i, y; \varepsilon)$$

where $h_i^*(\cdot)$ is the conditional demand for x_i, conditional on x_i and only x_i being non-zero. The conditional indirect utility function obtained by substituting these functions back into $u_i^*(\cdot)$ is $v_i^*(p_i, b_i, y; \varepsilon)$. Assuming that $u_i^*(\cdot)$ is a well-behaved direct utility function, these three functions possess all the standard properties, including satisfying Roy's identity. The explanation for the error term here is the random utility hypothesis. The quantities x_i, z and v_i^* are known numbers to the consumer but, because his preferences are incompletely observed, they are random variables from the point of view of the econometric investigator.

All of the foregoing is conditional on the consumer's selecting site i. The discrete choice of which site to select (remember we are modelling the extreme corner solution case which presumes that only one site will be selected) can be represented by a set of binary valued indices $d_1, \ldots, d_N$ where $d_i = 1$ if $x_i > 0$ and $d_i = 0$ if $x_i = 0$. The choice may be expressed in terms of the conditional indirect utility functions as

$$(3) \qquad d_i(p,b,y;\varepsilon) = \begin{cases} 1 & \text{if } v_i^*(p_i,b_i,y;\varepsilon) \geq v_j^*(p_j,b_j,y;\varepsilon) \text{ all } j \\ 0 & \text{otherwise.} \end{cases}$$

To the observer, the discrete choice indices are random variables with a mean $E\{d_i\} \equiv \pi_i$ given by

$$(4) \qquad \pi_i(p,b,y;\varepsilon) = Pr\{v_i^*(p_i,b_i,y;\varepsilon) \geq v_j^*(p_j,b_j,y;\varepsilon), \text{ for all } j\}.$$

Now consider the original, unconditional utility maximization problem, which consists of (1) augmented, if necessary, by the constraints in (1a). The complete set of the "unconditional" ordinary demand functions associated with this problem will be denoted $h^i(p,b,y;\varepsilon)$, for all $i = 1,\ldots,N$ and $z(p,b,y;\varepsilon) \equiv y - \Sigma\, p_i h^i(p,b,y;\varepsilon)$, and the resulting unconditional indirect utility function is $v(p,b,y;\varepsilon)$. The unconditional functions can be defined by their relationship with the corresponding conditional ones:

$$(5) \qquad h^i(p,b,y;\varepsilon) = d_i(p,b,y;\varepsilon)\, h_i^*(p_i,b_i,y;\varepsilon), \quad i = 1,\ldots,N$$

and

$$(6) \qquad v(p,b,y;\varepsilon) = \max[v_1^*(p_1,b_1,y;\varepsilon),\ldots,v_N^*(p_N,b_N,y;\varepsilon)].$$

Three general points about this approach to modelling extreme corner solutions are worth noting. First the key building blocks are the conditional indirect utility functions, $v_1^*(\cdot),\ldots,v_N^*(\cdot)$. Once these have been specified, the discrete choice indices can be derived from them via (3), the conditional demand functions can be derived via Roy's identity, and the unconditional demand functions via (5). Thus we can construct an extreme corner solution model directly from the $v_i^*(\cdot)$'s without having to bother with the underlying direct utility function $u(x,b,z;\varepsilon)$.

Second, the unconditional demand functions (5) embody an implicit switching regression model (i.e. a generalization of Tobit's model), since they can be expressed equivalently in the form (using the case of $N = 2$ for simplicity):

$$(7) \quad x = \begin{cases} -\dfrac{\partial v_1^*(p_1,b_1,y;\varepsilon)/\partial p_1}{\partial v_1^*(p_1,b_1,y;\varepsilon)/\partial y} & \text{if } v_1^*(p_1,b_1,y;\varepsilon) \geq v_2^*(p_2,b_2,y;\varepsilon) \\[2ex] -\dfrac{\partial v_2^*(p_2,b_2,y;\varepsilon)/\partial p_2}{\partial v_2^*(p_2,b_2,y;\varepsilon)/\partial y} & \text{otherwise.} \end{cases}$$

Thus, the random utility extreme corner solution demand model can be estimated by any of the statistical techniques developed for use with switching regression models while taking advantage of the additional restrictions inherent in the random utility formulation.

The third point is a caveat: the practical application of these models rests on the ability to devise specific functional forms for the conditional indirect utility functions and the joint density $f_\epsilon(\epsilon)$ which yield reasonably tractable formulas for the discrete choice probabilities and the conditional demand functions. Hanemann (1984a) presents a variety of demand functions suitable for extreme corner solutions which offer considerable flexibility in modelling price, income, and quality elasticities. Several of these models are applied to the Boston recreation data set in Hanemann (1983a) for the subset of households (approximately one quarter of the sample) who visited only one site over the summer and, therefore, displayed evidence of an extreme corner solution in their behavior. The remaining households visited either no sites - which can also be handled within the framework of an extreme corner solution model - or more than one site. However, none of the latter visited every site and, therefore, a general corner solution is required to model their behavior.

Theoretical Models of Corner Solution Decisions - The General Corner Solution Problem

The generalized corner solution differs from the extreme corner solution in that more than one alternative (site) is chosen and has a nonzero level of demand. One approach to characterizing general corner solutions is a straightforward generalization of that adopted above for extreme corner solutions. Instead of modelling the discrete choice decision as to which site to visit, alternative discrete choices can be viewed as combinations of sites. For example, suppose that the consumer decides to visit sites 2 and 3, but not sites 1 or 4,...,N. Conditional on this discrete choice, his utility is

$$u_{23}^{*} = u(0,x_2,x_3,0,\ldots,0,b,z;\epsilon)$$

$$\equiv u_{23}^{*}(x_2,x_3,b_2,b_3,z;\epsilon),$$

and he determines how many times to visit sites 2 and 3 by maximizing $u_{23}^{*}(\cdot)$ subject to $p_2x_2 + p_3x_3 + z = y$, $x_2 \geq 0$, $x_3 \geq 0$, and $z \geq 0$. By analogy to the previous discussion, there exist conditional demand functions for these sites and conditional indirect utility functions. Note that these functions, also, satisfy Roy's identity.

Let us consider the theoretical properties of this model of behavior. First, recall the utility maximization problem in (1)

$$\max u(x,b,z;\varepsilon)$$
$$\text{subject to } \Sigma p_i x_i + z = y$$
$$\text{and } x_i \geq 0,\ z \geq 0.$$

This problem can be rewritten as

(8) $$\max u(x,b,y - \Sigma p_j x_j - z;\varepsilon)$$
$$\text{subject to } 0 \leq x_i \leq y/p_i \quad i = 1,\dots,N.$$

The Kuhn-Tucker conditions associated with (8) are

(9) $$x_i \left| \begin{array}{l} = 0 \\ > 0 \\ = y/p_i \end{array} \right. \text{ as } \frac{\partial u}{\partial x_i} - p_i \frac{\partial u}{\partial z} \left\{ \begin{array}{l} \leq 0 \\ = 0 \\ \geq 0. \end{array} \right.$$

The Kuhn-Tucker conditions are thus nicely behaved. However a problem arises in defining demand functions and thus indirect utility functions when we allow the possibility that $\partial u/\partial x_i - p_i\, \partial u/\partial z$ may not be zero at the optimum.

To understand this, consider the following maximization problem which is a modification of (1):

(10) $$\max_x u(x,b,z;\varepsilon) \text{ subject to } \Sigma\, p_i x_i + z = y.$$

We refer to this as the "unconstrained" maximization problem. There are no non-negativity constraints and thus the unconstrained demands are implicitly allowed to take on negative values.

Additionally, we can imagine a series of "partially constrained" problems which involve equality constraints, of the form

(11) $\max_x u(x,b,z;\varepsilon)$ subject to $\Sigma\, p_i x_i + z = y$ and $x_1 = 0$

or

(12) $\max_x u(x,b,z;\varepsilon)$ subject to $\Sigma\, p_i x_i + z = y$, $x_1 = 0$ and $x_2 = 0$, etc.

These partially constrained problems are identical to the unconstrained problems except that some subset of the x's is taken, <u>a priori</u>, to be zero. The solution vectors are denoted by ${}_1x^*$ or ${}_{12}x^*$, etc., the demand functions by ${}_1h_i^*(\cdot)$ or ${}_{12}h_i^*(\cdot)$, etc. and the indirect utility functions by ${}_1v^*(\cdot)$ or ${}_{12}v^*(\cdot)$, etc. Note that ${}_jx_i^*$ is the quantity of x_i demanded when x_j and only x_j is constrained to zero. It is important to note that not all of the prices appear as arguments in these partially constrained demand and indirect utility functions. For $x_1 = 0$, the partially constrained demand function for x_i is ${}_1h_i^*(p_2,\ldots,p_N,b_2,\ldots,b_N,y)$ where ${}_1h_i^*(\cdot)$ will not be a function of p_1 or b_1. Note that this is precisely analogous to the extreme corner solution results presented above. Theorems regarding the properties of these functions are rigorously presented in Appendix 8.1, but are summarized here.

This fact that, at corner solutions, the demands for the goods which <u>are</u> being consumed are independent of the prices and qualities of those which <u>are not</u> is an interesting one. In effect, when one takes the non-negativity constraints in (1) seriously, the constrained demand function is segmented, having the general form:

$$h_1(p,b,y;\varepsilon) = \begin{cases} h_1^*(p_1,\ldots,p_N,b_1,\ldots,b_N,y;\varepsilon) \\ {}_2h_1^*(p_1,p_3,\ldots,p_N,b_1,b_3,\ldots,b_N,y;\varepsilon) \\ {}_{23}h_1^*(p_1,p_4,\ldots,p_N,b_1,b_4,\ldots,b_N,y;\varepsilon) \\ \vdots \\ \text{etc.} \end{cases} \tag{13}$$

It is instructive to relate the unconstrained case, which can be solved by simple calculus, with the more complicated set of constrained problems. In Appendix 8.1 it is shown that, if at some point in (p,y)-space the unconstrained demand functions are all positive, $h_i(p,y) > 0$ $i = 1,\ldots,N$, then the constrained maximization problem (1) has an interior solution, and conversely. Suppose, instead, that the solution to (1) includes a corner solution. Say only $x_1 = 0$. This implies that the unconstrained demand for good one is non-positive, while the partially constrained demands for all the other goods (i.e. the demands conditional on having $x_1 = 0$) are all positive: $h_1(p,b,y) \le 0$, ${}_1h_i^*(p_2,\ldots,p_N,b_2,\ldots,b_N,y) > 0 \quad i = 2,\ldots,N$.

It is interesting to observe that if we know the unconstrained indirect utility function, $v^*(p_1,\ldots,p_N,y)$, everything else - both the constrained and partially constrained indirect utility functions and the constrained and partially constrained demand functions - could be derived from this function. The key is the observation that Roy's Identity applies to the unconstrained and all the partially constrained indirect utility functions, just as to the constrained indirect utility function. Thus

$$(14) \qquad \begin{aligned} x_i^* &= - \frac{\partial v^*(p,b,y)/\partial p_i}{\partial v^*(p,b,y)/\partial y} && i = 1,\ldots,N, \\ {}_1x_i^* &= - \frac{\partial\, {}_1v^*(p_2,\ldots,p_N,b_2,\ldots,b_N,y)/\partial p_i}{\partial\, {}_1v^*(p_2,\ldots,p_N,b_2,\ldots,b_N,y)/\partial y} && i = 1,\ldots,N, \end{aligned}$$

etc.

There are some simple rules for deriving the various partially constrained indirect utility functions from $v^*(p,y)$. For example to obtain ${}_1v^*(p_2,\ldots,p_N,y)$ one solves

$$\partial v^*(p_1,\ldots,p_N,y)/\partial p_1 = 0$$

for $p_1 = \phi_1(p_2,\ldots,p_N,y)$ and substitutes ϕ_1 into v^* to obtain

$${}_1v^*(p_2,\ldots,p_N,y) = v^*[\phi_1(p_2,\ldots,p_N,y),\ p_2,\ldots,p_N,y].$$

Similarly, to obtain ${}_{12}v^*(p_3,\ldots,p_N,y)$ one solves

$$\partial\, {}_1v^*(p_2,\ldots,p_N,y)/\partial p_2 = 0$$

for $p_2 = \phi_2(p_3,\ldots,p_N,y)$ and substitutes ϕ_2 into ${}_1v^*$ to obtain

$${}_{12}v^*(p_3,\ldots,p_N,y) = {}_1v^*[\phi_2(p_3,\ldots,\ p_N,y),\ p_3,\ldots,p_N,y].$$

Proceeding in this manner, the full set of 2^N-1 partially constrained indirect utility functions can be obtained. The constrained indirect utility functions can then be constructed in the same manner as the constrained ordinary demand functions:

$$(15) \qquad v(p,y) = \begin{cases} v^*(p_1,\ldots,p_N,b_1,\ldots,b_N,y) \\ {}_2v^*(p_1,p_3,\ldots,p_N,b_1,b_3,\ldots,b_N,y) \\ {}_{23}v^*(p_1,p_4,\ldots,p_N,b_1,b_4,\ldots,b_N,y) \\ \vdots \\ \text{etc.} \end{cases}$$

Estimating General Corner Solution Models

The previous section presents the properties of demand functions and indirect utility functions when individuals can choose corner solutions (zero levels of some goods). Of course when individuals can be assumed to choose interior solutions then all the well behaved, continuous properties of neoclassical demand theory persist. Demands for systems of quality differentiated goods arise in many economic problems, however, and are characterized by general corner solution decisions. When general corner solutions prevail, we have seen that demand and indirect utility functions become discontinuous as individuals switch among different consumption regimes. The properties of these discontinuous functions can be laid out, but this is a far cry from using observed data to estimate such functions.

In nonmarket benefit analysis we use data on observed behavior to estimate models of behavior which can be linked, theoretically, to preferences. Information on preferences gives information on welfare gains and losses associated with changes in the consumer's economic environment.

In the first part of this report, we saw how observed behavior could be linked to welfare measures through estimated behavioral functions. The spirit of the task is the same here. Knowing what we do about the nature of demand in this decision making setting, we need to estimate behavioral relationships which we can subsequently relate to preferences (and thus welfare measures). Unfortunately, estimation is much more difficult when demand functions are discontinuous and when the relevant piece of the demand function is conditional on discrete choices.

There are actually three possible routes one could take in using observed behavior to estimate parameters necessary to provide information about preferences. The first is to estimate the analog to the extreme corner solution problem, as discussed at the beginning of the last section, where each "discrete alternative" is a unique combination of nonzero quality differentiated goods. The problem with this approach is that if there are N quality differentiated goods, there are 2^N combinations or discrete alternatives. In recreational examples where the number of sites can be 20 or 30, the problem soon becomes astronomical.

A second route to take is to avoid, temporarily, the discontinuities of the demand functions and focus on the Kuhn-Tucker conditions presented in (9). Suppose one observes an individual who purchases quantities $\bar{x}_1,\ldots,\bar{x}_Q$ of goods $1,\ldots, Q(\leq N)$, and

$$y - \sum_1^Q p_j\bar{x}_j$$

of the Hicksian composite commodity, but nothing of goods $Q+1,\ldots,N$. Define the N random variables $\eta_1,\ldots,\eta_N$ by

$$(16)\qquad \begin{aligned}\eta_i &= \eta_i(\bar{x},p,b,y;\varepsilon)\\ &\equiv \frac{\partial u}{\partial x_i}(\bar{x},0,b,y-\sum_1^Q p_j\bar{x}_j,\varepsilon) - p_i\frac{\partial u}{\partial z}(\bar{x},0,b,y-\sum_1^Q p_j\bar{x}_j,\varepsilon) \quad i=1,\ldots,N\end{aligned}$$

and let $f_\eta(\eta_1,\ldots,\eta_N)$ be their joint density, obtained from $f_\varepsilon(\varepsilon)$ by an appropriate change of variables. By virtue of (9), the probability of observing this consumption event is given by

$$(17)\qquad \begin{aligned}&\Pr\left\{\begin{array}{ll} x_i = \bar{x}_i, & i = 1,\ldots,Q\\ x_i = 0, & i = Q+1,\ldots,N\end{array}\right.\\ &= \Pr\left\{\begin{array}{ll} \eta_i = 0, & i = 1,\ldots,Q\\ \eta_i \leq 0, & i = Q+1,\ldots,N\end{array}\right.\\ &= \int_{-\infty}^0 \cdots \int_{-\infty}^0 f_\eta(0,\ldots,0,\eta_{Q+1},\ldots,\eta_N)\,d\eta_{Q+1}\cdots d\eta_N.\end{aligned}$$

Note that if instead the consumer purchased none of the goods, so that Q = 0 and z = y, the probability of this event would be

$$(18) \qquad \Pr\{z = y\} = \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} f_\eta(\eta_1,\ldots,\eta_N)\,d\eta_1\ldots d\eta_N.$$

If he purchases some quantity of every good (i.e. an interior solution) so that Q = N and

$$y > \sum_1^N p_j\bar{x}_j,$$

the probability would be

$$(19) \qquad \Pr\{x_i = \bar{x}_i,\ i = 1,\ldots,N\} = f_\eta(0,\ldots,0).$$

Given an entire sample of consumers located at different corner solutions, the likelihood function would be the product of individual probability statements each having the form of (17), (18) or (19). Two specific examples of this approach both based on Linear Expenditure System utility models are presented in Appendix 8.2.

Two general points emerge from this analysis which are worth emphasizing. First, the probability expressions such as (17) generally require the evaluation of an (N-Q)-dimensional cumulative distribution function - i.e. a multiple integral whose dimensionality corresponds to one less than the number of commodities not consumed. In the recreation case, where the number of sites (N) may equal perhaps 20, but the number of sites visited by an average individual (Q) will be 2 or 3, the evaluation of these integrals, while not impossible, is cumbersome.

The dimensionality problem (N-Q) is fundamental in that it is rooted in the logic of the utility maximization problem. However this way of treating the problem represents an improvement over the first approach. The discrete choices implied by the analog of (5) for general corner solutions involve, in principle, up to a (2^N -1) dimensional cumulative distribution function.

A third estimation alternative is to attempt to estimate the partially constrained demand function. Consider the probability statements for the observed consumption outcomes. If we observe an individual consuming positive quantities, $\bar{x}_i$, of every good, the probability of this event is

(20) $$\Pr[x_i=\bar{x}_i,\ \text{all } i] = \Pr[h_i^*(p,b,y;\varepsilon) = \bar{x}_i,\ \text{all } i].$$

Similarly, if we observe an individual consuming nothing of good 1 but positive quantities, $\bar{x}_i$ of every other good, the probability of this event is

(21) $$\Pr\begin{cases} x_1 = 0 \\ x_i = \bar{x}_i, & i=2,\ldots,N \end{cases} = \Pr\begin{cases} h_1^*(p,b,y;\varepsilon) \le 0 \\ {}_1h_i^*(p_2,\ldots,p_N,b_2,\ldots,b_N,y;\varepsilon) = \bar{x}_i & i = 2,\ldots,N. \end{cases}$$

In general

(22) $$\Pr\begin{cases} x_i = 0 & i=1,\ldots,Q \\ x_i = \bar{x}_i, & i=Q+1,\ldots,N \end{cases} = \Pr\begin{cases} h_i^*(p_i,p_{Q+1},\ldots,p_N,b_i,b_{Q+1},\ldots,b_N,y;\varepsilon) \le 0, & i=1,\ldots,Q \\ {}_Ah_i^*(p_{Q+1},\ldots,p_N,b_{Q+1},\ldots,b_N,y;\varepsilon) = \bar{x}_i, & i=Q+1,\ldots,N \end{cases}$$

where A is the set of x_i's not consumed.

The expressions in (22) are the probability statements associated with the indirect Kuhn-Tucker conditions and are logically equivalent to those based on the direct Kuhn-Tucker conditions, such as (17). Unfortunately, they are susceptible to the same problem of dimensionality since (22), like (17), requires in principle the evaluation of an (N-Q) dimensional cumulative distribution function. However, the probability statements derived from the indirect Kuhn-Tucker conditions may still prove advantageous. For example, there are cases when a given indirect utility function does not have associated with it a closed-form representation of the direct utility function. Thus, probability statements such as (17) cannot be employed, whereas those such as (22) are still available.

There is a second consideration in the choice among estimation techniques but one which has implications for prediction as well. There tends to be a basic trade-off between achieving simplicity in the (direct) Kuhn-Tucker conditions and in the demand functions. For example, if the Kuhn-Tucker conditions involve a simple random variable, the demand functions will typically involve ratios of random variables. (See Appendix 8.2) for an example). If a simple distribution for the random variables exists, then a simple distribution for their ratios will not, and vice versa. Thus the likelihood function for (17) may be easy to form, while the associated likelihood function for (22) will not be.

By choosing a utility function and error distribution that provides a relatively simple assessment of (17), we do not escape the need for evaluating the demand functions. If we wish to consider a hypothetical change in the individuals' environment, (prices, qualities, etc), then we must predict what his new decisions will be under the hypothetical circumstances. Prediction requires the calculation of expected demand and thus involves the assessment of probability statements as well.

Whatever the approach to estimation, we cannot escape the combinations implicit in the 2^N discrete choices when we come to construct the marginal probability distributions of the demands for individual sites, which would be needed to predict, say, the change in the expected demand for a site resulting from a change in its price or quality. Let $f_{x_1}(x_1)$ be the marginal density of $h_1(p,b,y)$. Heuristically, this density will have the general form

$$(23) \qquad f_{x_1}(x) = \int_0^{y/p_2} \cdots \int_0^{y/p_N} \Pr \begin{matrix} x_1 = x \\ x_i = \bar{x}_i, \; i=2,\ldots,N \end{matrix} \; d\bar{x}_2 \cdots d\bar{x}_N$$

where the probability statement inside the integral is given by expressions like (20), (21) and (22), depending on the region of $(x_2,\ldots,x_N)$-space. The evaluation of this marginal density and its mean, $E[x_i] = \int x f_{x_1}(x)\,dx$, may require numerical techniques.

Concluding Comments

This chapter presents the corner solution analog to the standard utility theoretic model of consumption. The introduction of corner solutions is shown to complicate the characterization of demand functions and indirect utility functions, which become discontinuous as the individual switches among different consumption patterns.

In the last section of this chapter, the implications of these discontinuous demand functions are drawn out. Three methods of estimation are outlined conceptually, but each suffers from severe dimensionality problems. Against this backdrop, the next chapter presents an overview of existing modelling techniques, each falling short of capturing the complete decision problem but each empirically quite feasible.

APPENDIX 8.1

PROPERTIES OF THE UNCONSTRAINED AND PARTIALLY CONSTRAINED PROBLEM

That it is possible to analyze corner solutions exclusively in terms of indirect utility functions has recently been proved by Lee and Pitt (1983) and Hanemann (1984d). In this appendix, we shall summarize these theoretical developments. For simplicity we switch, temporarily, to a deterministic utility setting (i.e. the vector ε is omitted) and the numeraire good, z, and the quality variables, b are ignored. Thus we can consider an indirect utility version of the Kuhn-Tucker conditions for the maximization problem:

(A1) $$\max_x u(x_1,\ldots,x_N) \text{ subject to } \Sigma\, p_i x_i = y,\ x_i \geq 0 \quad i = 1,\ldots,N$$

As before, we denote the resulting ordinary demand functions by $h^i(p,y)$, i = 1,...,N, and the indirect utility function by v(p,y), which are referred to as the "constrained" demand and indirect utility functions. The consumption vector which solves (A1) is denoted $\bar{x}$. In order to proceed, we need to introduce several additional maximization problems which are companions to (A1). One of these problems is (A1) minus the inequality constraints:

(A2) $$\max_x u(x_1,\ldots,x_N) \quad \text{subject to } \Sigma\, p_i x_i = y,$$

the "unconstrained" maximization problem. The resulting demand functions and indirect utility function are the "unconstrained" demand and indirect utility functions. The consumption vector which solves (A2) is denoted x^*. The logic of (A2) is that the consumer is implicitly allowed to purchase <u>negative</u> quantities of goods, which is meaningless economically but serves as a useful artifact for our analysis.

The remaining utility maximization problems all involve equality constraints, of the form

$$\text{(A3)} \qquad \max_{x} u(x_1,\ldots,x_N) \text{ subject to } \Sigma\, p_i x_i = y \text{ and } x_1 = 0$$

or

$$\text{(A4)} \qquad \max_{x} u(x_1,\ldots,x_N) \text{ subject to } \Sigma\, p_i x_i = y,\ x_1 = 0 \text{ and } x_2 = 0, \text{ etc.,}$$

which are "partially constrained" problems and we denote the solution vectors by ${}_1x^*$ or ${}_{12}x^*$, etc., the demand functions by ${}_1h_i^*(\cdot)$ or ${}_{12}h_i^*(\cdot)$, etc. and the indirect utility functions by ${}_1v^*(\cdot)$ or ${}_{12}v^*(\cdot)$, etc. It is important to note that ${}_1x_1^* = {}_{12}x_1^* = {}_{12}x_2^* = 0$ (i.e. the first element of ${}_1x^*$ and the first two elements of ${}_{12}x^*$ are zero) and that not all of the prices appear as arguments in these partially constrained demand and indirect utility functions. Thus,

$$\text{(A5)} \qquad \begin{aligned} {}_1x_i^* &= {}_1h_i^*(p_2,\ldots,p_N,y) \quad i = 1,\ldots,N, \\ {}_1u^* &= {}_1v^*(p_2,\ldots,p_N,y) \\ {}_{12}x_i^* &= {}_{12}h_i^*(p_3,\ldots,p_N,y) \quad i = 1,\ldots,N, \\ {}_{12}u^* &= {}_{12}v^*(p_3,\ldots,p_N,y). \end{aligned}$$

With this notation in hand, we can now describe Hanemann's findings through a series of theorems, whose proofs are omitted. These involve the following regularity conditions on the direct utility function:

(D-1) u is a continuous real-valued function defined over R^N

(D-2a) u is non-decreasing in each argument, and is strictly increasing in at least one argument.

(D-2b) If $x_i = 0$ for some, but not all, indices i, u is increasing in at least one argument x_j where $x_j > 0$.

(D-3) u is strictly quasiconcave

The purpose of the theorems is to establish various relations between the solution to the unconstrained or partially constrained problems.

Theorem 1. If $x_i^* > 0$ $i = 1,\ldots,N$, then $\bar{x}_i = x_i^* > 0$ $i = 1,\ldots,N$. Let u satisfy (D-1), (D-2a), and (D-3); then $\bar{x}_i > 0$ $i = 1,\ldots,N$ implies that $x_i^* = \bar{x}_i > 0$, $i = 1,\ldots,N$.

Theorem 2. Let u satisfy (D-1), (D-2a,b) and (D-3). Suppose $\bar{x}_1 = 0$, $\bar{x}_i > 0$ $i = 2,\ldots,N$. Then (i) $x_1^* \leq 0$, and (ii) $x_i^* = \bar{x}_i > 0$, $i = 2,\ldots,N$.

Theorem 3. Let u satisfy (D-2a) and (D-3). Suppose that $x_1^* \leq 0$ and ${}_1x_i^* > 0$, $i = 2,\ldots,N$. Then (i) $\bar{x}_1 = 0$ and (ii) $\bar{x}_i = {}_1x_i^* > 0$, $i = 2,\ldots,N$.

Given some set of indices, A, let ${}_{(A-i)}x^*$ denote the solution to

(A6) $$\max_x u(x_1,\ldots,x_N),\ \Sigma p_i x_i = y,\quad x_j = 0 \text{ all } j \varepsilon A,\ j \neq i,$$

where it is understood that the index i is a member of A.

Theorem 4. Let u satisfy (D-1), (D-2a,b), and (D-3). Suppose that, for some set of indices A, $\bar{x}_i = 0$ all $i \varepsilon A$ and $\bar{x}_i > 0$ all $i \not\varepsilon A$. Then, (i) $x_i^* \leq 0$ for at least one index $i \varepsilon A$

(ii) ${}_{(A-i)}x_i^* \leq 0$ for each $i \varepsilon A$

(iii) ${}_A x_i^* = \bar{x}_i > 0$ all $i \varepsilon A$.

Theorem 5. Let u satisfy (D-2a) and (D-3). Suppose that, for some set of indices A, $x_i^* \leq 0$ for at least one index $i \varepsilon A$, ${}_{(A-1)}x_i^* \leq 0$ for each $i \varepsilon A$, and ${}_A x_i^* > 0$ all $i \not\varepsilon A$.

Then, (i) $\bar{x}_i = 0$ all $i \varepsilon A$

(ii) $\bar{x}_i = {}_A x_i^* > 0$ all $i \not\varepsilon A$.

Theorem 1 says that if, at some point in (p,y)-space, the unconstrained demand functions are all positive, $h_i^*(p,y) > 0 \quad i = 1,\ldots,N$, then the constrained maximization problem (A1) has an interior solution, and conversely. Suppose, instead, that there is a corner solution. It turns out to make something of a difference whether the corner solution involves zero consumption of only one good or of several goods. If the former occurs, say only $x_1 = 0$, this implies that the unconstrained demand for good one is non-positive, while the partially constrained demands for all the other goods (i.e. the demands conditional on having $x_1 = 0$) are all positive: $h_1^*(p,b,y) \leq 0$, ${}_1h_i^*(p_2,\ldots,p_N,y) > 0 \quad i = 2,\ldots,N$. This is the content of Theorem 2, while Theorem 3 states the converse. Similarly Theorems 4 and 5 cover the case where the corner solution involves zero consumption of several goods.

APPENDIX 8.2

ESTIMATION OF GENERAL CORNER SOLUTIONS USING KUHN-TUCKER CONDITIONS

The Kuhn-Tucker approach to the estimation of general corner solutions was independently proposed by Hanemann (1978) and Wales and Woodland (1978). Two specific examples, both based on the Linear Expenditure System utility model, are

$$u(x,b,z;\varepsilon) = \sum_1^N \psi_j(b_j,\varepsilon_j)\ \ln(x_j + \theta_j) + \ln z \tag{A1}$$

$$u(x,b,z;\varepsilon) = \sum_1^N \theta_j \ln[x_j + a_j + \psi_j(b_j,\varepsilon)] + \ln z \tag{A2}$$

where

$$\psi_j(b_j,\varepsilon_j) = \exp[\sum_k \gamma_k b_{jk} + \varepsilon_j] \geq 0 \quad j = 1,\ldots,N$$

and the θ_j's, the γ_k's and the α_j's are the coefficients to be estimated, together with any parameters of the joint density $f_\varepsilon(\varepsilon)$. Define the constants $t_1,\ldots,t_N$ by

$$t_i = -\sum_k \gamma_k b_{ik} + \ln \frac{p_i(\bar{x}_i + \theta_i)}{y - \sum_1^Q p_j\bar{x}_j} \quad i = 1,\ldots,Q$$

(A3)

$$t_i = -\sum_k \gamma_k b_{ik} + \ln[p_i\theta_i/(y - \sum_1^Q p_j\bar{x}_j)] \quad i = Q + 1,\ldots,N.$$

Then, in the case of the utility model (A1), the probability statement (17) becomes

$$\Pr\left\{\begin{array}{ll} x_i = \bar{x}_i & i=1,\ldots,Q \\ x_i = 0 & i=Q+1,\ldots,N \end{array}\right. = \int_{-\infty}^{t_{Q+1}} \cdots \int_{-\infty}^{t_N} f_\varepsilon(t_1,\ldots,t_Q,\varepsilon_{Q+1},\ldots,\varepsilon_N)\,d\varepsilon_{Q+1}\cdots d\varepsilon_N. \tag{A4}$$

In particular if the ε_i's are independent extreme value variates, this probability has a closed-form expression:

$$\text{(A5)} \qquad \exp\left(-\sum_1^Q t_j\right)\exp\left[-\sum_1^N \exp(-t_j)\right].$$

It is interesting to note that while the probability statement is quite simple, the demand function associated with this example has the form

$$\text{(A6)} \qquad x_i = -\theta_i + \frac{\psi_i}{1+\sum_1^Q \psi_j} \; \frac{1}{p_i}\left(y + \sum_1^Q p_j\theta_j\right) \quad i = 1,\dots,Q$$

which involves the ratio $\psi_i(1+\Sigma\psi_j)^{-1}$. As we observed in the previous section, there is generally a tradeoff between having a simple distribution for the random variables $\psi_1,\dots,\psi_N$ and having a simple distribution for ratios formed from them.

In the case of the utility model (A2), define $t_1,\dots,t_N$ by

$$\text{(A7)} \qquad \begin{aligned} t_i &= -\sum_k \gamma_k b_{ik} + \ln\left[\frac{\theta_i}{p_i}\left(y - \sum_1^Q p_j\bar{x}_j\right) - \bar{x}_i - \alpha_i\right] \quad i = 1,\dots,Q \\ t_i &= -\sum_k \gamma_k b_{ik} + \ln\left[\frac{\theta_i}{p_i}\left(y - \sum_1^Q p_j\bar{x}_j\right) - \alpha_i\right] \quad i=Q+1,\dots,N. \end{aligned}$$

The probability statement (17) becomes

$$\text{(A8)} \qquad \begin{aligned} &\Pr\left[x_i = \bar{x}_i \;\; i=1,\dots,Q\right] \\ &\Pr\left[x_i = 0 \;\; i=Q+1,\dots,N\right] \\ &= \int_{t_{Q+1}}^{\infty}\cdots\int_{t_N}^{\infty} f_\varepsilon(t_1,\dots,t_Q,\varepsilon_{Q+1},\dots,\varepsilon_N)\,d\varepsilon_{Q+1}\cdots d\varepsilon_N. \end{aligned}$$

Once again, a closed form expression can readily be obtained when ε_i's are independent extreme value variates. These utility functions and stochastic specifications by no means exhaust the possibilities. For example, the approach adopted by Wales and Woodland (1978, 1983) is to assume that $u(x,b,z;\varepsilon) = \bar{u}(x,b,z) + \Sigma x_j\varepsilon_j + z\varepsilon_0$, where $\bar{u}(\cdot)$ is a nonstochastic function. Hence the $\eta_i(x,p,b,y;\varepsilon)$ functions take the form

$$\eta_i = \phi_i(\bar{x},p,b,y) + (\varepsilon_i - p_i\varepsilon_0) \quad i = 1,\dots,N$$

where $\phi_1,\ldots,\phi_N$ are nonstochastic functions. In this case it is convenient to assume that $\varepsilon_0,\ldots,\varepsilon_N$ are multivariate normal. Observe that it simplifies the modelling task if one assumes that, unlike $\partial u/\partial x_1,\ldots,\partial u/\partial x_N$, the term $\partial u/\partial z$ does <u>not</u> contain a stochastic element, a simplification which was not exploited by Hanemann (1978) or Wales and Woodland.

CHAPTER 9

A REVIEW AND DEVELOPMENT OF MULTIPLE SITE MODELING TECHNIQUES

In previous chapters we have described the nature of the recreation decision and called it a general corner solution problem. The last chapter set up a careful theoretical model of that problem - an extension of neo-classical constrained utility maximization. The latter assumes interior solutions and must be extended to incorporate the possibility of multiple corner solutions. One could attempt to estimate relevant parameters of the system by introducing stochastic elements into the theoretical model of Chapter 8, just as parameters are estimated by introducing stochastic elements into neoclassical demand functions. However, as we alluded to in Chapter 8, the direct estimation of the general corner solution model is extremely difficult involving the evaluation of a large number of integrals. While estimation by this approach is not impossible, it is costly and cumbersome. Since there already exist several ad hoc but less costly approaches to estimating demands in a multiple site framework, it is worth examining these approaches to see what characteristics of the general corner solution model are assumed away and how damaging these assumptions are.

In this chapter an indepth review of several alternative approaches are presented. The alternative models can be categorized in a number of ways. One way to subdivide the list is according to principle purpose. Some of the models were developed primarily to explain the allocation of visits among alternative sites. Others may explain allocation but are particularly applicable to the valuation of an additional site. Finally, many approaches were designed with the specific goal of valuing characteristics (principally, environmental) of sites.

A second way of subdividing the approaches is according to structure. The existing approaches can, by and large, be grouped into what might be called “demand models” and “share models”. The former explain the number of trips taken to each site while the latter take as the dependent variable the proportion of trips taken to each site. As we shall see, there is some correspondence between the subdivisions based on purpose and that based on structure.

Demand Systems in a Multiple Site Framework

Under this heading are included a number of related but distinct approaches including gravity models and multiple good analogs to the single site travel cost model. Also included is another extension of the single site travel cost model - the hedonic travel cost approach.

1. Gravity Models

One of the first treatments of multiple sites was in the context of zonal trip allocation models. In 1973 Cesario suggested the use of these gravity models for the specific purpose of explaining the allocation of trips from each zone to alternative sites. In these models visits between a zone and a site were explained on the basis of zonal and site characteristics and distance, with one set of parameters estimated for all combinations of zones and origins. For the most part such models have been used simply to estimate demand and predict use rates. Freund and Wilson (1974) provided one of the most careful applications of this approach in a study of recreation travel and participation in Texas.

In subsequent papers, such as Wennegren and Nielsen’s (1970), the gravity model was extended so that the zonal trips equation for site i included factors reflecting “competing opportunities” provided by other sites. In this example, trips were a function of price (p) and site capacity (b) and the trips to site i were assumed proportional to the following function of p_i, b_i, p_i, b_i:

$$x_i \propto (b_i^{\alpha}/p_i^{\beta}) \left(\sum_{j=1}^{N} b_j^{\alpha}/p_j^{\beta} \right)^{-1} \quad i=1,\ldots,N, \tag{1}$$

where α and β are parameters to be estimated. Presumably the introduction of the p_j, b_j made more explicit the substitutability among sites.

Gravity models have subsequently been employed in a few cases for benefit measurement. Two particularly sophisticated examples are the studies by Cesario and Knetsch (1976) and Sutherland (1982a). In Cesario and Knetsch, the zonal trips system is

$$(2) \qquad x_i = f(y,a)\phi_i^{\alpha}e^{-\beta p_i}\Big(\sum_{j=1}^{N} \phi_j^{\alpha}e^{-\beta p_j}\Big)^{-\gamma}, \quad i=1,\ldots,N,$$

where, y is income, a is a vector of individual characteristics and $\phi_1,\ldots,\phi_N$ are constructed indices of site quality. As long as $0< \gamma<1$, this formulation implies that a decrease in the cost of visiting a site and/or an increase in its quality index have <u>two</u> effects: not only is existing recreation actually diverted away from other sites to the site in question, but also some new recreation activity is generated (i.e. Σx_j increases). If $\gamma=1$, the latter effect vanishes and Σx_j remains constant. Cesario and Knetsch proceeded to treat the zonal visits equations as demand curves and take areas behind these curves as measures of consumer surplus.

The use of gravity models for benefit estimation has been limited, culminating in a rather complex paper by Sutherland published in 1982. Unlike his predecessors, Sutherland obtained predictions of <u>individual's</u> behavior rather than simply zonal aggregates. The model had four components which, while inextricably linked, were estimated independently.

In order to explain the components of the model, it is necessary to introduce a subscript for consumers, which has so far been suppressed. Let x_{it} be the demand for site i by consumer t, p_{it} the cost to consumer t of visiting site i, y_t his income, and a_t his other attributes (e.g. age, sex).

It should be noted that in part of his study (equation (28)), Sutherland treats each individual in his sample as the consumer; in the rest he treats the population of each "centroid" (a county, usually) as "the consumer". Furthermore he estimates separate demand systems for each of four recreation activities. Thus, there are four sets of x_{it}'s in the utility function, and four sets of demand functions corresponding to his equation (28). These demand systems are independent, in the sense that the demand by consumer t to visit site i for recreation activity 1 is independent of his demand to visit the same site for recreation activity 2. Moreover, each of the demand systems is modelled in a similar manner. For simplicity, therefore, we focus on the demand system for a single activity and omit activity-specific subscripts.

Thus, in terms of our previous discussion, the utility function becomes $u(x_i,b,z_t,a_t)$, and the demand functions for sites are

$$(3) \qquad x_{it} = h^i(p_{1t},\ldots,p_{Nt},b_1,\ldots,b_N,q_t,y_t;\ a_t) \qquad i = 1,\ldots,N.$$

A key element in the demand model is an impedance function $\theta_{it} = \theta(p_{it})$. Sutherland constructs the empirical probability density of the subset of p_{it}'s for which $x_{it}>0$ and then sets $\theta(\cdot)$ equal to a smoothed version of this empirical density. A feature of the function $\theta(\cdot)$, therefore, is that it is increasing over a (usually small) part of its domain, and decreasing over the remainder. The first two components of Sutherland's model are a trip production equation (i.e. a participation intensity equation)

$$(4) \qquad x_{\cdot t} \equiv \sum_i x_{it} = f(\sum_i b_i \theta(p_{it}), y_t; a_t)$$

and a site attractiveness model (i.e. an aggregate demand function for each site)

$$(5) \quad x_{i\cdot} \equiv \sum_t x_{it} = \alpha\, b_i^{\beta}(\sum_t \theta(p_{it}) \sum_i x_{it})^{\gamma}, \quad \alpha, \beta, \gamma > 0 \quad i = 1,\ldots,N$$

where b_i is a measure of facilities available at site i, $\Sigma_i b_i \theta(p_{it})$ is a measure of overall availability of recreation opportunities to consumer t, and $\Sigma_t \theta(p_{it})$ $(\Sigma_i x_{it})$ is intended as a measure of the overall accessibility of site i to the population of consumers.

Equations (4) and (5) are each estimated by OLS. Using the fitted regression equations the predicted values $\hat{x}_{\cdot 1},\ldots,\hat{x}_{\cdot T}$ are obtained from (4) and the predicted values $\hat{x}_{1\cdot},\ldots,\hat{x}_{N\cdot}$ are obtained from (5); these are substituted into the following gravity model

$$(6) \qquad \hat{x}_{it} = \hat{x}_{\cdot t}\left(\frac{\theta(p_{it})\, \hat{x}_{i\cdot}}{\sum_j \theta(p_{jt})\, \hat{x}_{j\cdot}}\right)$$

to obtain predictions, $\hat{x}_{it}$, of each consumer's visit to each site. Finally, for the purpose of valuing each site, the predicted site visits, $\hat{x}_{it}$ are regressed on the site costs via an equation of the form

$$(7) \qquad \ln(\hat{x}_{it}) = \delta_i + \lambda_i\, p_{it}$$

from which the Marshallian triangle is approximated.

By contrast with (3), the model (4) - (7) appears to be overfitting a demand system. Moreover, as with (1) and (2), the demand model does not appear to be desirable from a utility maximization standpoint, nor does it make any particular allowance for the appearance of corner solutions which certainly abound in the data set.

Sutherland's paper inadvertently exposed what is perhaps the most disturbing aspect of the gravity models. They are simply statistical allocation models based on no particular arguments about economic behavior. Consequently, when Sutherland used a gravity model to "allocate trips from zones to sites," he did not have a model of the requisite economic behavior to estimate benefits. He then was forced to re-estimate a relationship between trips and cost to capture the economic behavior implicit in a demand function. It is difficult to understand why one would wish to estimate a gravity model for benefit function and (b) if one believes that decisions are driven by economic considerations.

2. Systems of Demand Equations

Burt and Brewer (1971) were perhaps the first explicitly to specify multi-site demand models. Their motivation for going beyond the single site model was that they were interested in measuring the value of introducing a new recreational site. For such a potential value to be measurable, one needs to admit the existence of at least one other similar site. Once the existence of at least one alternative site is recognized, it seems appropriate to estimate the system of demands for all existing alternatives. Thus in deducing the value of the new site, Burt and Brewer set off to estimate how patterns of demand for existing sites would change with its addition.

The Burt and Brewer model was a straightforward extension of the single site travel cost model to a system of such demands

$$q_k = f_k(p_1, p_2, \ldots, p_m, y) \qquad k = 1, \ldots, m \tag{8}$$

where q_k is the number of trips taken to site k, p_k is the travel cost to the site k, y is income and m is the number of sites in the system considered. Any differences due to the quality characteristics of sites are assumed to show up in the estimated coefficients of the different demand functions. Unlike so many studies of this time, the authors used household rather than zonal data in their application - a study of water based recreation in Missouri.

A similar model (with the omission of income and based on zonal data) was employed by Cicchetti, Fisher, and Smith (1976) in their analysis of the Mineral King project in California. Once again the motivation was the valuation of a proposed new site. Similar to Burt and Brewer, the authors estimated a system of demands for alternative sites or site groups as functions of prices (i.e. the costs of traveling to each site). And, again, site characteristics were excluded from the model. Both versions are, in principle, consistent with the hypothesis of utility maximization. However, the version with non-zero income effects can only satisfy the integrability conditions (6) if the commodities (visits to recreation sites) are assumed to be perfect complements, consumed in fixed proportions. More importantly, it appears that each data set contains instances of corner solutions. This fact is ignored in the formulation and estimation of the demand systems.

In each case the benefits from the introduction of the new site were assessed by considering the benefits of a price change for the existing site most similar to the proposed site. Thus, gains from the new site accrued simply from reduced travel costs for some users. Hof and King (1982) asked the very pertinent question - Why do we need to estimate the system of demands in these cases? Why not just estimate the demand for the similar site (as a function of all prices) and evaluate the benefits in that market? In the context of the Burt and Brewer and the Cicchetti, Fisher and Smith papers, their arguments are cogent. If there is only one price change, its effect can be measured in one market (Just, Hueth, and Schmitz, 1982). Even if one expects seemingly unrelated regression problems, ordinary least squares will achieve the same results as generalized least squares when all equations include the same variables.

Hof and King further argued that Willig’s results provide bounds on compensating variation as functions of Marshallian consumer surplus. Thus, it is not necessary to estimate the entire demand system so as to impose cross-price symmetry and ensure path independence. In retrospect, this procedure of imposing symmetry (followed by both the Burt and Brewer and the Cicchetti, Fisher and Smith papers) seems inappropriate, since there is no reason for the Marshallian demands to exhibit such characteristics. Additionally this path independence property is not worth worrying about since the particular functional forms chosen for the system of demand functions in these papers do not meet integrability conditions (LaFrance and Hanemann, 1984). In any event, if we are interested in the effect of a single price change, there would seem no especially compelling reason to estimate an entire system of demands if they are to take the form suggested by Burt and Brewer or Cicchetti, Fisher and Smith.

All of the models mentioned so far included multiple sites to capture allocation of trips among substitute alternatives. Some of the gravity models attempted to capture the effect of site characteristics on this allocation, but were not concerned with the valuation of characteristics. The demand systems models did not even attempt to take explicit account of site heterogeneity. This was in part due to the purpose of the models and in part due to their structure. In estimating a demand system for m sites with n individuals or zones, an econometric modelling problem was encountered. While we have n observations on income and n observations on the price of each site (i.e. n x m travel costs), there is usually only one observation on the quality of each site (m total observations on quality). A site's quality characteristics do not vary over individuals (unless their perceptions are measured). Consequently site characteristics can not in general be introduced into demand systems such as (8). This does not imply that sites are considered perfect substitutes. What it does imply is that the differences in sites are not explicitly taken into account. As a result we can not predict the changes in visits to a site nor estimate the change in the value of a site, resulting from a change in a quality characteristic.

3. Varying Parameters Models

While site characteristics cannot be incorporated as separate variables in a system of demand equations, they can be incorporated by means of a varying parameters model (Freeman, 1979; Vaughan and Russell, 1982; Smith, Desvousges and McGivney, 1983; and Smith and Desvousges, 1985). The varying parameter model was first used in recreational modelling by Vaughan and Russell (1982) to determine the average value of a freshwater fishing day at fee-fishing sites. To accomplish this, they estimated a system of demand equations where the number of visits was specified only as a function of own price and income. Next, the 3 x N parameter values from these demand equations (constant, price coefficient and income coefficient for N sites) were regressed against the two observed characteristics of each site. By substitution Vaughan and Russell argued that estimation of the two stage model was equivalent to estimating one equation with observations pooled over sites. This equation is a function, though, of price, income, quality characteristics and cross product terms.

Smith, Desvousges and McGivney (1983) provided a theoretical basis for the varying parameters model based on a household production framework. They estimated the two steps separately using ordinary least squares for the first state demand functions specified as semi-long and weighted least squares for the second stage. In order to apply the two step procedure, it is necessary to have information on a fairly large number of sites, since the number of observations at the second stage equals the number of sites.

Smith et.al. estimated this model with information from participants so that the number of visits was always greater than zero. They noted that this truncation may bias the OLS estimates and employed Olsen's method of moments approximation to evaluate the importance of the bias introduced by the truncation. Based on these results, some of the demand equations from the analysis were excluded.

In a later paper, Smith and Desvousges (1985) proposed an alternative model for the first stage demand estimations. They employed a maximum likelihood estimator that explicitly reflects the truncation of the data from below. An additional truncation problem was present in their data. Any visits of six or more were lumped into one category censoring the upper bound at six visits. They found that the maximum likelihood parameter estimates were much different from the OLS estimates and that the resulting benefit estimates for most sites were three to thirty-three percent smaller with the maximum likelihood estimates. Ordinary consumer surplus measures were derived for changes in quality by determining the affect of a quality change on the predicted coefficients in the system.

4. Hedonic Travel Cost

The hedonic travel cost approach has as its sole focus the valuation of site characteristics. This approach to modelling (Brown and Mendelsohn, 1984; Mendelsohn, 1984) attempts to reveal shadow values for characteristics by estimating individuals' demand for the characteristics. This approach consists of two separate procedures. The first step entails regressing individuals' total costs of visiting a site on the characteristics of the site. If an individual visits more than one site, he is represented by more than one observation in the data set. That is, each observation is an individual/site-visited combination. It should be noted that since the costs of visiting any given site and the characteristics of the site are identical for all individuals visiting the site from the same origin, variation in the data must come from variation in the sites visited by those individuals from the same origin. With S origins, there will be S separate regressions, each presumably representing the cost function to individuals from that origin for obtaining more of the characteristics.

Brown and Mendelsohn estimated a simple linear regression for the cost function but a later piece by Mendelsohn employed a nonlinear Box-Cox transformation. The distinction is important since the partial derivatives of cost with respect to characteristics are then interpreted as the hedonic prices of the characteristics. The hedonic prices are used as prices in a second stage where the demand for characteristics is estimated. In the linear cost function case, hedonic prices are constant and do not vary over individuals from the same origin. When a nonlinear function is estimated, however, hedonic price gradients must be constructed from the first stage results. Hedonic prices will vary with characteristics levels.

Marginal value functions for quality characteristics are then estimating by regressing these derived hedonic prices for individuals from each origin to each site on the level of the quality characteristics at the relevant site and individual related variables. Brown and Mendelsohn also included an instrumental variable for the number of trips the individual took. Trips were initially regressed on the other individual-specific variables as well as dummy variables for origins. Then the predicted values were included in the marginal value functions for each characteristic.

Mendelsohn (1984) altered the second stage as well by estimating characteristic demand functions (i.e. quantity rather than hedonic price as the dependent variable). This procedure requires estimating instrumental variables for characteristic prices (in addition to visits) before including these prices in the characteristics demand function.

There are several apparent problems with this approach which may be of consequence in only some applications. The first is the absence of a good theoretical underpinning, leaving one confused as to what one is estimating. If we think about the nature of the problem, it differs substantially from the type of problem in which hedonic valuation is generally employed (i.e. in housing and labor markets). It is chance and not markets which provide the array of sites and their qualities in hedonic travel cost applications. Thus, it is unreasonable to expect costs of accessing all possible sites for all individuals to be an increasing function of even one characteristic. However the hedonic travel cost approach includes observations on costs and site characteristics only for those sites which are actually visited by individuals in the regression subsample. It is, of course, a logical result of constrained utility maximization that an individual will only incur greater costs to visit a more distant site if the benefits derived from the visit exceed those from a closer site. Nonetheless, it does not seem to follow that costs will be a single-valued, increasing function of each element of a vector of site characteristics.

The conceptual validity of the hedonic travel cost approach depends on two contentions which remain contestable and unproven. The first contention worthy of debate is whether the derivatives of the first stage regression legitimately reflect prices - the prices an individual perceives himself to have to pay to increase the level of the characteristics. If more than one characteristic is included in the function, or if important characteristics are omitted - and especially if sites are not continuous, it becomes quite possible for costs to be declining in at least one characteristic, thus producing a negative "hedonic price." This result repeatedly occurs in applications. Negative "prices" are produced in the first stage of the estimation. (see Brown and Mendelsohn, 1984; Mendelsohn, 1984; Bockstael, Hanemann and Kling, 1985). What do we do with these nonsensical prices?

Presuming for a moment that orderly prices for individual characteristics exist, the second debatable contention is that true demand functions for the characteristics can be statistically identified. This identification issue has been debated extensively in the context of the hedonic property value technique for valuing amenities, but many of the same points of controversy arise here. For a sampling of the arguments, see Brown and Rosen (1982), Mendelsohn (1983), and McConnell (1984).

The output of the final stage of the hedonic travel cost approach is a demand function for each characteristic. The demand function, although not derived from a utility maximizing framework, is interpreted to reflect the marginal willingness to pay per recreation day for an increase in the quality of the characteristic. There is an apparent inconsistency in the interpretation as we consider hypothetical movements away from the observed point. The demand functions are associated with characteristics and not sites and thus it does not seem possible to assess the value of a site specific change in quality (such as would be brought about by a regulation, etc.). A second concern is that there exists no logical relationship between characteristic demand functions and prices on the one hand and use rates of sites on the other. If we adopt a conventional measure of welfare associated with a change in a quality characteristic in the demand framework, does that reflect the value of the quality change per trip, or is it in some sense independent of the number of trips? These functions do not capture any information about how individuals' behavior (participation and site choice) would change with a change in quality. Without this latter information, it would not seem possible to assess the value of a change.

Share Models

The term "share models" denotes those models which attempt to explain the percentage of total demand allocated among discrete alternatives. In the context of this study, share models explain the allocation of total recreational visits among sites of different qualities and with different costs of access.

In what follows several existing modelling approaches are outlined which estimate share equations. In each case demand functions, $x_i = h_i(\cdot)$, for discrete alternatives $i = 1,\ldots,N$ are derived from an underlying utility function; share functions for these discrete alternatives are explicitly determined by $h_i(\cdot)/\Sigma h_j(\cdot)$. The alternative approaches are distinguished by the way in which the stochastic nature of the problem is captured. Before reviewing the existing models in this classification, we present a statement of the theory underlying share models.

1. The Theory of Share Models

In a standard neoclassical framework with a set of goods of interest x_i, $i=1,\ldots,N$ and a Hicksian good z, utility maximization yields demand functions of the form

$$x_i = h(p,b,q,y) \qquad i = 1,\ldots,N$$

where (p,q) are prices of (x,z) and b is a vector of quality characteristics. We could also derive partial demand functions (see eq. 17, Chapter 7) of the form

$$x_i = \tilde{h}_i^x(p,b,y_x) \tag{9}$$

where y_x is expenditure on the group of goods of interest, i.e. $y_x = \Sigma\ p_i \cdot x_i$.

Now suppose that we are interested in the <u>share</u> of recreation activity allocated to each site; we could consider either expenditure shares, $w_i \equiv p_i x_i / y_x$, and expenditure share functions

$$(10) \qquad w_i = w_i(p,b,y_x) = p_i \bar{h}_i^x(p,b,y_x)/y_x \qquad i = 1,\ldots,N$$

for quantity shares, $s_i = x_i/x_{\bullet}$, where $x_{\bullet} \equiv \sum_{1}^{N} x_j$, and quantity share functions

$$(11) \qquad s_i = s_i(p,b,y_x) \qquad i = 1,\ldots,N.$$

Note that the latter allocate total visitation, $x_{\bullet}$, among individual sites as a function of total expenditure on all sites, y_x, and not as a function of $x_{\bullet}$ itself.

The point worth emphasizing here is that modelling approaches based on the systems of share equations are entirely equivalent to modelling approaches based on the system of partial demand equations. Any system of share equations implies a corresponding partial demand system (up to a multiplicative factor), and conversely. Both systems convey the same amount of information about consumer preferences and behavior, with the exception that the share models do not contain information about the total demand (Σx_i) whereas the partial demand functions take that information as given.

The situation changes, however, as soon as we introduce stochastic elements and begin to think in terms of statistical models. Depending on the stochastic specification, it could make a considerable difference whether we choose to estimate share or demand systems. For example, suppose that we introduce additive, normally distributed disturbance terms into the partial demand system (9). The adding up restriction that $\Sigma p_i x_i = y_x$ induces a dependence among the disturbance terms. This forces us to assume that a subset of (N-1) of the x_i's have an (N-1)-dimensional multivariate normal distribution with mean vector $[\bar{h}_1^x(p,b,y_x),\ldots,\bar{h}_{N-1}^x(p,b,y_x)]$ and some covariance matrix. It is understood that the remaining consumption level is obtained via

(12) $$x_N = [y_x - \sum_1^{N-1} p_i x_i]/p_N$$

and hence is also normally distributed. It should be evident that it matters greatly whether we estimate the partial demand system or the share system in this case because the xi's are multivariate normal, but the distribution of the s_i's (the composition of dependent normal variates) is extremely complex and does not possess a closed form expression. Conversely, suppose we assume that the shares are multivariate normal. While this paves the way for direct estimation of the share equations, it rules out estimation of the partial demand system because there is no closed form expression for the distribution of the x_i's.

In these two examples one has to make a direct choice between a tractable distribution for the observed x_i's and a tractable distribution for the observed s_i's. When distributions other than the multivariate normal are used however, this dilemma can sometimes be avoided. For example, if we assume that $x_1,...,x_N$ are independently distributed gamma variates, the shares have the Dirichlet distribution. This approach to the modelling of shares was preposed by Woodland (1979). The estimation of the coefficients of $\bar{h}_i^x(p,b,y_x)$ can be based on either the observed quantities demanded or the observed shares. Similarly, suppose that the x_i's have a N-dimensional multivariate lognormal distribution. The first (N-1) shares have Aitchison and Chen's (1980) logistic normal distribution. Then $a = (a_1,...,a_N)$, $a_i \equiv \ln(s_i|s_N)$, $i = 1,...,N-1$, $\xi \equiv A\mu$, and $\Sigma = A\Omega A$, A being the (N-1)xN matrix $A = [I_{N-1},-e_{N-1}]$ where I_{N-1} is the identity matrix and e_{N-1} a vector of (N-1) i's.

Apart from considerations of ease of estimation, these four statistical models - normal demands, normal shares, gamma demands/Dirichlet shares, and lognormal demands/logistic normal shares - have different economic implications of importance. Both of the normal models can readily be employed when the data contain zero values for x_i and s_i; the problem is that, even if the estimated $\bar{h}_i(\cdot)$ or $s_i(\cdot)$ functions satisfy the non-negativity requirement (which is not guaranteed to happen), these models imply that one can have negative values of x_i or s_i with some probability, which is a mis-specification from an economic point of view. To be sure, this is less likely to be a serious problem for the normal demands model since the density in the negative orthant is likely to be negligible, as Woodland (1979, p. 362) points out. The gamma/Dirichlet model ensures that the estimated $\bar{h}_i(\cdot)$ and $s_i(\cdot)$ functions are positive, and it rules out the possibility of negative values for x_i and s_i. It can be applied when the data contain zero values for x_i and s_i but only if the predicted demand for the good does not exceed one. The distribution of $x_i = 0$ and the gamma density does not possess the standard bell shape. If $\gamma_i = \bar{h}_i^x(p,b,y) > 1$ the density possesses a conventional shape, but the domain over lognormal/logistic normal model does not impose any a priori restrictions on the sign or magnitude of $\bar{h}(\cdot)$, and it rules out the possibility of negative values for x_i and s_i, but its domain is restricted to $x_i > 0$ and $0 < s_i < 1$ $i = 1,\ldots,N$. Thus it, too, cannot be applied to data containing zero values for the x_i's or s_i's.

A special case of the share model arises when the multinomial distribution is employed. Total demand (Σx_i) is taken to correspond to the number of independent drawings or trials, the individual demands (x_i) are the counts, and the shares (s_i) are the probabilities. The models to be described below derive from this construct.

2. The Morey Model

Morey has used a modelling approach based on the multinomial distribution in a series of articles designed to explain the allocation of visits among alternative sites (Morey 1981, 1984a) and to value the introduction of a new site (Morey, 1984b). Despite the fact that the modelling approach focuses on the trip allocation and valuation of sites, it is closely aligned in spirit to many of the other approaches discussed in this section of the report. The concept of quality differentiated multiple alternatives is a central theme. Even though the approach has not been used in this way, it could be employed to value characteristics of sites, since the approach places a heavy emphasis on site characteristics.

The appealing feature of the multinomial distribution is that if the shares are assumed to follow such a distribution, then the implied x_i's are "counts" and therefore non-negative integers. Many commodities of interest to economists are indivisible and thus the x_i's are, strictly speaking, constrained to be integers. This is perhaps the nature of much recreational data: individuals typically consume a small number of indivisible units of the good (e.g. trips to sites). The standard scenario underlying the multinomial distribution is that R independent trials are held and, on each trial, N mutually exclusive outcomes may occur, with π_i being the probability of the i^{th} outcome where $\pi_i > 0$ and $\Sigma\pi_i = 1$. Let t_i be the number of times that the i^{th} outcome occurs in R trials. The probability of an outcome vector $(x_1,\ldots,x_N)$ is

(13) $$f(t_1,\ldots,t_N) = \frac{R!}{\prod_1^N t_j!} \prod_1^N \pi_j^{t_j} .$$

Applications of the multinomial distribution, such as Morey's, equate the count t_i with the observed demand for the i^{th} good x_i, π_i with the share function $s_i(p,b,y_x)$, and R with $x. \equiv \Sigma\, \hbar_i^x\, (p,b,y_x)$, and write the density of the observed demands as

(14) $$f(x_1,\ldots,x_N) = \frac{(x.)!}{\prod_1^N x_j!} \prod_1^N s_j(p,b,y_x)^{x_j}.$$

Parameters of the $s_j(\cdot)$ functions are then estimated by maximizing the likelihood function

(15) $$L = \prod_{m=1}^{M} \frac{x.!}{\prod_1^N x_j!} \prod_{j=1}^{N} s_j(p,b,y_x)^{x_j}$$

where M is the number of individuals in the sample and N is the number of sites.

The probability density in (14) requires that $s_i(p,b,y_x) > 0$, and hence that $\hbar_i\, (p,b,y_x) > 0$, $i = 1,\ldots,N$, but the domain of x_i includes 0. Although it can be applied to data sets containing zero values of the observed x_i's, a problem arises in that it will never predict zero demands or zero shares.

There is another conceptual problem in the application of the multinomial distribution to consumption data which needs to be recognized. The logic of the statistical model in (14) is that the number of trials, R, is exogenous and, therefore, this parameter is ignored in maximizing (15) to obtain estimates of $\pi_1,\ldots,\pi_N$. However, R equals $x_. \equiv \Sigma\, \hbar_j^x\, (p,b,y_x)$ which contains information on the coefficients to be examined, and should not be ignored in maximizing the likelihood function derived from this approach.

When applying this approach one must still choose a utility function to generate the share equations, and herein lies a third difficulty. From a practical standpoint, it is difficult to identify a utility function which has both desirable economic <u>and</u> statistical properties.

When applying the model to recreation data, Morey uses two different utility functions in trying to overcome this problem. In Morey (1981) he uses a CES subfunction

$$(16) \qquad u(x,b) = \left(\sum_{1}^{N} \phi_j(b_j)x_j^{\rho}\right)^{1/\rho} \quad \rho < 1, \quad \rho \neq 0$$

where $\phi_j(b_j)$ is the overall quality index for site j, and the parameters to be estimated are ρ and the coefficients of the $\phi_j(\cdot)$'s. This utility function implies homothetic partial demand functions and, therefore, share functions which are independent of y_x. The share functions look like

$$(17) \qquad s_i(p,b,y_x) = [\phi_i(b_i)p_i]^{\sigma} \{\Sigma[\phi_j(b_j)/p_j]^{\sigma}\}^{-1} \quad i=1,\ldots,N$$

where $\sigma \equiv (1 - \rho)^{-1} > 0$ is the (common) elasticity of substitution among different commodities (recreation sites).

Because of the homotheticity property, the above utility model implies that all commodities have a unitary income elasticity of demand, which is implausible in the recreation context. Recognizing this, Morey (1984) employs instead the following version of Pollak and Wales' (1978) Quadratic Expenditure System indirect utility function

$$(18) \qquad \bar{v}(p,b,y_x) = -\frac{g(p,b)}{y_x} + \frac{g(p,b)}{f(p,b)}$$

where

$$g(p,b) = \Sigma[\psi_j(b_j)^{\sigma}p_j^{1-\sigma}]^{1/(1-\sigma)}$$

and

$$f(p,b) = \Sigma[\phi_j(b_j)^{\sigma}p_j^{1-\sigma}]^{1/(1-\sigma)},$$

$\psi_j(b_j)$ and $\phi_j(b_j)$ being different quality indices for site j. The parameters to be estimated are now σ and the coefficients of the $\phi_j(\cdot)$'s. The resulting partial demand functions are

$$(19)\quad \tilde{h}_i(p,b,y_x) = y_x \frac{(\phi_i/p_i)^\sigma}{\Sigma\ \phi_j^\sigma\ p_j^{1-\sigma}} + \frac{y_x^2}{f(p,b)} \frac{(\phi_i/p_i)^\sigma}{\Sigma\ \phi_j^\sigma p_j^{1-\sigma}} - \frac{(\phi_i/p_i)^\sigma}{\Sigma\ \phi_j^\sigma p_j^{1-\sigma}},\ i = 1,\ldots,N.$$

Notice that these functions are no longer independent of y_x since the utility function is not homothetic. (The model actually estimated in Morey (1984) differs from (19) because the variable $x_\bullet$ is substituted for the variable y_x - i.e. the partial demand functions and share equations take the form $\tilde{h}_i(p,b,x_\bullet)$ and $s_i(p,b,x_\bullet)$). In any event, construction of share equations from these demand functions, while conceptually straightforward, is empirically cumbersome.

An additional problem with the QES model is that the demand function in (19) is based on the presumption of an interior solution but can in fact return negative or zero values for x_i and s_i. In contrast, the CES function presented earlier presumes interior solutions and completely precludes zero values of the x_i's for any set of finite prices. Neither set of characteristics is desirable. Recreational demand for sites is of course non-negative but most multiple site demand problems include zero levels of demand; corner solutions are frequently observed.

Regardless of whether they can be applied to data containing zero shares, none of the statistical models presented thus far is compatible with the economic phenomenon of corner solutions. In fact, irrespective of the form of the utility function chosen, the multinominal density is not necessarily a desirable tool for analyzing such data. This is true even though the multinominal density attaches a non-zero probability to the event that t_i=0 and can, therefore, be applied in practice to consumption data containing corner solutions. The problem is that in order for the density to be well defined, π_i, the probability of chosing i, must be strictly positive for all i = 1,...,N. If one identifies these parameters with the share system $s_1(p,b,y_x),\ldots,s_N(p,b,y_x)$, as Morey (1981, footnote 14) does, this implies that the true demand for each good is positive and the only reason for observing zero shares in some particular data set is sampling variation rather than a structural feature of economic behavior - in the same way that when one tosses a fair coin several times it is possible through sampllng variation to obtain a run of heads and no tails. Admittedly, the fitted $s_i(\cdot)$ functions could take very small values, so that the expected consumption of any particular good is very small, but this is not a satisfactory solution to the problem of corner phenomena from an economic point of view. Assuming that there are corner solutions for reasons other than sampling variation, economic theory requires that

$s_i(p,b,y_x) = 0$ for some range of (p,b,y_x)-space. As we showed in the last chapter, the internal structure of the $s_i(\cdot)$ functions changes when this occurs. The change in structure is not captured by (18) or (19), which are based on the presumption of an interior solution to the consumer's choice problem.

The CES utility model (16) precludes corner solutions since it implies that all goods are essential; hence the $s_i(\cdot)$ functions in (17) always satisfy $s_i(\cdot) > 0$. The CES model seems unsuited to recreation behavior; it is hard to believe that individual recreation sites are all essential goods.

3. Discrete Choice Models

An alternative approach is to retain the multinominal model but interpret the parameters, $\pi_1,\ldots,\pi_N$, not as shares per se but as choice probabilities arising from some structural economic model which at least implicity incorporates the possibility of corner solutions. Variations of this approach can be found in Hanemann (1978), Binkley and Hanemann (1978), Caulkins (1982), Feenberg and Mills (1980), and Morey and Rowe (1985). Recalling the expression for the multinominal distribution in (13)

$$f(t_1,\ldots,t_N) = \frac{R!}{\prod_1^N t_j!} \prod_1^N \pi_j^{t_j},$$

we now employ a different interpretation. Rather than treat the allocation of total demand, we are now concerned with the decision of what site to visit on each choice occasion. Thus π_j becomes the probability that alternative j is chosen on the given choice occasion and t_j equals 1 if j is chosen, 0 if it is not. In this way of structuring the problem, the expression $R!/\Pi t_j!$ disappears, since the number of repeated trials is 1. Finally the likelihood function takes the form

$$L = \prod_{m=1}^{M} \prod_{g=1}^{G} \prod_{j=i}^{N} \pi_{jgm}^{t_{jgm}} \tag{20}$$

where m indexes the individual, g indexes his choice occasions, and j indexes the alternatives. In this formulation π_i is still constrained to be strictly positive but this does not preclude s_i equaling zero since π is no longer a share but instead the probability of choosing alternative i on a given choice occasion.

This different way of employing the multinominal distribution implies a different underlying economic model. In the previous approach the consumer is assumed to select his entire portfolio of goods in an interdependent fashion - i.e. all at one time, as is implied by conventional utility models. However, implicit in (20) is the assumption that the individual makes a separate choice of which good to buy on each choice occasion. In the context of the recreational problem, he makes a separate choice of which site to visit each time he engages in the recreational activity.

This economic model is different from the Morey model because the macro-allocations now involve the determination of $x_.$ rather than y_x, and the micro-allocations involve the allocation of $x_.$ among individual sites. We should emphasize that the assumption of an allocation of $x_.$ rather than y_x is still not consistent with the conventional economic framework of utility maximization (except in very restricted cases). However, without estimating directly the complex model of Chapter 8, this inconsistency can not be avoided.

The micro-decision is to allocate a fixed total consumption, x, among N different goods, where we recognize that the goods are indivisible and can only be consumed one unit at a time. This indivisibility is explicitly recognized in the solution of the consumer's maximization problem, but in a special manner: instead of assuming that the consumer selects his entire portfolio of visits to the different recreation sites at a single instant (e.g. at the start of the recreation season), as is implied by conventional utility models, we now assume that he makes a separate choice of which site to visit each time he engages in the recreation activity. Given the predetermined total number of trips ("choice occasions"), $x_.$, we introduce the choice vectors $d_r = (d_{1r},\ldots,d_{Nr})$, $r = 1,\ldots,x_.$, where $d_{ir} = 1$ if the i^{th} site is selected on the r^{th} trip and $d_{ir} = 0$ otherwise.

If we wished, we could allow for the possibility that the quality attributes or costs of visiting the sites vary over different choice situations, but this complication in notation is avoided here. Let z_r denote the consumption of non-recreational goods and y_r the consumer's income on the r^{th} choice occasion. Finally, let $f(d_r,b,z_r)$ be the consumer's utility function relevant for the r^{th} choice occasion. Given the macro-allocation decision, the micro-decision is to

(21) $$\underset{d_1,\ldots,d_{x.},z}{\text{maximize}} \sum_{r=1}^{x.} f(d_r,b,z_r)$$

$$\text{subject to} \quad \sum_{j=1}^{N} p_j d_{jr} + z_r = y_r \qquad r = 1,\ldots,x_. \,.$$

This maximization problem can be decomposed into x separate problems, that is, a separate decision problem for each choice occasion of the form

(22) $$\text{maximize } f(d_r,b,z_r) \text{ s.t. } \sum_j p_j d_{jr} + z_r = y_r.$$

In order to describe the solution, suppose that on the r^{th} choice occasion, the individual has selected site i. Conditional on this decision, his utility is

(23) $$u_{ir} = f(0,\ldots,0,1,0,\ldots,0,b_i,y_r-p_i)$$
$$\equiv v_i(b_i,y_r-p_i),$$

if we assume that $f(\cdot)$ satisfies weak complementarity. We will refer to $v_1(\cdot),\ldots,v_N(\cdot)$ as <u>conditional</u> indirect utility functions. Since the consumer selects the site which yields the highest utility, the solution to (22) can be expressed in terms of these conditional indirect utility functions as

(24) $$d_{ir} = \begin{cases} 1 & \text{if } v_i(b_i,y_r-p_i) > v_j(b_j,y_r-p_j) \text{ all } j \\ 0 & \text{otherwise.} \end{cases}$$

For estimation purposes, it is necessary to introduce a stochastic element into this demand model. In the context of discrete choices, such as arise here, this is commonly done by introducing a random element directly into the utility function producing what is known as a random utility maximization (RUM) model. The idea is that, although the consumer's utility function is deterministic for <u>him</u>, it contains some elements which are unobservable to the econometric investigator and are treated by the investigator as random variables. These elements will be denoted by the random vector ε, and the utility function will be written $f(d_r, b, z_r; \varepsilon)$, where we are implicitly assuming that the distribution of ε is stable across choice occasions.

If the model satisfies weak complementarity, the conditional indirect utility functions have the general form $u_{ir} = v_i(b_i, y_r - p_i, \varepsilon)$, $i = 1, \ldots, N$. Furthermore if we assume that the random elements enter the utility functions in such a way that they, too, are affected by weak complementarity, then we can write the conditional indirect utility functions as functions of the scalar ε_i so that[2]

$$u_{ir} = v_i(b_i, y_r - p_i; \varepsilon_i) \qquad i = 1, \ldots, N. \tag{25}$$

The consumer's utility maximizing choice is still expressed in terms of the conditional indirect utility functions along the lines of (24), except that the discrete choice indices $d_{1r}, \ldots, d_{Nr}$ are now random variables with a mean $E\{d_{ir}\} \equiv \pi_{ir}$ given by

$$\pi_{ir} = \Pr\,\{v_i(b_i, y_r - p_i; \varepsilon_i) > v_j(b_j, y_r - p_j, \varepsilon_j) \quad \text{all } j\}. \tag{26}$$

If the random variables $\varepsilon_1 \ldots, \varepsilon_N$ are assumed to be independently and identically distributed extreme value variates, and the conditional indirect utility functions, (25), take the form

$$u_{ir} = v_i(b_i, y_r - p_i) + \varepsilon_i \qquad i = 1, \ldots, N, \tag{27}$$

then the logit model of discrete choices is generated

$$\pi_{ir} = e^{v_i} / \sum_{j=1}^{N} e^{v_j} \qquad i = 1, \ldots, N. \tag{28}$$

Alternatively the random variables could follow McFadden's (1978) Generalized Extreme Value Distribution

$$(29) \qquad \Pr\{\varepsilon_1 < s_1, \ldots, \varepsilon_N < s_N\} = \exp[-G(e^{-s_1}, \ldots, e^{-s_N})]$$

where G is a positive, linear homogeneous function of N variables.[3] When combined with (27) this yields discrete choice probabilities of the form

$$(30) \qquad \Pi_{ir} = e^{v_i} G_i(e^{v_1}, \ldots, e^{v_N}) / G(e^{v_1}, \ldots, e^{v_N}) \qquad i = 1, \ldots N$$

where $G_i(\cdot)$ is the partial derivative of $G(\cdot)$ with respect to its i^{th} argument. In either case, the formulas for the choice probabilities may be substituted into the multinomial density (13) for maximum likelihood estimation of the parameters in the $v_i(\cdot)$ functions and of any other parameters that may have been introduced into the joint density of the ε_i's.

In estimating these parameters, the total number of trips, $x_.$, is treated as an exogenous constant. However its determination is now ignored but is the subject of the macro-allocation decision, to which we now turn. The only way to determine $x_.$ in a manner logically consistent with an overall utility maximizing choice is to assume that $x_.$ emerges from a sequence of separate decisions: on each day the individual decides both whether to participate in recreation on that day and which site to visit if he does participate. This is conceptually the approach taken by Caulkins and by Feenberg and Mills who consider the choice occasion to be each day of the year and each day of the recreational season respectively.

To explain, suppose the season has R days, and let θ_r represent the participation decision on the r^{th} day, where $\theta_r = 1$ if the individual participates and $\theta_r = 0$ if he does not. As before, $d_{1r}, \ldots, d_{Nr}$ represents the choice of a site on the r^{th} day, conditional on a decision to participate. In addition let $u_{0r} = v_0(y_r; \varepsilon)$ measure the individual's utility if he does not particpate in recreation on the r^{th} day. His overall decision problem is to

(31) $$\underset{\theta,d,z}{\text{maximize}} \sum_{r=1}^{R} [\theta_r f(d_r,b,z_r;\varepsilon) + (1-\theta_r)\, v_0(y_r;\varepsilon)]$$

subject to

$$d_{ir} = 0 \text{ or } 1, \quad \sum_i d_{ir} = 1, \qquad r = 1,\ldots,R$$

$$\theta_r = 0 \text{ or } 1 \qquad r = 1,\ldots,R$$

$$\sum_i p_i d_{ir} + z_r = y_r \qquad r = 1,\ldots,R$$

which can be decomposed into R separate problems of the form

(32) $$\underset{\theta_r,d_r,z_r}{\text{maximize}} \quad \theta_r f(d_r,b,z_r;\varepsilon) + (1-\theta_r)\, v_0(y_r;\varepsilon)$$

subject to the constraints. On any one day the probability that the individual participates in recreation, π_{*r}, is given by

(33) $$\pi_{*r} \equiv \Pr\{\theta_r=1\} = \Pr\{v_0(y_r;\varepsilon) < \max[v_1(b_1,y_r-p_1;\varepsilon),\ldots,v_N(b_N,y_r-p_N;\varepsilon)]\}$$

while the probability that he visits site i, conditional on deciding to participate, is given by

(34) $$\begin{aligned}\pi_{ir} &= \Pr\{d_{ir} = 1 \mid \theta_r=1\} \\ &= \Pr\{v_i(b_i,y_r-p_i;\varepsilon), \\ &\quad j=1,\ldots,N \mid v_i(b_i,y_r-p_i;\varepsilon) > v_0(y_r;\varepsilon)\}\end{aligned}$$

This type of problem is encountered frequently in economics and can be handled in a straightforward manner with conventional discrete choice models.

The expected number of visits to all sites over the season can be determined by $E[x_\bullet] = \Sigma\pi_{*r}$ while the expected number of visits to the i^{th} site would be $E[x_i] = \Sigma\, \pi_{ir}\pi_{*r}$. The logic of this formulation is that the participation decision (which ultimately determines how many trips are made over the season) and the site choice are interdependent and are made simultaneously by the individual.

From an econometric view, one can estimate the participation and site choice decisions simultaneously or, with some loss of efficiency, separately using a GEV model. Again, suppose that

$$u_{or} = v_0(y_r) + \varepsilon_0 \tag{35}$$

and let the joint density of $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_N$ be GEV with

$$G(t_0, t_1, \ldots, t_N) = t_0 + [\Sigma(t_i^{1/(1-\sigma)})]^{1-\sigma} \tag{36}$$

where $\sigma \varepsilon [0,1)$ is, in effect, the common index of correlation for $\varepsilon_1, \ldots, \varepsilon_N$. The probability of selecting the i^{th} site conditional on a decision to participate in recreation on any given day is

$$\pi_i = (e^{v_i/(1-\sigma)}) \left[\sum_j e^{v_j/(1-\sigma)}\right]^{-1} \qquad i = 1, \ldots, N. \tag{37}$$

Note that this looks like the simple extreme value model in (28) except for the normalizing constant $(1-\sigma)^{-1}$. Define the "inclusive value", I, by

$$I \equiv \ln(\sum_j e^{v_j/(1-\sigma)}). \tag{38}$$

The probability of participation in recreation on any day is then

$$\pi_* = e^{(1-\sigma)I}/(e^{v_0} + e^{(1-\sigma)I}). \tag{39}$$

Given an estimate of I from the analysis of site choices, the analysis of participation intensity based on (39) yields estimates of σ and $v_0(\cdot)$.

Variants of this type of model have been applied to recreation demand by Feenberg and Mills (1980) and Caulkins et al. (1984), but with some differences. For the purpose of analyzing site choices, Feenberg and Mills set $\sigma = 0$ in (36) and (37) - i.e. they employed the standard logit model (28). With this model, their inclusive value index was

$$(40) \qquad I' \equiv \ln\left(\sum_{1}^{N} e^{v_j}\right).$$

They proceeded to estimate two alternative regression models of the form

$$x_{.} = -\bar{v}_0 + \gamma I' \qquad \gamma \neq 1$$

$$\ln x_{.} = -\bar{v}_0 + \gamma I' \qquad \gamma \neq 1,$$

but neither of these is consistent with the total recreation demand model which would result from their model.

Caulkins et al. (1984) also set $\sigma = 0$ in their analysis of site choices, and they employed a binary logit model estimated by maximum likelihood for intensity of recreation participation. But, their logit model was based on a participation probability of the form

$$(41) \qquad \pi_* = e^{\Upsilon}/(e^{v_0} + e^{\Upsilon})$$

where, instead of being given by (40) Υ is a linear function of the average price and quality characteristics of the various sites. Because of this difference, the site choices and the recreation participation decisions are not mutually consistent, in the sense of being derived from a single underlying utility maximization.

The above two studies have one characteristic in common, the total number of trips taken in a season (i.e. the macro-allocation decision) is determined indirectly by adding up the number of independent occasions upon which the individual chooses to participate in recreation. Treating the macro decision as the sum of totally uncoordinated micro decisions is not especially appealing. Without attempting to estimate the corner solutions model of the previous chapter however, there appears to be no consistent way to link independent discrete choice decisions and a macro decision for total trips with a common underlying utility maximization framework.

It is possible to construct a model which, while not rigorously derived from a single utility maximization problem, nonetheless captures the nature of both the micro and macro allocation. This approach is adopted in Chapter 10. A similar approach was used by Hanemann (1978).

Let us employ a discrete choice model such as those presented in (28) or (37) for the site choice decision, but then let us specify total recreation demand as some function of the cost and quality of the recreation opportunities available to the individuals, capturing these dimensions for example by an inclusive value index such as that given in (38). The macro relationship has much the same properties as the single equation models presented in the first part of this report. For example, it is important to capture the fact that some individuals do not participate in recreation at all. Consequently the estimation approach must take into account the fact that $x_{\bullet}$ may be either positive or zero.

The expected number of visits to all sites over the season may be cast in the form

$$E\{x_{\bullet}\} = E\{x_{\bullet} | x_{\bullet} > 0\} \Pr\{x_{\bullet} > 0\}, \tag{42}$$

where the second term on the right-hand size is the probability that the individual participates at all in recreation. The expected number of visits to the i^{th} site is given by

$$E\{x_i\} = \pi_i E\{x_{\bullet} | x_{\bullet} > 0\} \Pr\{x_{\bullet} > 0\} \tag{43}$$

where π_i is given by (28) or (37), i.e. the discrete site choice model.

A comparison of this approach with the Feenberg and Mills or Caulkins et al. models exposes an important difference. In this model the probability that an individual is not a recreationalist, i.e. he does not participate at all in the recreational activity j is estimated directly. Drawing on the discussion of Chapter 3, either Tobit or Heckman procedures can be used to estimate this equation. This latter procedure is particularly appropriate if factors such as old age, ill health or preferences for other activities cause an individual never to recreate. In the previous approach where total visits are determined by the summation of independent decisions on sequential choice occasions, nonparticipants happen, in a sense, by accident. They are predicted to be those individuals who happen to have a string of zero predicted responses to a sequence of N independent micro decisions. Modelling the macro allocation separately would appear to be a more realistic and useful description of individual behavior.

Welfare Measurement Given The Nature of Recreation Decisions

One can certainly argue with features of all of the models outlined above. Here we will be concerned with only one criteria, albeit an extremely important one, for assessing alternative models. The criteria is how adequately each model captures the appropriate benefits which accrue from art environmental change, given the nature of recreational decisions in a multiple site framework.

It is important at this point to reiterate and to develop more fully what we mean by the nature of recreational decisions. Suppose we are interested in valuing an improvement in water quality, and we attempt to do this by looking at recreational behavior over an array of recreational sites with different water quality in the region of interest. Any sample of the relevant population will turn up a fair number of individuals who do not participate in water recreation at these sites at all. Of those who do participate in the activity, it will be unusual to find anyone who visits all sites. It will also be unusual if the entire data set consists of individuals each of whom visit only one site. Additionally, we are interested in how many trips an individual takes to each site. Thus we observe either that an individual did not participate in the activity at all or that he participated but took no trips to several sites and a positive number of trips to some subset of sites.

Recreational behavior is complicated to model because of this mix of continuous and discrete decisions and because decisions result invariably in corner solutions. Nonparticipants are, of course, at a corner solution with respect to the total trips decision. Participants are also at corner solutions since they take zero trips to at least some sites (i.e. there are zero levels of these commodities). One of the drawbacks of the straightforward demand systems modelling of Burt and Brewer and Cicchetti, Fisher and Smith is that the models are predicated on the assumption of interior solutions to the utility maximization process. Once we admit to corner solutions, the nature of demand systems changes.

This criticism is in some ways applicable to the share models as well. The share models treat the total number of trips as fixed. Additionally most of these models implicitly presume a nonzero share (however small) for all sites. The share models can be transformed into demand systems and estimated in that form, providing predictions of total number of trips. However such models suffer from the same problem as the Burt and Brewer type models in that they presume interior solutions. Many of the discrete choice approaches get around the problem by estimating decisions <u>per</u> <u>choice occasion.</u> This ignores interdependence across trip decisions and provides estimates of total trips demanded only in an indirect and unsatisfactory way. The final discrete choice model suggested above attempts to mitigate the second of these criticisms, but does so in a way which is not completely consistent with a utility maximization framework.

Given the complexities of the decision making process, a pertinent question at this point is: How important is it to model behavior, if we are interested simply in valuing changes in characteristics (e.g. environmental improvements)? The answer to this question is critical. The costs of obtaining good models of behavior in this context are high and we need to know whether they are worth it.

One can debate the importance of wholly consistent, utility theoretic models. What is much more certain is the importance of estimating effectively the complex dimensions of recreational demand. There are two reasons for this. Estimation can be biased if account is not taken of corner solutions (see for example the literature on truncated and censored samples). More important for our purposes here, welfare measurement in this context depends on the behavioral adjustments of individuals.

Consider once again the water quality example. Suppose there are N sites and water quality is improved at one of these sites, j. It is true that those who visit site j will benefit. How much they benefit will be affected by how many trips they take to site j - a decision which might change with the improvement of the site. Additionally, recreationalists who did not previously visit site j may now find it desirable and may move from corner solutions for visits to site j to positive demands. Finally, we may find the improvement of site j attracting previous nonparticipants into the recreational activity.

Now suppose more than one site's quality is improved, a more likely result of regional implementation of an environmental regulation. Then, depending on the pattern of improvements, all sorts of re-orderings may take place. Some sites may be improved but may generate no user net benfits because they actually lose visits to other more improved sites. Clearly the welfare gains to an individual at any one site are conditioned on his decision to visit the site and must be adjusted by the probability of that site being visited. Models which do not take into account changes in behavior can not accurately measure benefits.

Concepts of Welfare Evaluation in a Stochastic Setting

In this section we discuss some issues arising when the estimated multiple site demand models are used to derive money measures of the effect on an individual's welfare of a change in the prices or qualities of the available recreation sites. We assume that the demand functions are compatible with the hypothesis of utility maximization, so that the underlying indirect utility function can be recovered from them, and we are concerned with exact welfare measures rather than Marshallian approximations.

The basic theory of welfare measurement for quality changes was developed by Maler (1971, 1974) in the context of a deterministic (i.e. non-random) utility function which ignored the possibility of corner solutions. Given an indirect utility function, $v(p,b,y)$, and some change in the set of prices and qualities facing an individual consumer from (p',b') to (p'',b''), two appealing measures of the effect of this change on his welfare are the compensating and equivalent variations, C and E, defined respectively by

$$v(p'',b'',y-C) = v(p',b',y) \quad (44)$$

$$v(p'',b'',y) = v(p',b',y+E).$$

Defined in this way, C and E measure not only the direction of the change in welfare, i.e,

$$\text{sign}(C) = \text{sign}(E) = \text{sign}[v(p'',b'',y) - v(v',b',y)],$$

but also the magnitude of the change. The link between the C and E measures for pure quality changes and the conventional compensating and equivalent variations for pure price changes is explored in Hanemann (1980a), where it is shown that standard results on the sign of (C-E) and the relation between C or E and the usual Marshallian measure of consumer's surplus carry over from price to quality changes in some cases.

The task of performing welfare evaluations is more complex when one works in a random utility setting. The theory of welfare measurement in this context has been developed by Hanemann (1982c), and revised and extended in Hanemann (1984c). We will provide a sketch of this theory here, leaving the reader to refer to these papers for a more detailed presentation. Both deal with extreme, rather than general, corner solutions but these can involve either purely discrete choices as in the logit models (28) and (30), or mixed discrete continuous choices.

The starting point for this welfare analysis, as for demand analysis is, the set of N conditional indirect utility functions discussed in Chapter 8, $v_1(p_1,b_1,y_1;\varepsilon),\ldots,v_N(p_N,b_N,y;\varepsilon)$, from which the unconditional indirect utility function $v(p,b,y;\varepsilon)$ may be obtained. This gives the utility attained by the consumer when confronted with the choice set (p,b,y). The utility known to <u>him</u>, but a random variable for the econometric investigator, who can, at best, derive a cumulative distribution function for the v's: $F_v(w) \equiv \Pr\{v(p,b,y,;\varepsilon) < w\}$. For example, when the $v_i(\cdot)$'s have the form given in (27), i.e. additive stochastic terms, then

(45) $$F_v(w) = F_\varepsilon(w-v_1,\ldots,w-v_N).$$

By analogy with (44), the compensations required by the individual to offset the change from (p',b') to (p'',b'') are given by

$$(46)\qquad v(p'',b'',y-C;\varepsilon) = v(p',b',y;\varepsilon)$$

$$v(p'',b'',y;\varepsilon) = v(p',b',y+E;\varepsilon).$$

The problem in the random utility context is that C and E are now random variables, since they depend implicitly on ε. How, then, do we obtain a single number representing the compensating or equivalent variation for the price/quality change?

Hanemann (1984c) presents three different approaches to welfare evaluations in the random utility context, only one of which has previously been recognized. That approach is based on the expectation of the individual's unconditional indirect utility function,

$$V(p,b,y) \equiv E[v(p,b,y;\varepsilon)].$$

In terms of this function, the measure of compensating variation is the quantity C' defined by

$$(47)\qquad V(p'',b'',y-C') = V(p',b',y).$$

This measure has been employed by Hanemann (1978, 1982c, 1983a), McFadden (1981), and Small and Rosen (1982). The formulas needed to calculate V( ) for some common logit and probit additive-error random utility models are summarized in Hanemann (1982c). For example, in the GEV logit model (30),

$$(48)\qquad V(p,b,y) = \ln G(e^{v_1},\ldots,e^{v_N}) + 0.57722\ldots,$$

which is simply the inclusive value index (apart from Euler's constant, 0.57722. . .). This formula will be used in the empirical application in the next chapter.

Another possible welfare measure is

$$C^{+} \equiv E[C], \tag{49}$$

i.e. the mean of the individual's true (but random) compensation. The distinction between C^{+} and C' is subtle, but important. C^{+} is the observed expectation of the maximum amount of money that the individual could pay after the change and still be as well off as he was before it. By contrast, C' is the maximum amount of money that the individual could pay after the change and still be as well off, in terms of the observer's expectation of his utility, as he was before it.

A third possible welfare measure is derived as follows. One might want to know the amount of money such that the individual is just at the point of indifference between paying the money and securing the change or paying nothing and foregoing the change. For the observer, this could be taken as the quantity C* such that

$$Pr[v(p'',b'',y-C^{*};\varepsilon) > v(p',b',y;\varepsilon)] = 0.5, \tag{50}$$

i.e. there is a 50:50 chance that the individual would be willing to pay C* for the change. It can be shown that, while C^{+} is the mean of the distribution of the true compensation C, C* is the median of this distribution.

The procedures for calculating C^{+} and C* are described in Hanemann (1984 c). Here we wish to emphasize that the three welfare measures, C', C^{+} and C*, are in principle different, and the choice between them requires a value judgment on the part of the analyst. However, there are some circumstances in which some or all of them coincide. For example, in additive-error GEV models (which includes the standard logit model as a special case) Hanemann (1984c) proves that C' = C*. Similarly, in cases where the conditional indirect utility functions have the special form

$$v_i(p_i,b_i,y;\varepsilon) = \phi_i(p_i,b_i;\varepsilon) + \gamma y \quad i = 1,\ldots,N \tag{51}$$

where $\gamma > 0$ is a constant that does not vary with i, he shows that $C' = C^{+}$. Thus, when $\phi_i(\cdot)$ in (51) involves an additive stochastic term that is a GEV variate, $C^{*} = C' = C^{+}$. However, (51) is a highly restrictive assumption since it implies that both the discrete choices and the continuous choices (i.e. the conditional demand functions) are independent of the individual's income. (This assumption of no income effects is employed by both McFadden (1981) and Small and Rosen (1981)).

If a) there are income effects or b) there are no income effects but the conditional indirect utility functions do not involve additive GEV variates, the difference between C^{+} and C^{*} can be substantial because the distribution of C, the true but random compensation, tends to be rather skewed, being bounded by zero at the low end but by income or $-\infty$ at the high end (Hanemann, 1984). Thus its mean, C^{+}, may substantially exceed its median, C^{*}.

Hanemann (1983b, 1984c) also compares these welfare measures with an alternative calculation that was performed by Feenberg and Mills (1980) and Meta Systems (1983) using a logit model of recreation site choice. These authors were concerned with evaluating the benefits from an improvement in quality at an individual site - say, site 1. Thus, b_1 changes from b_1' to b_1'' while $b_2,\ldots,b_N$ and $p_1,\ldots,p_N$ remain constant. Using the nonstochastic component of the conditional indirect utility function for this site, they calculated the quantity $\tilde{C}$ defined by

$$v_1(p_1,b_1'',y-\tilde{C}) = v_1(p_1,b_1',y). \tag{52}$$

By contrast, Hanemann shows that

$$C' \approx \pi_1' \tilde{C} \tag{53}$$

$$C^{+} = \pi_1' \tilde{C} + \text{other terms.} \tag{54}$$

where π_1' is the probability that the individual selects site 1 when faced with (p_1,b_1,y).

The point is this: if we know for sure that the individual would select site 1, then C would indeed be the approprite welfare measure. But, we can never be sure in the random utility context: there are only probabilities of site selection (i.e. $\pi_1' < 1$), and we must weight the benefit C by the probability that the individual would have selected this site in the first place. (In the data set used by Feenberg and Mills and Meta Systems, this probability was on the order of 0.05 to 0.2 for most sites.) Alternatively, if we wish to use the C^+ measure, equation (54) says that to the quantity $\pi_1 \tilde{C}$ we must add some other terms which measure the expected benefit to the individual if he originally selected some site other than 1 but switched to site 1 as a consequence of the improvement in its quality. In principle, the net effect of the two corrections to $\tilde{C}$ implied by (54) is $\tilde{C}$ could be larger or smaller than C^+. From some numerical simulations reported in Hanemann (1984c), it appears more likely that $\tilde{C}$ overstates C^+ as well as C'.

In the next chapter, an application is presented using one of the discrete choice models outlined earlier in this chapter. While not perfect, this approach allows many of the crucial aspects of welfare measurement discussed here to be captured.

1. Although we treat z_r as the numeraire, the specification of these two variables for practical purposes is somewhat troublesome. Although it might seem natural to require that $\Sigma z_r = z$ and $\Sigma y_r = y$, it may not be desirable to assume that $z_r = z/x_.$ and $y_r = y/x_.$ because this implies that, if $x_.$ changes, y_r and z_r change. The effective budget constraint facing the individual changes with $x_.$, which introduces an income-change component into welfare evaluations for price/quality changes, as explained in Chapter 8. It would be more convenient to assume that y_r is, say, monthly or weekly income (i.e. $y_r = y/12$, etc.) and z_r is correspondingly proportional to z.

2. This simplication is not crucial; it is omitted, for example in random coefficients versions of the discrete choice model on the lines of Hausman and Wise (1978).

3. One could also assume that the $\varepsilon_1,\ldots,\varepsilon_N$ have a multivariate normal distribution, which produces a probit model of discrete choice. However, the choice probability formula corresponding to (29) involves an (N-1) dimensional multivariate normal integral. In the application below where N = 30 this is computationally infeasible.

CHAPTER 10

ESTIMATION OF A MULTIPLE SITE MODEL

The estimation of a multiple site model and the calculation of benefits from a hypothetical improvement in water quality at different sites are the subjects of this Chapter. The purpose of the estimation is to make operational previous modelling discussions rather than to test specific hypotheses about swimming behavior in Boston. In what follows we demonstrate the application of the two-part discrete/continuous choice model presented in Chapter 9. In making any approach operational, the difficulties of measuring and incorporating information about environmental quality becomes immediately apparent. The first part of this Chapter is dedicated to their discussion. While no definitive statement is made as to the "proper" set of water quality measures, results add to the optimism that welfare changes can be observed from physical measures of water quality. It is hoped that subsequent research will provide a more substantive guide to this perplexing problem.

Measurement of Water Quality Change

Up to this point, water quality has been blithely treated as an easily measured and universally accepted vector, b. Confronted with having to estimate the effect of water quality changes on individual behavior, the problem must now be addressed of measuring a water quality vector consistent with the objective of assessing welfare changes from improved water quality.

1. Objective Measures and Perceptions

By far, the most frequently used measures of water quality in the natural science literature are scientifically measurable concentrations of elements or organisms in the water. Because the techniques of measurement are consistent in many attributes, they are referred to as objective measures. They are frequently combined into indices which are intended to, reveal general levels of contamination of the water.

Yu and Fogel (1978) note that there are more than 100 water quality indices in use throughout the United States. An example is the index developed by St-Louis and Legendre (1982) for seven lake beaches in Quebec. This index was based on monitoring data from ten years using three groups of bacteria (total coliforms, fecal coliforms, and streptococci).

Often, the index is used in conjunction with epidemiological data to determine empirically the biological effects of water quality on human health. Hendry and Toth (1982) used total coliform data and swimmer health data to determine effects of land use on the bacteriological water quality in an Ontario lake. The frequency of ear infections in the population of swimmers was significantly associated with the amount of swimming in the lake. Another example is in Cabelli et al. (1983), a three year epidemiological study of beaches in New York, Louisiana, and Massachusetts. Telephone interviews of weekend swimmers eight-to-ten days after their swimming inquired as to possible gastrointestinal illness. The authors concluded that swimming in polluted water does increase risk of acute gastroenteritis, and that the risk occurs even at beaches that are far cleaner than the existing recommended guideline (1000 total coliforms/100 ml). The research also provided evidence that some organisms such as enterococci bacteria are good indicators of health hazards, and others, such as fecal coliform, are not.

In contrast to these pursuits of physical and biological scientists, the 1970's saw the arrival of social scientists and psychologists in the water quality field and a new emphasis on behavioral changes limited to water quality perceptions. In the period that followed, a number of strategies were pursued to elicit individuals' perceptions of water quality.

Barker (1971) determined what criteria Toronto lake users and lakeside residents used to identify water pollution. Around fifty percent of users and residents stated that the appearance of water indicated pollution; about one in seven cited odor. Nearly twenty five percent of both groups responded that they could not identify polluted water, and the remainder relied on posted signs or the publication of test results to assess the quality of water. David (1971) asked Wisconsin households to describe polluted water. Algae, murky water and debris were most frequently cited. Both Barker and David, thus found that visual characteristics were the most likely to be cited as indicating water pollution.

Ditton and Goodale (1973) compared differences in perceptions among three groups of Wisconsin recreators - swimmers, fishermen and boaters. Respondents generally tended to characterize Green Bay in terms appropriate to that part of the Bay where they lived. Moreover forty seven percent indicated “unpleasant smell”, and fifteen percent chose “too many weeds” as Green Bay’s most bothersome physical characteristic. “Wind”, “waves” and “water temperature” were selected by between four and seven percent of respondents. Green Bay’s most bothersome water quality characteristic was determined to be “dead fish”, followed by “bacteria”, “foam”, “cloudiness” and “chemicals”.

Ditton and Goodale also distinguished perception on the basis of Bay use. For example, users were most likely to cite “cold water”, “winds”, and “cloudiness” than nonusers, and nonusers were most likely to cite factors such as “foam”, “chemicals”, and “bacteria”. User perceptions differed somewhat among subgroups. Swimmers were more likely to describe the Bay as “dirty” and to be bothered by “cold water” and “junk on the bottom”. Boaters were more frequently bothered by “winds” and “weeds”. The responses of fishermen were generally between those of swimmers and boaters. However, no differences were found among groups with respect to the most bothersome water quality characteristic.

Kooyoomjian and Clesceri (1974) also compared perceptions across users including fishermen, residents, and recreationists (swimmers, boaters, and sightseers). In general, the users of low quality lakes appeared to have more intense complaints than did users of high quality lakes, and the water quality problems of the smaller lakes were more intense than the larger lakes. Similar to the Ditton and Goodale study, different problems bothered particular user groups more than others. The fishermen as a group objected to surface effects (roughness, oil films) and crowding effects (too many boats and waterskiers). The residents as a group objected to shoreline problems, odors, color, and taste. The recreationists objected to water contact factors, such as cold water temperature and bottom conditions.

Given that visual cues are apparently of particular importance, one final study which relies exclusively upon visual pollution stimuli is worth discussion. Dinius (1981) designed and used what he called a Visual Perception Test (VPT) to examine water quality perceptions. He found that increased water discoloration was perceived as indicating increased pollution. The water quality of increasingly littered sites was also judged more negatively, even when the water itself had not actually been altered. In general then, litter and water discoloration were found to interact in producing perceptions of undesirable water quality.

2. The Correlation between Perceptions and Objective Measures

There is a potential dilemma inherent in the results of the aforementioned studies: water quality policy is directed toward changing objective measures whereas benefits from the policy are argued to arise from changes in perceptions. If there is an inconsistency between objective measures and perceptions, then there is a major obstacle to valuing the benefits from "improved" water quality. It is possible that improvements in water quality by objective standards may not be perceived by individuals. Individuals not perceiving the improvement will not alter their behavior, and economists using indirect market methods (which depend upon behavior) to measure the benefits will not detect any change. For example, if bacteria counts at beaches are lowered and there is less illness but nobody ascribes it to the beaches, measuring benefits by looking at changes in beach use will be fruitless.

It is therefore critical to establish a link between water quality perceptions and objective measures. If it happened that objective measures were highly correlated with the sorts of stimuli that people perceive and provoke behavior change, then the core of the problem would be eliminated. If not, then perhaps alternative methods of benefit analysis should be pursued.

There exists a literature, albeit inconclusive, on the relationship between perceptions of water quality and objective measures. Bouwes (1983), for example, conducted a state-wide telephone survey in Wisconsin in which respondents were asked to give water quality ratings on specific lakes. The "ratings" ran from O (good) to 23 (bad). Objective ratings, Utformark's Lake Condition Index (LCI), were regressed on the subjective ratings. A low correspondence was found. It is not completely clear why an index is used in this study since it arbitrarily forces subjective weights on individual objective measures.

Binkley and Hanemann (1978) consider individual objective measures as well as less complicated yet more specific subjective ratings. Twelve water quality characteristics and five subjective variables were used. They also sought information on the characteristics which people perceived important to good water quality. Respondents reported clarity and absence of floating debris as the most important characteristics. The correlation between perceived and objective water quality measures was statistically significant although not much of the variance in individual perceptions was explained. Other studies (e.g. Dornbusch (1975); Bouwes and Schneider (1979)) have provided greater correlation, but it seems appropriate to conclude that the link between perceptions and objective measure of water quality remains in question.

In scrutinizing the existing attempts to establish a relationship between perceived and objective measures of water quality, it is clear that these studies have suffered from less than perfect design and less than high quality data. For example, the Bouwes (1983) and Bouwes and Schneider (1979) studies are based on a small sample size and also employ rather arbitrary indexing. This indexing may obscure any correspondence which actually exists. Also the subjective measure is a peculiar ranking of O to 23 (from good to bad). This is not a scale to which people would easily adapt. Additionally, adding responses to such a ranking across people is suspect.

In most studies it is presumed that individuals have accurate perceptions of the water quality at sites which they use. As a consequence only users' perceptions are regressed on objective measures. There is a sample selection bias inherent in this procedure. If water quality does indeed affect recreational decisions and if tastes for and/or perceptions of water quality vary over the population, then we will be collecting a biased sample when we interview users or interview on site. At the extreme, one should be able to predict the answer to the following question asked of Mr. Z found visiting Beach A: "Do you find the water quality here acceptable?" Of course he does or he would not have come.

Perhaps the most critical flaw in past studies lies in the measurement of perceptions. It is difficult to draw a correspondence between objective measures and perceptions if we cannot precisely define perception. Questions which have been devised to elicit perceptions tend to be open-ended and qualitative. Because of the "fuzzy" nature of perceptions, quantitative analysis of this sort may be doomed to failure.

There is one redeeming factor in all of this confusion. As long as we can be convinced that a relationship does exist between perceptions and objective measures, it is not necessary to know this relationship to measure benefits from policies based on objective measures. Suppose we have information which shows a policy change measured in objective terms influences behavior. For purposes of benefit analysis, it is not necessary to know the structure of linkage between objective and subjective measures. It is sufficient to know the reduced form effect.

3. Quality in the Proposed Model

Maybe a better rationale for being optimistic about individual response to objective (or subjective) water quality measures is that they tend to be statistically significant factors in determining demand for water-based recreation. Hanemann (1978), for example, found that objective quality measures (such as color, turbidity and fecal coliform) do affect the decision as to which beach to attend even though they do not appear to affect how often one visits beaches. Caulkins, Bishop and Bouwes (1982) also found objective measures of quality cause individuals to choose particular beaches and also affect the frequency of recreation trips. The previously discussed Lake Classification Index (LCI) was used to measure water quality in this study. Finally, Russell and Vaughan (1982) have had substantial success in using catch or bag rates in determining the demand for sportfishing. This represents a different, but conceptually similar, way of capturing quality through objective measures.

We are, thus, somewhat comforted in knowing that objective measures of quality have been used successfully in recreational demand models. While we do not yet know as much as we would like about the perceptions issue, this does not prevent us from proceeding. In the application to follow, objective measures of water quality are employed.

Specification of the Discrete/Continuous Choice Model of Recreational Demand

In this section the discrete/continuous choice model outlined in Chapter 9 is applied to recreational swimming data from the Boston area. While the current study supported collection of new recreational data for the Chesapeake, analysis of this data could not be accomplished in time for this report and will be published in a subsequent volume. The estimation presented here has two components. The first is the macro-decision: does an individual participate in the activity of interest (swimming at beaches in the Boston-Cape Cod area), and if so how many trips does he take in a season? The second component is a site allocation decision: on each choice occasion, which site does he visit? This latter is structurally equivalent to a share decision: what proportion of the total visits are made to each site? Because the micro decision generates information necessary for estimation of the macro-decision, we deal with the micro-decision first.

1. The Micro Allocation Decision - Choice Among Sites

The first part of the model involves the estimation of the household's choice among sites. It will be important here to capture those elements - hopefully cost and water quality - which affect the site chosen. McFadden (1976) provides a utility theoretic framework for employing the multinomial logit model which is applicable to a discrete choice problem of this sort. Suppose we call v_i^* a latent variable denoting the level of indirect utility associated with the i^{th} alternative. The observed variable Y_i has the property that

(1) $$Y_i = 1 \quad \text{if } v_i^* = \max(v_1^*, v_2^*, \ldots, v_M^*)$$
$$Y_i = 0 \quad \text{otherwise.}$$

Indirect utility associated with the i^{th} alternative is some function of z_i, a vector of attributes of the i^{th} alternative, so that $v_i^* = v_i(z_i) + \varepsilon_i$. The random component is generally attributed to the systematic, but unmeasurable, variation in tastes and omitted variables. Thus, each household has a level of error which, in a sense, remains with it over time. If the ε's are independently and identically distributed with type I extreme value distribution (Weibull), then it is well known that

(2) $$\text{Prob}\ (Y_i=1 \mid z) = \frac{e^{v_i}}{\sum_{j=1}^{M} e^{v_j}},$$

(see equation 28, Chapter 9, as well as Maddala 1983; McFadden, 1973; Domencich and McFadden, 1975). The likelihood function for the sample is

(3) $$L = \prod_{i=1}^{M} \left(\frac{e^{v_i}}{\sum_j e^{v_j}} \right)^{gi}$$

where $g_i = 1$ if i is chosen, $g_i = 0$ otherwise.

The multinomial logit has a property which, while in some circumstances is useful, in others is unrealistic. The model persented above implicitly assumes independence of irrelevant alternatives, i.e. the relative odds of choosing any pair of alternatives remains constant no matter what happens in the remainder of the choice set. Thus, this model allows for no specific pattern of correlation among the errors associated with the alternatives; it denies - and in fact is violated by - any particular similarities within groups of alternatives.

McFadden (1978) has shown that a more general nested logit model, specifically incorporating varying correlations among the errors associated with the alternatives, can also be derived from a stochastic utility maximization framework. If the ε's have a generalized extreme value distribution then a pattern of correlation among the choices can be allowed. The GEV (generalized extreme value) model is presented in equation (30) Chapter 9, but its derivation can be found in McFadden (1978) and Maddala (1983). For our purposes we merely state the results. McFadden defines a probabilistic choice model

$$(4) \qquad P_i = \frac{e^{V_i} G_i(e^{V_1},\ldots,e^{V_N})}{G(e^{V_1},\ldots,e^{V_N})}$$

where G_i is the partial of G with respect to the i^{th} argument and $G(e^{V_1},\ldots e^{V_N})$ has certain properties which imply that

$$(5) \qquad F(\varepsilon_1,\ldots,\varepsilon_n) = \exp\{-G(e^{-\varepsilon_1},\ldots,e^{-\varepsilon_n})\}$$

is a multivariate extreme value distribution. When $G(e^{V_1},\ldots,e^{V_N})$ is defined as Σe^{V_i}, then the model reduces to the ordinary multinomial logit (MNL) described above. However when

$$(6) \qquad G(Y) = \sum_{m=1}^{M} a_m \left(\sum_{i \in S} e^{V_i/(1-\sigma_m)} \right)^{1-\sigma_m}$$

where there are M subsets of the N alternatives and $0 < \sigma_m < 1$, then a general pattern of dependence among the alternatives is allowed. The parameters, σ_m, can be interpreted as indices of the similarity within groups.

Suppose we were to classify the alternatives into these M groups designated $S_1,\ldots,S_M$ as implied above, and we were interested in the probability of choosing alternative i, then

$$(7) \qquad P_i = \sum_{m=1}^{M} P(i|S_m)P(S_m),$$

where

$$P(i|S_m) = \begin{cases} \dfrac{e^{v_i/(1-\sigma_m)}}{\sum_{j\varepsilon S_m} e^{v_{jm}/(1-\sigma_m)}} & \text{if } i=S_m \\ 0 & \text{Otherwise} \end{cases} \quad (8)$$

and

$$P(S_m) = \frac{a_m \left(\Sigma_{j\varepsilon S_m} e^{v_{jm}/(1-\sigma_m)} \right)^{1-\sigma_m}}{\sum_{n=1}^{M} a_n \left(\Sigma_{k\varepsilon S_n} e^{v_{kn}/(1-\sigma_n)} \right)^{1-\sigma_n}} \quad (9)$$

The above GEV model is useful in many applied discrete choice problems. Frequently, alternatives group themselves in obvious patterns of substitutability. If they do, it is both convenient and appropriate to estimate the GEV model. It is appropriate because the results of an MNL will be invalid if such a pattern actually exists. It is convenient because it reduces the number of alternatives included at each stage.

To make the estimation process explicit, let us consider the following form of v_{im}

$$v_{im} = \theta' Z_{im} + \phi' W_m \quad (10)$$

where the Z's denote attributes associated with all sites and the W's are associated solely with the choice with the subset S_m. Also let us assume that σ_m is identical within all groups and equal to σ. Define the "inclusive value" of group m as

$$I_m = \ln \sum_{i\varepsilon S_m} e^{\theta' Z_{im}/(1-\sigma)} , \quad (11)$$

the probabilities can be rewritten as

$$P_{i|m} = \frac{e^{\theta' Z_{im}/(1-\sigma)}}{\sum_{k\varepsilon S_m} e^{\theta' Z_{km}/(1-\sigma)}} \quad (12)$$

and

(13) $$P_m = \frac{e^{\phi' W_m + (1-\sigma) I_m}}{\sum_{j=1}^{M} e^{\phi' W_j + (1-\sigma) I_j}}.$$

These probabilities can be estimated using MNL procedures. First, the $P_{i|m}$ are estimated with M independent applications of the multinomial logit. Note that at this stage θ is not recoverable, but can be estimated only up to a scale factor of $1-\sigma$. From the results of (12), the inclusive prices (11) are calculated and incorporated as variables in the second level of estimation (13). Here the ϕ's and the σ are estimated. A σ outside the unit interval is inconsistent with the underlying utility theoretic model and suggests misspecification.

It should be noted here that two step estimation i.e. the estimation of (12) and (13) Independently is not necessarily efficient. Amemiya (1973) explores this property of the model and presents a correction factor. However, even Ameniya suggests that the cost in computational complexity is probably not worth the gains. We consider McFadden's estimation method adequate and use it to estimate a GEV model in the next section.

2. The Macro Allocation Decision - Participation and Number of Trips

This part of the model is a single activity model. As such, at least some of the concerns in Part I of this Volume are applicable. Of particular importance is Chapter 4 which discusses the treatment of participants and nonparticipants. Here we have the sort of problem which commanded our attention in the Chapter. How do we estimate a demand curve for an activity for which a substantial portion of the population chooses zero visits? Because we have data on nonparticipants as well as participants, either of two methods can be employed, the Tobit model or the Heckman model. They are both plausible explanations of behavior and reasonably easy to apply.

The Tobit model presumes that individual's decisions can be described as

(14) $$\begin{aligned} x_i &= h(z_i) + \varepsilon_i && \text{if } h(z_i) + \varepsilon > 0 \\ x_i &= 0 && \text{if } h(z_i) + \varepsilon_i < 0. \end{aligned}$$

There is some sense that $h(z_i) + \varepsilon_i$ is defined and has meaning in the negative quadrant, but that its realization, x, is constrained by real world constraints to be nonnegative. The important thing is that the decisions - whether or not to participate and how much to participate - are dictated by the same forces. The likelihood function for model (14) is

$$(15) \qquad L_T = \prod_{i \varepsilon s} f\left(\frac{-h_i/\sigma}{\sigma}\right) \prod_{i \varepsilon s} F(-h_i/\sigma)$$

where s is the set of individuals who participate.

In contrast, the Heckman model takes the following form

$$(16) \qquad \begin{aligned} x_i &= g_2(z_{i2}, \beta_2) + \varepsilon_{i2} > 0 \qquad &\text{if } g_1(Z_{i1}, \beta_1) + \varepsilon_{i1} > 0 \\ x_i &= 0 &\text{if } g_1(Z_{i1}, \beta_1) < 0 \end{aligned}$$

which is in some sense more general. Here different factors can affect the participation decision and the demand for trips decision. Even if the same variables are believed to affect both decisions, they may enter with different coefficients.

The model in (16) can be handled by first estimating a probit whose likelihood function is

$$(17) \qquad L_p = \prod_{i \varepsilon S} (1-F(-g_{i1}/\sigma_{11}) \prod_{i \varepsilon s} F(-g_{i1}/\sigma_{11}))$$

and then estimating an OLS equation of the form

$$(18) \qquad x_i = g_2(z_{i2}, \beta_2) + \frac{\sigma_{12}}{\sigma_{22}} \lambda_i + \nu_{12}$$

where $\lambda_i = f(-g_{i1}/\sigma_{11})/(1-F(-g_{i1}/\sigma_{11}))$ and is calculated from the parameters of (17), and ν_{12} is a random term.

It should be noted that if we had data only on participants, the Heckman-type model could not be used. However, the Tobit can be used because the likelihood function in (15) can be modified to reflect the fact that the probability of drawing a particular observation is conditioned on participation (see Chapter 3 for a discussion).

The Data and Model Estimation

1. Micro-Allocation Model

In this section the data set is described and the specific models to be estimated are defined. The data set is one collected in the Boston area some ten years ago. There are advantages to using this Boston data in our illustration of modelling technique. It has been analyzed by three other research efforts (Binkley and Hanemann, 1978; Hanemann, 1978; Feenberg and Mills, 1980). Its repeated use is a testimony to its quality. The data set contains information on both participants and nonparticipants, as it is based on random household interviews in the Boston SMSA. For each participant, a complete season's beach use pattern is reported, including the number of trips to each beach in the Boston area. Observations are defined by households although socio-economi c information about the respondent is collected. This, of course, begs the question of the individual's role in a household's decision process. Although the topic is worthy of study, no attempt is made to address it at this time. Throughout, the terms household and individual are used interchangeably, abstracting from complex and potentially important distinctions.

The data set has strengths but also a number of shortcomings. For one thing, there is no useable data to capture the value of time. Consequently the concepts introduced in Chapter 4 can not be applied. Additionally, there are a large number of sites - 30 for which we have objective measures of water quality and which account for about 70% of household visits. It is unlikely that all of these sites are known to all households *in* the survey, but there is no way of determining what each household's actual choice set is. Finally there is no recorded data on quality perceptions for any but the beaches most frequently visited by the household. Thus we do not know how the individual perceives beaches which he does not visit but with which he might be familiar. Difficulties encountered applying multiple site models to the Boston data set provided guidance in designing the Chesapeake swimming survey.

The GEV model developed in the last section is particularly appropriate for the Boston data set. Among the thirty sites, eight are beaches at fresh water lakes and twenty-two are salt water beaches. It would seem reasonable to suppose that the odds of choosing fresh water site A over salt water site B will be disrupted by the addition of another fresh water lake site. Put another way, fresh water and salt water sites are probably viewed as closer substitutes within groups than across groups.

In this section we employ the GEV model of the previous section to individuals' choices where individuals are viewed (a) as choosing between fresh and salt water and (b) as choosing among fresh water sites conditioned on the fresh water choice and choosing among salt water sites conditoned on the salt water choice. In actuality, the problem is set up so that the individual chooses the "best" within the group of salt water sites and the "best" within the group of fresh water sites and then chooses between these two "best" alternatives on each choice occasion.

The determinants of most interest in choosing among sites are (a) the site characteristics which vary over alternatives and (b) the costs of gaining access to the sites. Costs were constructed as the sum of travel and entrance costs, where travel costs were simply estimated at a fixed rate of 7 cents per mile (1974 data). Mileage was determined from an adjustment of straight line distances.

Twelve objective measures of quality were available for each site: oil, turbidity, color, pH, alkalinity, phosphorus, nitrogen, ammonia, COD, fecal coliform, total coliform and temperature. While these measures are not necessarily the best measures of water quality, they are preferred to the quality measures provided in more recent analysis of this data because they were collected in the relevant year and because they provide information on more than one dimension of quality. Consistent with Hanemann's results, correlation among groups of water quality variables was found. Consequently, not all parameters could be included in the model. The patterns of correlation helped in choosing quality variables for inclusion. Also, in light of the discussions earlier in this Chapter, those objective measures which are most likely to be either directly observable or highly publicized were selected. The quality variables chosen for this model included oil, turbidity, fecal coliform, chemical oxygen demand and temperature.

Three other variables were identified as potentially valuable in the site choice model, each of which is a restricted variable of sorts. The variable "Noise" was set to one for all beaches which were in particularly noisy, congested areas close to freeways (zero otherwise). The variable "Ethnic" was set to one if the beach was especially popular with a particular ethnic group and the individual was not of that group (zero otherwise). Several beaches were so designated in the study. Finally, "Auto" was set to one if a beach was not accessible by public transportation and the household did not own a car.

Because of the nature of the logit model, variables which are present in the indirect utility function but do not change across alternatives (i.e. individual specific) tend to cancel out upon estimation - that is, their coefficients cannot be recovered. This is true unless it is argued that an alternative specific variable has a different effect depending on the value of a socioeconomic variable, in which case the two variables could be entered interactively.

Income is a special individual specific variable because we know from utility theory (see discussion in previous chapter) that income and price must enter the indirect utility function in the form Y-p. Thus if $Y-p_i$ enters linearly into Vi, income will cancel out upon estimation (because the estimated model will be a function of $V_i - V_j$), but the coefficient on price will be income's implicit coefficient as well. This will be important in calculating benefits.

The models estimated in the first stage of the GEV estimation are

$$P_i | \text{ salt} = \frac{e^{\Sigma\theta_{ks} Z_{ik}/(1 - \sigma)}}{\sum_{j \in J_S} e^{\Sigma\theta_{ks} Z_{jk}/(1 - \sigma)}} \qquad (19)$$

and

$$P_i | \text{ fresh} = \frac{e^{\Sigma\theta_{kf} Z_{ik}/(1 - \sigma)}}{\sum_{j \in J_F} e^{\Sigma\theta_{kf} Z_{jk}/(1 - \sigma)}} \qquad (20)$$

where Z_k, k=1,...,K include: oil, turbidity, fecal coliform and temperature readings at sites; noise, ethnic and public access restricted variables; and travel costs. J_F and J_S are the fresh water and salt water choice sets respectively.

Estimation of the second stage of the model requires the calculation of inclusive values from each of the first stage estimations, where the inclusive price is as defined in (11). This "inclusive value" captures the information about each group of sites in Stage I. Thus if water quality were to change at some sites, the inclusive values would change. Additionally we postulate that other variables besides the inclusive value may enter at this stage - variables which affect the salt-fresh water decision but do not vary over alternatives within each group. Also, since the fresh-salt water decision is dichotomous, it is straightforward to enter individual specific variables which are believed to affect salt water and fresh water decisions differently. Besides a constant term and the inclusive price, the size of the household, the proportion of children and whether or not the household has access to a swimming pool are included.

$$(21) \quad P_{salt} = \frac{1}{1 + e^{\phi_0 + (1 - \sigma)(I_F - I_S) + \phi_1(-W_1) + \phi_2(-W_2) + \phi_3(-W_3)}}$$

where I_F = freshwater inclusive value, I_S = saltwater inclusive value, W_1 = household size, W_2 = precent of children in household, and W_3 = access to swimming pool.

2. Macro-Allocation Model

Several variables were selected to help explain the macro-allocation decision of Boston households. With some prior testing and consideration of Hanemann's results, it was determined that the following household characteristics were most likely to affect this decision:

- income
- size and composition of household
- education
- length of work week of household head
- ownership of water sports equipment.

Additionally, variables must be included which reflect the cost and quality of the swimming activities available. Herein lies one of the major difficulties with this "second best," two part approach. How does one choose appropriate variables for the cost and quality of swimming excursions, if those trips are, or can be, taken to different sites with different costs and quality characteristics? Ideally the decision of how much and where to go should be model led simultaneously (see Chapter 8). However, "second best" models are unable to handle these problems simultaneously and require some approximations.

To capture the effect of travel cost, Hanemann used the average distance from the individual's home to all sites as a proxy for this cost. He omitted water quality altogether from this equation because only objective measures of quality were available and, thus, the average quality for sites did not vary over individuals in the sample. In any event, a simple average distance over all sites would seem problematic.

Indeed, we wished to include variables which reflected the quality and costs of the best alternatives for each individual, not necessarily the characteristics of the closest site or the average characteristics over sites. The inclusive value concept has an appealing interpretation since it represents, in a sense, the value of different alternatives weighted by their probabilities of being chosen. Defining an inclusive value from both stages of the GEV estimation gave us

$$I_p = \ln \left[\sum_{j \in J_S} e^{v_j} + \sum_{j \in J_F} e^{v_j} \right] \tag{22}$$

where J_S is the set of salt water sites, J_F is the set of fresh water sties and $v_j = \theta' Z_j + \phi' W_j$ where the Z's are explanatory variables in the first stage and the W's are explanatory variables in the second stage.

Inclusion of I_p in our macro allocation model is intuitively appealing but not perfectly correct. I_p, after all, is defined on choice occasions and the macro allocation decision is an annual or seasonal decision. In fact, as discussed in Chapter 9, there is no obvious way to make this model, or any of the related models, perfectly consistent between micro and macro decisions as well as economically plausible. However, since sufficient information is not available to determine how an individual's choice occasions might change over the season, there is only one such inclusive price for each individual in the data set. As such it may offer a good, albeit ad hoc, reflection of the value of the swimming alternatives available to the individual. It is, however, not consistent with a McFadden type utility theoretic model, and as such, its coefficient is not theoretically bounded by zero and one.

Results of the Estimation

1. Micro-Allocation Model

Table 10.1 presents the estimated coefficients and test statistics for the first stage of the GEV model and Table 10.2 presents the second stage results. Goodness of fit measures for logit models are not especially decisive. For each model we present Chi-square statistics based on likelihood ratio tests. In each case the statistic is significant at the 1% level of significance.

Another "goodness of fit" test is the "proportion predicted correctly". This is a misleading statistic because it counts an observation "correctly predicted" only if the alternative with the largest predicted probability is identical to the alternative chosen. This statistic gives no credit to being close as is implicit in the R^2 of multiple regression analysis. As such, it can make a good model look poor. In other cases, when the actual choices are skewed towards one alternative, it suggests more explanatory power than the model possesses. Finally when using the logit to predict shares, the statistic is not accommodating at all. Suppose individual i picks alternative j six times and alternative k four times in the sample. Then predicted probabilities of .6 and .4 will be quite accurate, yet the "proportion predicted correctly" statistic would count 6 correct predictions and 4 incorrect predictions. When the individual tends to select only one or two of the possible alternatives, this statistic is a better indicator. It is presented here for what it is worth. Note that the percent predicted correctly by the model is extremely high for the fresh water choice (72.5% for 8 freshwater sites) and the second stage choice (87.8% for the salt versus freshwater choice), but much poorer for the choice among the 22 saltwater beaches (26.3%). Although this certainly beats the probability of correct prediction from a model with no information (this probability would be 1/22 or 4.5%).

In the first stage of the GEV, the estimated coefficients all are significant at the 5% significance level and of the expected sign (with the possible exception of temperature and turbidity in the fresh water equation). The explanatory variables include five quality variables (except for temperature, each designates levels of undesirable chemicals or bacteria), three restricted variables (each defined so that when it takes the value 1 it reflects a nuisance of some sort), and travel cost. The results are encouraging because they bear out expectations. *Ceteris paribus* individuals visit closer beaches, avoid noisy areas and are discouraged by beaches heavily populated by ethnic groups different from their own. Additionally, individuals who do not own cars are less likely to visit beaches not serviced by public transportation.

TABLE 10.1:

First Stage GEV Model Estimates of Choice Among Freshwater and Saltwater Beaches

Boston - Cape Cod, 1975

Beach Characteristic	Saltwater Estimate (t-ratio)	Freshwater Estimate (t-ratio)
Oil	-.036 (-10.01)	-.100 (-2.62)
Fecal Coliform	-.049 (-4.12)	-.486 (-5.47)
Temperature	-.056 (-5.32)	-.281 (-3.58)
COD	-.022 (-17.67)	-.169 (-14.31)
Turbidity	-.047 (-8.48)	.273 (9.10)
Noise	-.109 (-9.90)	-.938 (-8.47)
Public Transportation	-1.103 (-12.91)	-1.275 (-4.07)
Beach Ethnicity	-1.784 (-27.58)	-1.321 (-5.51)
Trip Cost	-.572 (-35.89)	-2.166 (-26.61)
Llkelihood	-10850	896.
Precent right	26.3	72.5
Chi-squared with 9 degrees of freedom	4084.2	1804.7

TABLE 10.2

Second Stage GEV Model Estimates of Choice between Saltwater and Freshwater Beaches

Boston - Cape Cod, 1975

	Constant	Inclusive Price $(1-\sigma)$	No. of People in Household	% of Children in Household	Access to Swimming Pool
Estimated Coefficient (t-ratio)	16.520 (22.9)	.854 (23.6)**	-.162 (-10.9)	.420 (2.33)	.861 (9.16)

Likelihood = -1780.

Percent Right = 87.8

Chi-squared with 5 degrees of freedom = 3421.0

* t-ratios in parentheses

** This t-ratio tests significant difference from zero. A more appropriate test is significant difference from 1; the relevant t-ratio is -4.044.

These logical results lend credence to the water quality coefficient estimates, results about which we have less *a priori* information. The water quality variables were chosen to reflect easily perceived or highly publicized quality factors. Except for turbidity in the fresh water equation, the results seem to suggest individuals are responsive to these factors. In retrospect, the separation of the fresh and salt water decisions may have been appropriate because a certain level of turbidity, oil, etc. may be interpreted differently on a fresh water lake than at a salt water beach. (The levels of many of these variables tend to be quite different at salt versus fresh water sites).

While we may expect a different response at salt water beaches versus fresh water beaches, this can not directly be deduced from the coefficients reported in Table 10.1. Estimated coefficients of discrete choice models are notoriously difficult to interpret. If the model is

$$P_i = \frac{e^{\Sigma\theta_k Z_{ik}/(1-\sigma)}}{\sum_j e^{\Sigma\theta_k Z_{jk}/(1-\sigma)}} \tag{23}$$

then the interpretation of the estimated coefficient $\theta_k/(1-\sigma)$ is the following

(24) $$\frac{\theta_k}{(1-\sigma)} = \frac{\partial \ln(P_i/P_j)}{\partial(Z_{ik} - Z_{jk})} \qquad \text{for all } j.$$

Thus, if Z_{ik} were to increase, $\theta_k/(1-\sigma)$ would reflect how the log of the odds of choosing alternative i over each other alternative would change relative to the change in Z_{jk} (assuming Z_{jk} remains constant for all j). Consequently, it is inappropriate to compare coefficients across (saltwater and freshwater) equations, since the number of alternatives available will affect the magnitude of the coefficient. Nonetheless, these results establish the direction of effects as well as provide a means of predicting changes in behavior resulting from changes in values of explanatory variables.

From the first stage results can be calculated the "inclusive" values to be introduced in stage two. The expression for the inclusive values is presented in equation (11). For each individual, the inclusive value together with household size, percent children, and swimming pool access are used to explain the decision (on any one choice occasion) as to whether to visit a salt water or freshwater beach. The results are presented in Table 10.2 and can be interpreted as follows. Because of the way in which the constant term, household size, percent children, and swimming pool are entered into the estimation, their coefficients reflect the log of the odds ($\ln(P_S/P_F)$) of choosing a salt water site over a fresh water site. Thus larger families tend to go to lakes but families with a larger portion of children tend to go to salt water beaches. Those who have access to a swimming pool are more likely to visit salt water beaches than freshwater lakes.

The inclusive value term captures the effect of all of the variables used to explain site choice. By its design (see discussion of GEV model earlier), the coefficient of inclusive value must be positive and in the unit interval. This coefficient is an estimate of $1-\sigma$ from equation (11) where σ is an index of similarity (or correlation) among alternatives within groups not present across groups. For a $\sigma = 0$, no particular similarity exists within groups and there is no gain from the GEV model. For a $\sigma = 1$, alternatives within groups are perfect substitutes. In our problem, $1-\sigma$ equals .854 implying a σ of .146, which is significantly different from both 0 and 1. Thus we can expect that there are gains from using the GEV specification. Again, the results of stage two can be used to predict responses to changes in explanatory variables-- including those incorporated in stage 1, through the change in inclusive value.

2. The Macro-Allocation Model

The Tobit model (14) was estimated using the Berndt, Hall, Hall and Hausman maximum likelihood routine. Parameters were initialized at OLS estimates. The results of the simple OLS regression and the Tobit model are reported in Table 10.3.

The model presented in Table 10.3 includes income, household size, household composition, a restricted variable for ownership of specific water sports equipment (such as boat, water skis, wet suits, etc.), and the inclusive value variable (expression (22)) discussed earlier. The coefficients on other variables, such as education and length of work week, were not significantly different from zero by any reasonable test in the models employed, nor did their exclusion significantly alter the other results.

TABLE 10.3

Estimates of Tobit Model of Boston
Swimming Participation and Intensity

Variable	Tobit Estimates	Initial Value (OLS estimates)
Constant	26.01 (2.57)*	35.98 (4.59)
"Inclusive Value"	.897 (1.86)	1.02 (2.74)
Income	-1.19 (-.56)	-.07 (1.79)
Size of Household	-24.10 (-2.76)	-8.1 (-2.08)
Percent Children	-6.18 (-1.22)	-14.71 (-2.02)
Water Sports Equipment	13.05 (3.44)	6.42 (2.05)

Chi-squared statistic = 262.
*t-ratios in parentheses

The estimated coefficients for all variables except income are statistically significant from zero at the 2.5 percent level (for a one tail test). The coefficient on income is negative but not significantly different from zero even at the 20 percent significance level (in a two tail test), suggesting the absence of an income effect. It is, however, possible that the negative effect on participation of household size and number of children may in part be reflecting the fact that large families tend to be poorer, inner city families. However, re-specification of the model does not shed further light on this speculation. The "inclusive value" variable, included to reflect the value of recreational opportunities, is significant and positive as expected. This variable, derived as it is from quality and cost aspects of alternative sites, facilitates the prediction of changes in macro-allocation decisions arising from policy changes.

The statistic $(-2 \ln(L_0/L_\beta))$, where L_0/L_β is the likelihood ratio, is distributed as a chi-square. This statistic is significant at the 2.5 percent level for the Tobit model.

Benefit Measurement in the Context of the Multiple Site Model

The multi-layered model described and estimated above can be used in an interesting and revealing manner, both to predict behavioral responses to water quality changes and to value these changes in terms of compensating variation measures. It is possible, for example, to introduce a hypothetical change in one or more quality variables at some or all of the sites and then to predict changes in beach use in response to this quality change.

The model estimated in the previous sections allows us to capture three types of changes in beach use. The discrete-continuous macro-allocation model (estimated as a Tobit earlier in this chapter) permits the prediction of two aspects of the beach use decision: whether or not to participate and, if so, how many trips to take. Both aspects of the decision are functions of site qualities included in the participation function.

Finally, quality improvements, particularly if they have differential effects on sites, may cause individuals to reallocate trips among sites. The estimated parameters of the GEV models are combined with site qualities, individuals' costs and other variables to predict each household's probability of visiting each site. A predicted probability can be interpretted as a predicted share of the household's total trips. Thus a change in the quality at one or more sites can change a) whether or not a household participates in the recreational activity, b) the total number of trips taken, and c) the allocation of trips among sites.

The ultimate purpose of the modelling effort however is to estimate the benefits associated with improvements in water quality. Formulas for deriving welfare measures in the context of discrete choice models of random utility maximization have been developed by Hanemann (1982c, 1984c) and are outlined at the end of Chapter 9. It is generally the compensating and/or equivalent variation of the quality change which is taken as a useful measure of benefits. Selecting, for demonstration purposes, the compensating variation (C), this measure can be defined by the following expression:

$$v(p^o,b^o,y) = v(p^o,b^1,y-C)$$

where again v is the indirect utility function, p and b are vectors of site prices and qualities, and y is income.

In Chapter 9, the complications which arise in attempting to define this measure in a stochastic discrete choice setting were discussed. The compensating variation is now defined by

$$v(p^o,b^o,y;\varepsilon) = v(p^o,b^1,y-C;\varepsilon),$$

were ε is random and as a result C is now a random variable. Depending on how one chooses to take account of this randomness, three different measures of compensating variation can be defined. One is the C which equates the expected values of the indirect utility functions (C'), one is the mean of the distribution of C (C^+), and one is the median of that distribution (C^*). Hanemann argues that in many cases the distribution of C is likely to be substantially skewed resulting in the mean greatly exceeding the median. In the case of GEV models, however, the median value (C^*) coincides with C', the measure which equates expected utilities. It is this measure which we calculate in the subsequent illustration.

In our problem

(25) $$G(e^{v_1},\ldots e^{v_N}) = \left(\sum_{j\in J_S} e^{v_j/(1-\sigma)}\right)^{1-\sigma} + \left(\sum_{j\in J_F} e^{v_j/(1-\sigma)}\right)^{1-\sigma}$$

where J_S is the set of salt water sites and J_F the set of freshwater sites and where $v = \sigma' Z_{im} + \phi' W_m$, Z_{im} are factors which vary over sites and W_m are factors which vary between salt and freshwater sites. Thus

$$G(e^{v_1},\ldots,e^{v_N}) = \sum_{S=1}^{2} e^{\phi' W_s} + (1-\sigma) I_s$$

where $$I_s = \ln\left(\sum_{i\varepsilon s} e^{\sigma' Z_{is}/(1-\sigma)}\right).$$

Then the expected value of the indirect utility function equals

(26) $$v(p^o,b^o,y) = \ln G(e^{v_1^o},\ldots,e^{v_N^o}) + k,$$

where k is a constant.

Now consider a change in quality from b^o to b^1. The C' measure defined above is given by

(27) $$V(p^o,b^o,y) = V(p^o,b^1,y-C')$$

or

(28) $$\ln G\left(e^{v_1^o(y,z^o,w^o)},\ldots,e^{v_N^o(y,z^o,w^o)}\right)$$

$$\models \ln G\left(e^{v_1^1(y-C',z^1,w^1)},\ldots,e^{v_N^1(y-C',z^1,w^1)}\right).$$

There is no closed form solution for compensating variation in this case, but Hanemann (1982) shows that the compensating variation of a change from (p^o,b^o) to (p^o,b^1) can be approximated by

$$(29) \quad C \approx \left(\sum_s e^{\phi' w_s^0 + (1-\sigma) I_s^0} - \sum_s e^{\phi' w_s^1 + (1-\sigma) I_s^1}\right) / \sum_s \gamma_s e^{\phi' w_s^1 + (1-\sigma) I_s^1}$$

where $s = 1,2$ denotes the salt and fresh water alternatives (w_s^0, I_s^0) and (w_s^1, I_s^1) represent values of variables before and after the policy change respectively, and γ_1 and γ_2 are the implicit income coefficients in the salt and fresh water models.

The calculation of CV according to (29) yields an estimate of the compensating variation per choice occasion for the household. To obtain annual or seasonal benefit estimates, this number must be multiplied by the number of trips the individual takes. One should note that even if the individual takes no more trips in response to the quality change (either because he is constrained or because a more substantial quality change is necessary to increment the number of trips), the benefits of improvements are still measureable. That is, even if a quality change is insufficiently large to prompt an individual to alter his behavior in any way, the benefits he experiences if he is a user of the improved sites can be calculated.

In Table 10.4 the estimated benefits (in 1974 dollars) of a series of hypothetical water quality changes are reported. The estiamtes are only as good as the underlying data. While the Boston survey was carefully conceived, we have earlier noted a number of shortcomings in the data set. Thus, the results of this experiment are presented as an illustration with the hope that similar and more detailed experiments with the Chesapeake swimming data will generate more defensible benefit estimates.

The hypothetical water quality changes introduced include a 10% and a 30% reduction in each of the following water quality parameters individually: oil, chemical oxygen demand (COD) and fecal coliform. These reductions were introduced uniformly across all sites. In Table 10.4 the results are presented. A 10% reduction in COD, for example, is predicted to generate an average compensating variation of 12¢ (in 1984 dollars) per choice occasion and $2.85 (1974 dollars) per household per season when averaged over all households (users and nonusers) in the sample. The relevant sample included those 373 households with useable income information. Also reported in Table 10.4 is a 30% reduction in each pollutant at all sites.

Table 10.5 considers a 30% change in all four water quality parameters simultaneously. The benefits associated with this sort of change for all beaches is compared to the same sort of pollutant reductions if they affected only beaches in Boston harbor. Reductions in pollutants at downtown Boston beaches (8 of the 30 sites) generate more than half the benefits reported when all sites are uniformly improved.

TABLE 10.4

Average Compensating Variation Estimates of Reductions in Specific Pollutants at Boston Area Beaches (in 1974 dollars)

	10% reduction at all sites		30% reduction at all sites	
	per choice occasion	per season	per choice occasion	per season
oil	$ .05	$ .96	$ .20	$ 4.66
COD	.12	2.65	.29	7.15
fecal coliform	.02	.19	.12	2.85

TABLE 10.5
Average Compensating Variation Estimates of Water Quality Improvements for Boston City Beaches and All Boston Area Beaches

	30% reduction at all sites		30% reduction at downtown Boston Beaches	
	per choice occasion	per season	per choice occasion	per season
oil, turbidity, COD and fecal coliform	$ .50	$12.04	$ .27	$ 6.13

These examples are offered to demonstrate the sorts of questions which can be answered with a model such as the one estimated in this chapter. The model is admittedly a “second best” model. Approximations are adopted to make operational a theoretically consistent and behaviorally-plausible model. This approach requires statistical procedures which are easily obtainable but nonetheless the approach provides a complete description of the individual’s decision framework. The treatment of the micro- and macro-allocation decisions, while only an approximation to a completely consistent underlying utility framework, is a reasonable description of behavior.

CHAPTER 11

CONCLUSIONS

Theoretical acceptability and empirical tractability are the criteria most emphasized in this study. These are also the standards which may ultimately decide its worth. Without theoretical consistency, the work presented here will likely not stand the test of time. Without empirical tractability, the developments will remain largely academic, having little impact on benefit estimation.

The two standards set out here are often in conflict; theoretical descriptions of complex behavior may require complex empiricism. While most of the developments are quite operational, some of the results are not easily implemented at present. However, as our technical knowledge continues to advance, we may soon accomplish easily what now appears to be difficult. Moreover by developing more rigorous methods, the less demanding "short-cuts" can be tested to determine if they provide acceptable approximations.

Proper specifications and estimation would not be critical if benefit estimates were not sensitive to these considerations. Yet this and other research before it have revealed how sensitive benefit estimates are to specification and estimation approaches.

Whatever approach is taken to obtain welfare evaluations for policy makers, it must be undertaken with utmost care. Just as with contingent valuation studies, the worth of benefit measures generated by a recreational demand study is linked to the quality of the design of the study. In this Volume, a number of ways are offered in which anomalies in recreational demand analysis can be resolved and the quality of benefit estimates enhanced.

The Traditional Single Site/Activity Model

Part I of this Volume is devoted to the conventional recreational demand model for a single site or recreational activity. This is important because it is one of the most frequently used models and because often it is an essential component of more complex approaches such as that estimated in Chapter 10.

An analysis of existing literature suggests several weak linkages between the traditional "travel cost" model and economic theory. At the risk of making too universal a statement and thus appearing to ignore a few studies which have addressed these issues, the following weaknesses are apparent:

1) the lack of attention to individual behavior and utility maximization as foundations of welfare analysis;

2) the inability of behavioral models to allow for corner solutions with respect to participation, valuation of time and visitation of sites;

3) the neglect of theoretical and statistical properties of welfare estimates in the context of the travel cost model.

The belief that models of individual behavior and the tenets of welfare analysis are the foundations for good benefit analysis can not be overstated. Without credible models of behavior, we are left making leaps of faith in accepting benefit estimates. With a foundation of individual behavior, empirical tests of behavioral hypotheses can be achieved and, based on those results, inferences about welfare changes from water quality improvements are possible. Models of individual behavior and the theory of welfare provide the soundest structure for assessing benefits.

Within this context, several specific accomplishments of Part I are offered:

o Compilations of closed-form solutions for compensating and equivalent variation for some specific functional forms;

o Presentation of numerical algorithms for compensating and equivalent variation for flexible demand functions;

- A behavioral model which describes both the decision to participate in a recreation activity as well as how much to participate;

- An explanation of contradictions between the zonal travel cost model and a general model of individual behavior;

- An operational model of behavior in the presence of realistic constraints on income and leisure time, including the case in which the time constraint can not be collapsed into the budget constraint;

- Empirical evidence that traditional approaches to handling the value of time cause benefits to be understated

- Theoretical proof that different assumptions about the source of regression error lead to different benefit estimates;

- Theoretical proof that the small-sample property of unbiasedness is violated for most consumer surplus estimates - benefits computed from estimated linear and semi-log demand functions are biased upward;

- Procedures to correct the biasedness in benefit estimates based on small samples.

All of the shortcomings of the implicit market approach are far from resolved, but a large portion of the disbelief in the sound theoretical foundations of these procedures can be suspended. The theoretical foundation provides a structure in which hypotheses can be formed and tested.

Water Quality and the Multiple Site Model

The impetus for this study is a desire to obtain via indirect market methods accurate benefit measures of improvements in regional water quality. The above findings are all essential, but preliminary, steps to addressing this fundamental research goal. The measurement of water quality benefits is addressed in Part II, a section of the Volume largely devoted to multiple site modelling.

The connection between water quality valuation and multiple site modelling arises from both a practical and a substantive source. As early modellers discovered, the variation in quality at a single site is often insufficient to determine individuals' responsiveness to quality changes. Multiple site models offer the potential for observing large quality variation. Perhaps more important, however, is the realization that most environmental problems and surely water quality problems, affect regions - not single sites. Individuals choose to use sites based on the availability and quality of other alternatives. Approaches which hope to capture values through observations on behavior must accurately capture the choice environment of the individual.

Here again, review of the existing literature indicates several major problems in use of implicit market approaches. These include:

- a potential inconsistency between the quality variables to which recreators respond and the quality variables that policy-makers can control;
- a lack of general utility theoretic model of multiple site behavior against which the existing models can be assessed;
- a host of models which have, in one form or another, implausible assumptions about human behavior.

Part II is directed toward analyzing these problems.

The apparent conflict between the use of perceptions and objective measures of water quality may be less severe a problem than it appears. There is no doubt that a recreator's perception of water quality may not correspond directly to changes in quality parameters which are scientifically measureable. However, perceptions are not easily measured, and it is not clear that the objective measures are any less representative than poorly constructed subjective indices. Moreover, water quality parameters appear to be highly correlated so that the level of perceivable quality characteristic (e.g. turbidity) may be indicative of an unobservable pollutant. Finally, there is empirical evidence of relationships between objective measures and behavior.

Because both multiple site modelling and the modelling of quality and demand are relatively newer, less developed and more complex topics than single site modelling, our developments in Part II are of a different nature than those in Part I. In Part I incremental theoretical contributions are made to an already well-established estimation approach. There, the intent is to make benefit estimates more defensible, and in all cases the emphasis is on making developments operational. Because multiple site modelling is a newer and less well understood area, the contributions in Part II are of a more theoretical nature. While the practical issues of implementation have been directly addressed, the modelling framework is less well developed and the issues of concern more complex.

A few researchers have already ventured into the realm of multiple site modelling. Until now, the inherent differences in these approaches have not been analyzed. While different approaches produce different benefit measures, the models' characteristics have not been traced to differences in underlying assumptions. In Chapters 8 and 9, the nature of multiple site models is explored both from a statistical and an economic behavior viewpoint and bring to light the assumptions about behavior which are implicit in different modelling methods. Also presented is a general utility theoretic model, which is not yet operational.

One aspect of the nature of individuals' decisions turns out to have severe implications for modelling. Almost universally, individuals have access to a number of sites and visit more than one, but less than all, accessible sites. This "generalized corner solution" problem is not an academic construct but a reflection of the nature of people's behavior. Few researchers have recognized the importance of capturing this general corner solution-type behavior. Those who have, have attempted to do so in a purely statistically manner, by using statistical distributions with probability masses at zero. This is not satisfactory from an economist's point of view since the fundamental nature of demand is being ignored. Important changes in demand structures occur at corners, i.e. an individual's demand for site i no longer is sensitive to marginal price or quality changes for site j if j is not visited. This phenomena arises in a number of other economic settings and complicates demand analysis.

In Chapter 10, some of the most restrictive requirements of the multiple site model are relaxed and a plausible model of behavior which is relatively easy to estimate is presented. Estimation of the model using a Boston swimming data set provides an illustration of how the model can be used to predict changes in behavior. Benefit estimates are derived for a variety of hypothetical water quality improvements.

The Future

The emphasis on careful specification and theoretical consistency is really a disguised emphasis on behavior. The intent of this Volume's developments is to improve the modelling of the individual's decision process so that it can better approximate the structure of behavior. It is after all the basic tenet of indirect market methods that valuation can be revealed through behavior. To the extent that this is true, indirect market methods have a distinct advantage because the predictions of these models cart be tested against actual behavior. Thus we have a context in which our hypotheses can be tested.

The results to date are promising. The conceptual developments have been tested on a Boston data set which was collected without the additional conceptual guidance provided by the work in the Volume. Even with the limitations of this data, however, welfare gains from objective water quality measures can be observed. Subsequent analysis with the Boston data and with data on the Chesapeake Bay (collected in the course of this study) will, we hope, provide the empirical payoff from the development of our models of individuals' behavior.

RECREATIONAL DEMAND BIBLIOGRAPHY

Abbott, M., and Ashenfelter, O. 1976. "Labour Supply, Commodity Demand and the Allocation of Time". Review of Economic Studies 43:3:389-411.

Adams, R. L., R. C. Lewis and B. H. Drake. 1973. Outdoor Recreation, An Economic Analysis Bureau of Outdoor Recreation, U. S. Department of the Interior. Washington, D. C.

Aitcheson, J. and S. N. Chen. 1980. "Logistic - Normal Distributions: Some Properties and Uses". Biometrics 67:2:261-272.

Aitcheson, J. 1982. "The Statistical Analysis of Compositional Data". Journal of the Royal Statistical Society, Series B. 44:2:139-177.

Allen, P. Geoffrey, Thomas H. Stevens, and Scott A. Barrett. 1981. "The Effects of Variable Omission in the Travel Cost Technique." Land Economics 57:2(May):173-80.

Amemiya, T. 1973, "Regression Analysis When the Variance of the Dependent Variable is Proportional to the Square of Its Expectation", Journal of the American Statistical Association 68:928-934.

Anderson, F. J., and N. C. Bonser. 1974. "Allocations, Congestion and the Valuation of Recreational Resources." Land Economics 50:51-57.

Arrow, K. J., and A. C. Fisher. 1974. "Environmental Preservation, Uncertainty and Irreversibility." Quarterly Journal of Economics 88(May):313-19.

Ashenfelter, O., and J. Ham 1979. "Education, Unemployment and Earnings." Journal of Political Economy 89:S99-166.

Ashenfelter, O. 1980. "Discrete Choice in Labor Supply: The Determinants of Participation in the Seattle-Denver Income Maintenance Experiments". Industrial Relations Section, Princeton University. Working Paper No. 136.

Barker, M. L. 1971. "Beach Pollution in the Toronto Region." In Perceptions and Attitudes in Resource Management. W. Sewell and I. Burton (eds.). Ottawa. 37-47.

Barten, A. 1969. "Maximum Likelihood Estimation of a Complete System of Demand Equations." European Economic Review: 1:7-73.

Batie, S. S., R. B. Jensen, and L. G. Hogue. 1976. "A Lancasterian Approach for Specifying Derived Demands for Recreational Activities." Southern Journal of Agricultural Economics 8:101-7.

Beardsley, Wendell. 1971. "Bias and Non-Comparability in Recreation Evaluation Models." Land Economics 47:2(May):175-180.

Becker, G. S. 1965. "A Theory of the Allocation of Time." Economic Journal 75(Sept):493-517.

Becker, G. S., and H. G. Lewis. 1973. "On the Interaction Between the Quantity and Quality of Children". Journal of Political Economics 81:5279-5288.

Bell, Frederick, W. and E. Ray Canterbury. 1976. Benefits from Water Pollution Abatement. A report to the National Commission on Water Quality (Tallahassee, Florida State University). Available from NTIS as PB 252 172-01; 217 pps.

Binkley, Clark S., and W. Michael Hanemann. 1978. The Recreation Benefits of Water Quality Improvement: Analysis of Day Trips in an Urban Setting. Washington, D. C.: Environmental Protection Agency. EPA-600/5-78-010.

Bishop, John and Charles Cicchetti. 1975. "Some Institutional and Conceptual Thoughts on the Measurement of Indirect and Intangible Benefits and Costs." In Henry M. Peskin and Eugene P. Seskin, (eds.) Cost-Benefit Analysis and Water Pollution Policy Washington, D. C.: The Urban Institute.

Bishop, R. C., and T. A. Heberlein. 1979. "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" American Journal of Agricultural Economics 6:926-30.

Bishop, R. C., and T. A. Heberlein. 1980. "Simulated Markets, Hypothetical Markets and Travel Cost Analysis: Alternative Methods of Estimating Outdoor Recreation Demand." Working Paper #187, 45 pps.

Bishop, R. C., and K. Samples. 1985. "Estimating the Value of Variations in Anglers' Success Rates: An Application of the Multiple-Site Travel Cost Method." Marine Resource Economics 2:1:55-74.

Bockstael, N. E., and K. E. McConnell. 1980. "Calculating Equivalent and Compensating Variation for Natural Resource Environments." Land Economics 56:l(Feb):56-63.

Bockstael, N. E., and K. E. McConnell. 1981. "Theory and Estimation of the Household Production Function for Wildlife Recreation." Journal of Environmental Economics and Management 8:199-214.

Bockstael, N. E., and K. E. McConnell. 1983. "Welfare Measurement in the Household Production Framework." American Economic Review 73:806-814.

Bockstael, N. E., and I. E. Strand. 1985. "Distribution Issues and Non-market Benefit Measurement." Western Journal of Agricultural Economics forthcoming.

Bockstael, N. E., and I. E. Strand. 1986. "The Effect of Common Sources of Regression Error on Benefit Estimates". University of Maryland. Manuscript.

Bockstael, N. E., I. E. Strand, and W. M. Hanemann. 1985. "Time and the Recreational Demand Model." Department of Agricultural and Resource Economics, University of Maryland, College Park. Manuscript.

Bouwes, N. W. 1983. An Economic Analysis of Lake Rehabilitation/ Stabilization Projects: Mirron/Shadow Lakes and White Clay Lake. ERS Staff Report No. AGES830620.

Bouwes, N. W., and R. Schneider. 1979. "Procedures in Estimating Benefits of Water Quality Change." American Journal of Agricultural Economics 61:535-539.

Bowes, M. David, and J. B. Loomis. 1980. "A Note on the Use of Travel Cost Models with Unequal Zonal Populations." Land Economics 56:4:465-470.

Brown, G., J. Charbonneau, and M. J. Hay. 1978. "The Value of Wildlife Estimated by the Hedonic Approach." U.S. Fish and Wildlife Service Working Paper #6, 25 pps.

Brown, G. M. and R. Mendelsohn. 1984. "The Hedonic Travel Cost Method." Review of Economics and Statistics 66:427-33.

Brown, R. J., J. Durbin and J. M. Evans. 1975. "Techniques for Testing the Constancy of Regression Relationships Over Time." Journal of the Royal Statistical Society 37(2):149-63.

Brown, M. G., and F. Nawas. 1973. "Impact of Aggregation on the Estimation of Outdoor Recreation Demand Functions." American Journal of Agricultural Economics 55:246-249.

Brown, W. G., Ajner Singh and Emery N. Castle. 1965. "Net Economic Value" of the Oregon Salmon-Steelhead Sport Fishery." Journal of Wildlife Management 29:266.

Brown, W. G., C. Sorhus, B. Chou-Yang, S. Richards. 1983. "Using Individual Observations to Estimate Recreation Demand Functions." American Journal of Agricultural Economics 65:154-157.

Brown, W. G., C. N. Sorhus, and D. A. Meyer. 1982. "Estimation Of Oregon Sport-Caught Salmon and Steelhead Values Via the Travel Cost Method." Memeo, ASSA Meetings, 25 pps.

Bruvold, W. H. 1975. "Human Perception and Evaluation of Water Quality." CRC Critical Reviews in Environmental Control. 5:153-231.

Burt, O. R., and D. Brewer. 1971. "Estimation of Net Social Benefits from Outdoor Recreation." Econometrics 39:5:813-827.

Burtless, G., and J. A. Hausman 1978. "The Effect of Taxation on Labor Supply: Evaluating the Gary Indiana Negative Income Tax Experiment." Journal of Political Economy 86:1103-30.

Cabelli, V. J., A. P. Dufour, L. J. McCabe, and M. A. Levin,. 1983. "A Marine Recreational Water Quality Criterion Consistent with Indicator Concepts and Risk Analysis." Journal of the Water Pollution Control Federation 55:1306-1314.

Carson, R. T., and R. C. Mitchell. 1983. "A Re-estimation of Bishop and Heberlein's Simulated Market - Hypothetical Markets - Travel Cost Under Alternative Assumptions." Resources for the Future, Manuscript.

Caulkins, P. P. 1982. An Empirical Study of the Recreational Benefits Generated by Water Quality Improvement. Unpublished doctoral dissertation. University of Wisconsin-Madison. 235 pps.

Caulkins, P. P., R. C. Bishop, and N. W. Bouwes. 1982. "The Travel Cost Model for Lake Recreation: A Comparison of Two Methods for Incorporating Site Quality and Substitution Effects." Mimeo, AEA Annual Meeting, 25 pps.

Caulkins, P. P., R. C. Bishop and N. W. Bouwes, Sr. 1984. "The Travel Cost Model for Lake Recreation: A Comparison of Two Methods of Incorporating Site Quality and Substitution Effects." Environmental Protection Agency.

Cesario, F. J. 1973. "A Generalized Trip Distribution Model." Journal of Regional Science 13:233-247.

Cesario, F. J. 1976. "Value of Time in Recreation Benefit Studies." Land Economics 52(Feb):32-41.

Cesario, F. J. 1980. "Congestion and Valuation of Recreation Benefits." Land Economics 56:329-38.

Cesario, F. J., and Jack L. Knetsch. 1970. "Time Bias in Recreation Benefit Estimates." Water Resources Research 6(June):700-704.

Cesario, F. J., and Jack L. Knetsch. 1976. "A Recreation Site Demand and Benefit Estimation Model." Regional Studies 10:97-104.

Cheung, H. K. 1972. "A Day-Use Park Visitation Model." Journal of Leisure Research 4:139-156.

Christensen, J. B., and C. Price. 1982. "A Note on the Use of Travel-Cost Models with Unequal Zonal Populations: Comment." Land Economics 58:395-99.

Cicchetti, C. J., A. C. Fisher, and V. K. Smith. 1973. "Economic Models and Planning Outdoor Recreation." Operations Research 21:5(Sept-Oct):1104-1113.

Cicchetti, C. J., A. C. Fisher, and V. K. Smith. 1976. "An Econometric Evaluation of a Generalized Consumer Surplus Measure: The Mineral King Controversy." Econometrics 44:6:1259-1276.

Cicchetti, C. J., and A. Myrick Freeman, III. 1971. "Option Demand and the Consumer Surplus: Further Comment." Quarterly Journal of Economics 85: (Aug):528-39.

Cicchetti, C. J., Joseph J. Seneca, and Paul Davidson. 1969. The Demand and Supply of Outdoor Recreation. New Brunswick, New Jersey: Bureau of Economic Research, Rutgers University.

Cicchetti, C. J., and V. K. Smith. 1973. "Congestion, Quality Deterioration and Optimal Use: Wilderness Recreation in the Spanish Peaks Primitive Area." Social Science Research 2:15-30.

Cicchetti, C. J., and V. K. Smith. 1976. "The Measurement of Individual Congestion Costs: An Economic Application to Wilderness Recreation." In Theory & Measurement of Economic Externalities. Lin and A. Y. Steven (eds.), Academic Press: N. Y. 183-203.

Cicchetti, C. J., and V. K. Smith. 1976. The Costs of Congestion. Cambridge, Massachusetts: Ballinger Publishing Co.

Cicchetti, C. J., V. K. Smith, J. L. Knetsch, and R. A. Patten. 1972. "Recreation Benefit Estimation and Forecasting: Implications of the Identification Problem." Water Resources Research 8:4(Aug):840-850.

Clawson, M. 1959. Methods of Measuring the Demand for and Value of Outdoor Recreation. Washington, D. C. Resources for the Future Reprint 10.

Clawson, M., and J. L. Knetsch. 1966. Economics of Outdoor Recreation. Washington, D. C.: Resources for the Future, Inc.

Common, M. S. 1973. "A Note on the Use of the Clawson Method for the Evaluation of Recreation Site Benefits." Regional Studies 7:401-406.

Cook, Philip J. 1972. "A 'One-Line' Proof of the Slutsky Equation." American Economic Review 62:139.

Cummings, R. G., D. S. Brookshire, and W. D. Schulze. 1984. Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Valuation Method. Draft report presented to the Environmental Protection Agency.

Darling, A. H. 1973. "Measuring Benefits Generated by Urban Water Parks." Land Economics 49:22-24.

David, Elizabeth L. 1971, "Public Perceptions of Water Quality." Water Resources Research 7:3(June):453-457.

Davidson, P., F. G. Adams, and J. Seneca. 1966. "The Social Value of Water Recreational Facilities Resulting from an Improvement in Water Quality: The Delaware Estuary." In A. V. Kneese and S. C. Smith (eds.), Water Research Baltimore: John Hopkins Press.

Davis, Robert K. 1963. "Recreation Planning as an Economic Problem." Natural Resources Journal 3:2(Oct):239-49.

DeSerpa, A. C. 1971. "A Theory of the Economics of Time." Economic Journal 81:828-846.

Desvousges, W. H., V. K. Smith and M. P. McGivney. 1983. "A Comparison of Alternative Approaches for Estimation of Recreation and Related Benefits of Water Quality Improvements." Washington, D. C.: Environmental Protection Agency, Office of Policy Analysis.

Deyak, Timothy A., and V. Kerry Smith. 1978. "Congestion and Participation in Outdoor Recreation: A Household Production Approach." Journal of Environmental Economics and Management 5(Mar) 63-80.

Dinius, S. H. 1981. "Public Perceptions in Water Quality Evaluation." Water Resources Bulletin 17:116-121.

Ditton, B., and T. L. Goodale. 1973. "Water Quality Perceptions and the Recreational Users of Green Bay." Water Resources Research 9:3:569-79.

Dornbusch, D. M. 1975. The Impact of Water Quality Improvements on Residential Property Prices. Prepared for the National Commission on Water Quality. NTIS, pbs 248-805.

Domencich, T. A. and D. McFadden, 1975. Urban Travel Demand: A Behavioral Analysis". North-Holland, Amsterdam.

Driver, B. L., and Perry J. Brown. 1975. "A Social-Psychological Definition of Recreation Demand, with Implications for Recreation Resource Planning." In Assessing the Demand for Outdoor Recreation. Washington, D. C.: National Academy of Sciences.

Dworkin, J. M. and K. D. Pijawka. 1982. "Public Concern for Air Quality: Explaining Change in Toronto, Canada, 1967-1978." International Journal of Environmental Studies 10:728-744.

Dwyer, J. F., and Michael D. Bowes. 1978. "Concepts of Value for Marine Recreational Fishing." American Journal of Agricultural Economics 60:1008-12.

Dwyer, J. F., J. R. Kelly, and M. D. Bowes. 1977. Improved Procedures for Valuation of the Contribution of Recreation to National Economic Development. Urbana-Champaign: University of Illinois.

Edwards, J. A., K. C. Gibbs, L. J. Guedry, and H. H. Stoevner. 1976. The Demand for Non-Unique Outdoor Recreational Services: Methodological Issues. Oregon State University Agricultural Experiment Station Technical Bulletin No. 133 (May).

Environmental Research Group. 1972. A Guide to the Data Collecting and Processing System Economic Evaluation of Wildlife and Wildlife Oriented Resources in the Southeastern U. S. Southern Georgia State University, Atlanta.

Federal Water Pollution Control. 1966. Delaware Estuary Comprehensive Study: Preliminary Report and Findings. Philadelphia; Department of Interior.

Feenberg, Daniel, and Edwin S. Mills, 1980. Measuring the Benefits of Water Pollution Abatement. New York: Academic Press.

Fisher, Ann. 1983. "The Role of Risk Assessment in Evaluating the Benefits of Hazardous Waste Management." Mimeo, Southern Economics Meetings, 29 pps.

Fisher, A. C., and J. V. Krutilla. 1972. "Determination of Optimal Capacity of Resource Based Recreation Facilities." In J. V. Krutilla (ed.) Natural Environments: Studies in Theoretical and Applied Analysis. Balt.: Johns Hopkins Press.

Fisher, Ann, and Robert Raucher. 1982. "Comparison of Alternative Methods of Evaluating the Intrinsic Benefits of Improved Water Quality." Paper presented at the American Economics Association Annual Meeting. New York, New York.

Fisher, Franklin and Carl Shell. 1971. "Taste and Quality Change in the Pure Theory of the True Cost of Living Index." In Griliches (ed), Price Indexes and Quality Change. Harvard University Press.

Freeman, A. Myrick, III. 1979a. The Benefits of Environmental Improvement: Theory and Practice. Baltimore: Johns Hopkins Press for Resources for the Future, Inc.

Freeman, A. Myrick, III. 1979b. The Benefits of Air and Water Pollution Control: A Review and Synthesis of Recent Estimates. Prepared for Council on Environmental Quality, Washington, D. C.

Freund, R. J. and R. R. Wilson. 1974. "An Example of a Gravity Model to Estimate Recreation Travel." Journal of Leisure Research 6:241-256.

Gibbs, Kenneth C. and John F. McGuire III. 1973. Estimation of Outdoor Recreation Values. Gainesville, Florida: University of Florida, Food and Resource Economics Dept.

Gillespie, G. A. and D. Brewer. 1968. "Effects of Nonprice Variables Upon Participation in Water-Oriented Outdoor Recreation." American Journal of Agricultural Economics 50:82-90.

Gillespie, G. A. and D. Brewer. 1969. An Econometric Model for Predicting Winter-Oriented Outdoor Recreation Demand. USDA Bulletin No. ERS-402. Washington, D. C.: U.S. Government Printing Office (March).

German, W. M. 1976. "Tricks with Utility Functions." In Essays in Economic Analysis. M. Artis and R. Nobay (eds.), Cambridge University Press, Cambridge.

Grieg, P. J. 1983. "Recreation Evaluation Using a Characteristics Theory of Consumer Behavior." American Journal of Agricultural Economics 65(Feb.):90-97.

Grubb, Herbert W., and James T. Goodwin. 1968. Economic Evaluation of Water Oriented Recreation in the Preliminary Texas Water Plan. Austin, Texas Water Develop. Board Ref. No. 84.

Gum, R. L., and W. E. Martin. 1975. "Problems and Solutions in Estimating the Demand for and Value of Rural Outdoor Recreation." American Journal of Agricultural Economics 57:558-566.

Ham, J. C. 1982. "Estimation of a Labour Supply Model with Censoring Due to Unemployment and Underemployment." Review of Economic Studies 49:333-54.

Hanemann, W. Michael 1978. "A Methodological and Empirical Study of the Recreation Benefits from Water Quality Improvement," Ph.D. dissertation, Department of Economics, Harvard University.

Hanemann, W. Michael. 1979. "Consumer's Surplus Without Apology: Another Comment." University of California, Berkeley.

Hanemann, W. Michael 1980a. "Quality Changes, Consumer's Surplus, and Hedonic Price Indices." University of California, Berkeley, CA. Economics Working Paper No. 116, (November).

Hanemann, W. Michael. 1980b. "Measuring the Worth of Natural Resource Facilities: Comment." Land Economics 56(Nov)4:482-486.

Hanemann, W. Michael 1982a. "Quality and Demand Analysis". In New Directions in Econometric Modelling and Forecasting in U.S. Agriculture, Gordon C. Rausser (ed.), Elsevier/North-Holland Publishing Company, New York.

Hanemann, W. Michael 1982b. "Welfare Analysis Based on Systems of Partial Demand Functions." University of California, Berkeley: California Agricultural Experiment Station Working Paper No. 243.

Hanemann, W. Michael 1982c. “Applied Welfare Analysis w/Qualitative Response Models.” University of California, Berkeley. California Agricultural Experiment Station Working Paper No. 241.

Hanemann, W. Michael 1982d. "Some Further Results on Exact Consumer's Surplus". University of California, Berkeley, Working paper, 9 pps.

Hanemann, W. Michael 1983a. “Econometric Models of the Demand for Quality-Differentiated Goods.” University of California, Berkeley: Draft, 26 pps.

Hanemann, W. Michael 1983b. “Marginal Welfare Measures for Discrete Choice Models.” Economic Letters 13:129-136.

Hanemann, W. Michael 1983c. “Information and the Concept of Option Value.: University of California, Berkeley: California Agricultural Experiment Station Working Paper No. 228.

Hanemann, W. Michael. 1984a. “Discrete-Continuous Models of Consumer Demand.” Econometrics 52(May):541-561.

Hanemann, W. Michael. 1984b. “Entropy as a Measure of Consensus in the Evaluation of Recreational Site Quality”. Journal of Environmental Management 18:3:241-252.

Hanemann, W. Michael. 1984c. “Welfare Analysis with Discrete Choice Models.” University of California, Berkeley, Working Paper #173.

Hanemann, W. Michael, 1984d. “Corner Solutions and Duality in Consumer Theory.” University of California, Berkeley, Manuscript.

Hanemann, W. Michael. 1984e. “Welfare Evaluation in Contingent Valuation Experiments with Discrete Responses.” American Journal of Agricultural Economics 66:332-341.

Haveman, Robert H. and Burton A. Weisbrod. 1975. "The Concept of Benefits in Cost-Benefit Analysis: With Emphasis on Water Pollution Activities." In Henry M. Peskin and Eugene P. Seskin (eds.), Cost-Benefit Analysis and Water Pollution Policy. Washington, D. C. The Urban Institute.

Barrington, W. 1981. "The Distribution of Recreation Benefits from Improved Water Quality: A Microsimulation." Unpublished report to the National Science Foundation, Resources for the Future.

Haspel, Abraham, E., and F. Reed Johnson. 1982. "Multiple Destination Trip Bias in Recreation Benefit Estimation." Land Economics 58(Aug):364-72.

Hausman, J. A. 1981. "Exact Consumer's Surplus and Deadweight Loss." American Economic Review 71:4:662-676.

Hausman, J. A. and D. A. Wise. 1978. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." Econometrics 46:403-426.

Hay, Michael J., and Kenneth E. McConnell. 1979. "An Analysis of Participation in Nonconsumptive Wildlife Recreation." Land Economics 55:11:460-471.

Heckman, J. J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables." Annals of Social and Economic Measurement 5:475-492.

Heckman, J. J., M. R. Killingsworth, and T. E. McCurdy. 1981. "Empirical Evidence on Static Labour Supply Models: A Survey of Recent Developments," In Hornstein, Grice, and Webb. (eds.) 75-122.

Heintz, H. T. Jr., A. Hershaft and G. C. Horak. 1976. "National Damages of Air and Water Pollution." A Report to the U.S. Environmental Protection Agency. Rockville, Md., Environmental Control, Inc.

Hendry, G. S. and A. Toth. 1982. "Some Effects of Land Use on Bacteriological Water Quality in a Recreational Lake." Water Research 16:105-112.

Hiett, R. L. and J. Worrall. 1977. "Marine Recreational Fishermen's Ability to Estimate Catch and to Recall Catch and Effort over Time." Human Sciences Incorporated, McLean, Va.

Hof, J. G., and D. A. King. 1982. "On the Necessity of Simultaneous Recreation Demand Equation Estimation." Land Economics 58:547-52.

Hof, J. G., and D. A. King. 1983. "On the Necessity of Simultaneous Recreation Demand Equation Estimation: Reply." Land Economics 59:4:459.

Hof, J. G., and J. B. Loomis. 1983. "A Recreation Optimization Model Based on the Travel Cost Method." Western Journal of Agricultural Economics 8:1:76-85.

HoIt, Suzanne 1983. "Estimating the Value of Travel Time." Santa Cruz: Univ. of California. Manuscript, 9 pps.

HoIt, Suzanne. 1984. "The Shadow Price of Time Revealed." University of California, Santa Cruz, Manuscript 25 pp.

Hotelling, H. 1938. "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates." Econometrics 6:242-69.

Hotelling, H. 1948. Letter to the National Park Services in "Economics of Outdoor Recreation" - The Prewitt Report.

Houthakker, H. 1951-1952. "Compensated Changes in Quantities and Qualities Consumed." Review of Economic Studies 19:155-164.

Hueth, D. 1984. "The Measurement of Recreation Quality Change in a Household Production Framework: Theory and Application." Paper presented to the 1984 Annual Meeting, Cornell, N. Y. August 5-8, 1984.

Huppert, D. D. 1983. "NMFS Guidelines on Economic Valuation of Marine Recreational Fisheries." Washington, D. C. NOAA Tech. Memo. NMFS, SWF6-32.

Huppert, D. D. and Cynthia J. Thomson. 1984. "Demand Analysis of Partyboat Angling in California Using the Travel Cost Method." Southwest Fisheries Center Administrative Report, NOAA (February).

Hurwicz, Land H. Uzawa. 1971. "On the Integrability of Demand Functions." In Preferences, Utility and Demand. Chipman, Hurwicz, Richter and Sonnenschein (eds). New York: Harcourt Brace.

Johnson, B. M. 1966. "Travel Time and the Price of Leisure." Western Economic Journal 4:135-145.

Johnson, G. 1983. "Measuring the Cost of Time in Recreation Demand Analysis: Comment." American Journal of Agricultural Economics 65:169-71.

Just, Richard, Darrell Hueth and Andrew Schnity. 1982. Applied Welfare Economics and Public Policy. Prentice Hall, Englewood Cliffs, N. J.

Kalter, R. J. 1971, The Economics of Water Based Outdoor Recreation. Alexandria, Va.: Institute for Water Resources, Army Corps of Engineers.

Kalter, R. J. and L. Gosse. 1970. "Recreation Demand Functions and the Identification Problem." Journal of Leisure Research 2:43-53.

Katzner, D. 1970. Static Demand Theory. MacMillan, New York.

Killingsworth, Mark R. 1983. Labor Supply. Cambridge Surveys of Economic Literature, Cambridge University Press.

Knetsch, J. L. 1963. "Outdoor Recreation Demands and Benefits." Land Economics 39:387-96.

Knetsch, J. L. 1964. "Economics of Including Recreation as a Purpose of Eastern Water Projects." Journal of Farm Economics 46:5(Dec):l148-1157.

Knetsch, J. L. 1977 . "Displaced Facilities and Benefit Calculations." Land Economics 43:123-29.

Knetsch, J. L., Richard E. Brown and William J. Hansen. 1976. "Estimating Expected Use and Value of Recreation Sites." Planning for Tourism Development: Quantitative Approaches. C. Gearing, W. Swart, and T. Var (eds.). New York: Praeger Publishers.

Knetsch, J. L., and Robert K. Davis. 1966. "Comparison of Methods for Recreation Evaluation." In A. V. Kneese and S. C. Smith (eds.), Water Research. Baltimore: Johns Hopkins.

Kooyoomjian, K. J. and N. L. Clesceri. 1974. "Perceptions of Water Quality by Select Respondent Groupings in Inland Water-Based Recreational Environments." Water Resources Bulletin 10:728-744.

Krutilla, J. V. 1967. "Conservation Reconsidered." American Economic Review 57(Sept):777-86.

Krutilla, J. V., and Cicchetti, C. J. 1972. "Evaluating Benefits of Environmental Resources with Special Application to the Hells Canyon." Natural Resources Journal 12 (Jan):l-29.

Krutilla, J. V., and A. C. Fisher. 1975. The Economics of Natural Environments. Baltimore: John Hopkins Press for Resources for the Future, Inc.

Krutilla, J. V., and Jack L. Knetsch. 1970. "Outdoor Recreation Economics." Annals of the American Academy of Political and Social Sciences 389:63-70.

Kuhn, Thomas S. 1969. Structure of Scientific Revolutions. University of Chicago Press.

LaFrance, Jeffrey and W. M. Hanemann. 1985. "The Integration of Some Common Demand Systems." Economics Dept., Bozeman, Montana, Manuscript.

Lancaster, K. 1966. "A New Approach to Consumer Theory." Journal of Political Economy 74:2:132-157.

Lancaster, K. J. 1971. Consumer Demand: A New Approach. Columbia University Press, New York.

Lee, Lung-Fei and Mark M. Pitt. 1983. "Specification and Estimation of Demand Systems with Limited Dependent Variable." Discussion Paper #189, Center for Economic Research, University of Minnesota.

Levin, Henry. M. 1966. "Estimating the Municipal Demand for Public Recreational Land." Washington, D. C.: Economic Studies Division, The Brookings Institute.

Long, W. H. 1968. "A Sample Design for Investigating the Effects of Stream Pollution on Water-Based Recreational Expenditures." Water Resources Bulletin 4:3(Sep):19-26.

Longford, W. A., and D. J. Cocheba. 1978. "The Wildlife Valuation Problem: A Critical Review of Economic Approaches." Canadian Wildlife Service Occ. Paper #37, 37 pps.

Lucas, Robert C. 1963. "Bias in Estimating Recreationists' Length of Stay from Sample Interviews." Journal of Forestry 61(Dec):912-14.

Maddala, F. S. 1983. Limited-dependent and Qualitative Variables in Econometrics. New York. Cambridge University Press.

Majid, I., J. A. Sinden, and A. Randall. 1983. "Benefit Evaluation of Increments to Existing Systems of Public Facilities." Land Economics 59:4:377-392.

Maler, Karl G. 1971. "A Method of Estimating Social Benefits from Pollution Control." Swedish Journal of Economics 73:1:121-133.

Maler, Karl G. 1974. Environmental Economics: A Theoretical Inquiry. Baltimore: John Hopkins Press for Resources for the Future, Inc.

Maler, Karl G. and Ronald Wyzga. 1976. Economic Measurement of Environmental Damage. Paris, Organization for Economic Co-operation and Development.

Malm, W., K. Kelley, J. Molenar, and T. Daniel. 1981. "Human Perception of Visual Air Quality (Uniform Haze)." Atmospheric Environment 15:1875-1900.

McConnell, K. E. 1975. "Some Problems in Estimating the Demand for Outdoor Recreation." American Journal of Agricultural Economics 57(May):330-39.

McConnell, K. E. 1977. "Congestion and Willingness to Pay: A Study of Beach Use." Land Economics 53:2:185-95.

McConnell, K. E. 1979. "Values of Marine Recreational Fishing: Measurement and Impact of Measurement." American Journal of Agricultural Economics 61:5:921-923.

McConnell, K. E. 1980. "Valuing Congested Recreation Sites." Journal of Environmental Economics and Management 7(Dec):389-394.

McConnell, K. E. 1980. "Valuing Congested Recreation Sites Comment." Journal of Environmental Economics and Management 7(Dec):389-394.

McConnell, K. E. (ed.) 1984. Identification of Preferences in Hedonic Models. Draft report presented to the Environmental Protection Agency under Cooperative Agreement CR-811043-O1-O, Vol. I.

McConnell, K. E., and V. A. Duff. 1976. "Estimation of Net Benefits of Recreation under Conditions of Excess Demand." Journal of Environmental Economics and Management 2:224-230.

McConnell, K. E., and I. E. Strand. 1981. "Measuring the Cost of Time in Recreation Demand Analysis: An Application to Sport Fishing." American Journal of Agricultural Economics 63(Feb):153-56.

McConnell, K. E. and I. E. Strand 1983. "Measuring the Cost of Time in Recreation Demand Analysis, Reply." American Journal of Agricultural Economics 65:169-71.

McConnell, K. E., and J. G. Sutinen. 1983. "A Conceptual Analysis of Congested Recreation Sites." In Advances in Applied Micro-economics. V. K. Smith (ed), Vol 3.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." In Frontiers in Econometrics. P. Zarembka (ed.), Academic Press, New York.

McFadden, D. 1976. "Quantal choice analysis: A survey." Annals of Economic and Social Measurement 5:363-390.

McFadden D. 1978. "Modelling the Choice of Residential Location". In Spatial Interaction Theory and Planning Models. Anders Karlquist, Lars Lundqvist, Folke Snickars, and Jorgen W. Weibull (eds.), North-Holland Publishing Company, New York.

McFadden, D. 1981. "Econometric Models of Probabilistic Choice," In Structural Analysis of Discrete Data. C. F. Manski and D. McFadden (eds.), MIT Press, Cambridge, Mass.

McGaw, R. L., and D. W. Gillen. 1983. "Economic Value of Salmon Angling: Estimates of Willingness to Pay from Hedonic Price Functions." Fredericton: University of New Brunswick, Department of Economics Working Paper, 18 pps.

Mendelsohn, Robert. 1983. "Identifying Structural Equations with Single Market Data." Institute for Economic Research. University of Washington, Seattle.

Mendelsohn, Robert. 1984. "Estimating the Structural Equations of Implicit Markets Household Production Function," Review of Economic and Statistics 66:673-677.

Menz, F. C. 1981. "An Economic Evaluation of the St. Lawrence River-Eastern Lake Ontario Bass Fishery." Report prepared for New York Sea Grant Institute (June).

Menz, F. C. and D. P. Wilton. 1983. "Alternative Ways to Measure Recreation Value by the Travel Cost Method." American Journal of Agricultural Economics 65:332-336.

Merewitz, Leonard. 1966. "Recreational Benefits of Water Resource Development." Water Resources Research 2:625-40.

Miller, Jon R., and Frederic C. Menz. 1979. "Some Economic Considerations for Wildlife Preservation." Southern Economic Journal 45(Jan):718-29.

Mood, A., F. Graybill and D. Bees. 1963. Introduction to the Theory of Statistics. McGraw-Hill Book Co., New York.

Morey, E. R. 1981. "The Demand for Site-Specific Recreational Activities: A Characteristics Approach." Journal of Environmental Economics and Management 8:345-371.

Morey, E. R. 1984a. "The Choice of Ski Areas: Estimation of a Generalized CES Preference Ordering with Characteristics, Quadratic Expenditure Functions and Non-additivity." Review of Economics and Statistics 66:584-90.

Morey, E. R. 1984b. "Characteristics, Consumer Surplus and New Activities: A Proposed Ski Area." Paper presented at the American Economics Association Meetings, Dallas, Texas.

Morey E. R. and Robert Rowe. 1985. "The Logit Model and Exact Consumer's Surplus Measures: Valuing Marine Recreational Fishing." Paper presented at AERE Workshop in Recreational Demand Modeling.

Mossin, J., and M. Bronfenbrenner. 1967. "The Shorter Work Week and Labor Supply." Southern Economic Journal 33:322-31.

Muellbauer, J. 1974. "Household Production Theory, Quality, and the 'Hedonic Technique'." American Economic Review 64:977-994.

Munley, Vincent G., and V. Kerry Smith. 1976. "Learning by Doing and Experience.: The Case of Whitewater Recreation." Land Economics 52:4:545-53.

National Commission on Water Quality. 1975. Washington, D.C. Staff Report.

National Economic Research Associates (NERA). 1974. "Estimates of the Value of Changes in Aquatic Life Resulting from Reductions in Thermal Effluent." Attachment V to Edison Electric Institute Ad Hoc Water Quality Group and Utility Water Act Group, Comments on EPA's Proposed 304 Guidelines and 306 Standards of Performance for Stem Electric Power Plants. Report submitted to U. S. EPA, Washington, D. C.

National Park Service, Land and Recreation Planning Division. 1949. The Economics of Public Recreation. Washington, D. C.

National Academy of Sciences. 1975. Assessing Demand for Outdoor Recreation. Washington, D. C.

Nichols, L. M., M. Bowes, and J. F. Dwyer. 1978. Reflecting Travel Time in Travel-Cost-Based Estimates of Recreation Use and Value. Forestry Research Report #78-12. University of Illinois, Urbana-Champaign.

Norton, V. J., T. P. Smith, and I. E. Strand. 1983. Stripers: The Economic Value of the Atlantic Coast Commercial and Recreational Striped Bass Fisheries. Maryland Sea Grant Publication # UM-SG-TS-83-12.

Perlman, R. i966. "Observations on Overtime and Moonlighting." Southern Economic Journal 33: 237-44.

Peterson, G. L. and E. S. Neumann. 1969. "Modeling and Predicting Human Response to the Visual Recreational Environment." Journal of Leisure Research 1:219-237.

Pollak, R. A. 1971. "Conditional Demand Functions and the Implications of Separable Utility." Southern Economic Journal 37:423-33.

Pollak, R. A., and M. L. Wachter. 1975. "The Relevance of the Household Production Function and Its Implications for Allocation of Time." Journal of Political Economics 83:2:255-277.

Pollak, R. A. and T. J. Wales. 1969. "Estimation of the Linear Expenditure System." Econometrica 37:611-628.

Pollack, R. A., and T. J. Wales. 1978. "Estimation of Complete Demand Systems from Budget Data: The Linear and Quadratic Expenditure Systems." American Economic Review 68:348-359.

Pollack, R. A., and T. J. Wales. 1980. "Comparison of the Quadratic Expenditure System and Translog Demand Systems with Alternative Specifications of Demographic Effects." Econometrics 48:595-612.

Porter, Richard C. 1982. "The New Approach to Wilderness Preservation Through Benefit Cost Analysis." Journal of Environmental Economics and Management 9:59-80.

Randall, Alan, and John R. Stoll. 1980. "Consumer's Surplus in Commodity Space," American Economic Review 70(June):449-55.

Ravenscraft, David J., and John F. Dwyer. 1978. "Reflecting Site Attractiveness in Travel Cost-Based Models for Recreation Benefit Estimation." Forestry Research Report 78-6, Department of Forestry, University of Illinois at Urbana-Champaign, Urbana, Illinois, July.

Reiling, S. D., K. C. Gibbs, and H. H. Stoevner. 1973. Economic Benefits from an Improvement in Water Quality. Washington, D. C., U. S. Environmental Protection Agency.

Russell, Clifford S. and Williams A. Vaughn, 1982. "The NationalRecreational Fishing Benefits of Water Pollution Control." Journal of Environmental Economics and Management 9:328-354.

Saarinen, T. F. 1979. "Commentary-Critique of Bunting-Guelke Paper." Annals of the Association of American Geographers 69:464-468.

St-Louis, N. and P. Legendre. 1982. "A Water Quality Index for Lake Beaches." Water Research 16:945-948.

Samuelson, P. A. 1950. "The Problem of Intergrability in Utility Theory." Economica 17:335-385.

Seckler, D. W. 1966. "On the Uses and Abuses of Economic Science in Evaluating Public Outdoor Recreation." Land Economics 42:485-94.

Seckler, D. W. 1968. "Analytical Issues in Demand Analysis for Outdoor Recreation: Comment." American Journal of Agricultural Economics 50:147-51.

Sellar, C., J. P. Chavas and J. R. Stoll. 1983. "Specification of the Logit Model: The Case of Valuation of Non-Market Goods." National Resource Working Paper Services, Department of Agricultural Economics, Texas A & M University.

Sellar, C., J.R. Stoll, and Jean-Paul Chavas. 1983. "Validation of Empirical Measuring of Welfare Change: A Comparison of Nonmarket Techniques." Department of Agricultural Economics, Texas A & M University, Resources Group Working Paper.

Seneca, J. 1969. "Water Recreation, Demand and Supply." Water Resources Research 5:1177-1185.

Shechter, M., and R. C. Lucas. 1978. Simulation of Recreation Usefor Park and Wilderness Management. Johns Hopkins University Press, Baltimore.

Sinden, J. A. 1974. "A Utility Approach to the Valuation of Recreational and Aesthetic Experiences." American Journal of Agricultural Economics 56:61-72.

Small, K. A., and H. S. Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." Econometrica 49:105-130.

Smith, Terrence P. 1983. "An Application of a Numerical Algorithm to Measure Welfare Change." University of Maryland, College Park, manuscript.

Smith, Terrence P. 1983. "Incorporating Time into the Utility Maximization Problem."

Smith, V. K. 1975a. "The Estimation and Use of Models of the Demand for Outdoor Recreation." In Assessing the Demand for Outdoor Recreation. Washington, D.C.: National Academy of Sciences.

Smith, V. K. 1975b. "Travel Cost Demand Models for Wilderness Recreation: A Problem of Non-Nested Hypotheses." Land Economics 51:2:103-111.

Smith, V. K. 1981. "Congestion, Travel Cost, Recreational Demand Models and Benefit Evaluation." Journal of Environmental Economics and Management 8:92-96.

Smith, V. K. 1981. "Introduction to Advances in Applied Microeconomics and Some Perspectives." In Advances in Applied Microeconomics, V.K. Smith (ed.), Jai Press, 1:1-11.

Smith, V. K. 1983. "Option Value: A Conceptual Overview." Southern Economic Journal 49(Jan):654-68.

Smith, V. K., and W. H. Desvousges. 1983. "The Valuation of Risk Reductions Associated with Hazardous Waste Regulations." Draft, 54 pps.

Smith, V. K., and W. H. Desvousges. 1985. "The Generalized Travel Cost Model and Water Quality Benefits: A Reconsideration." Southern Economics Journal. 51:371-81.

Smith, V. K., W. H. Desvousges, and Matthew P. McGivney. 1983. "The Opportunity Cost of Travel Time in Recreation Demand Models" Land Economics 59:3:259-78.

Smith, V. K., and R. J. Kopp. 1980. "The Spatial Limits of the Travel Cost Recreational Demand Model." Land Economics 56:1: 64-72.

Smith, V. K., and J. V. Krutilla. 1982. "Toward Formulating the Role of Natural Resources in Economic Models." In V. K. Smith and J. V. Krutilla (eds.), Explorations in Natural ResourceEconomics. Baltimore: John Hopkins, pp. 1-43.

Smith, V. K., and V. G. Munley. 1978. "The Relative Performance of Various Estimators of Recreation Participation Equations." Journal of Leisure Research 10:165-176.

Smith, V. K., and D. Waldman. 1982. "A Comparison of Ordered Logit and Ordered Probit: Some Monte Carlo Experiment Results." Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, manuscript.

Smith, R. J. and N. J. Kavanaugh. 1969. "The Measurement of Benefits of Trout Fishing: Preliminary Results of a Study at Grafham Water, Great Ouse Water Authority, Huntingdonshire." Journal of Leisure Research 1:4(Aug):316-332.

Steiner, Peter O. 1968. "Comment on 'The Value of Travel Time'," in Samuel B. Chase (ed), Problems in Public Expenditure Analysis. Washington, D. C. Brookings Institution.

Stevens, Joe B. 1986. "Recreation Benefits from Water Pollution Control." Water Resources Research 2:2:167-182.

Stoevener, H. H., and W. G. Brown. 1967. "Analytical Issues in Demand Analysis for Outdoor Recreation." Journal of Farm Economics 49:1295-1304.

Stoevener, H. H., and W. G. Brown. 1968. "Analytical Issues in Demand Analysis for Outdoor Recreation." American Journal of Agricultural Economics 50:151-3.

Stoll, J. R. 1983. "Recreational Activities and Nonmarket Valuation: The Conceptualization Issues." Southern Journal of Agricultural Economics 15:2:119-25.

Stoll, J. R., R. N. Shulstad, and W. M. Smathers, Jr. 1983. Nonmarket Valuation: Current Status, Future Directions. Proceedings of a Regional Workshop, Southern Natural Resources Economics Committee.

Strong, E. J. 1983a. "A Note on the Functional Form of Travel Cost Models with Zones of Unequal Population." Land Economics 59:342-49.

Strong, E. J. 1983b. "Measuring Benefits of Outdoor Recreation Services: An Application of the Household Production Function Approach to the Oregon Steelhead Sport Fishery." Unpublished Master's Thesis, 155 pps.

Sutherland, R. J. 1982a. "A Regional Approach to Estimating Recreation Benefits of Improved Water Quality." Journal of Environmental Economics and Management 9:229-247.

Sutherland, R. J. 1982b. "The Sensitivity of Travel Cost Estimates of Recreation Demand to the Functional Form and Definition of Origin Zones." Western Journal of Agricultural Economics 7:1:87-98.

Sutherland, R. J. 1982c. "Recreation Benefits and Displaced Facilities." Journal of Leisure Research 14:248-62.

Takayama, Akira. 1974. Mathematical Economics. The Dryden Press. Hinsdale, IL. 744 pps.

Takayama, Akira. 1977. "Sensitivity Analysis in Economic Theory." Metroeconomica 29:9-37.

Thayer, Mark A. 1981. "Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence." Journal of Environmental Economics and Management 8:27-44.

Theil, H. 1951-1952. "Qualities, Prices and Budget Enquiries." Review of Economic Studies 19:129-47.

Theil, H. 1971. Principles of Econometrics. John Wiley and Sons, Inc.

Trice, Andrew H., and Samuel E. Wood. 1958. "Measurement of Recreation Benefits." Land Economics 34:195-207.

U. S. Congress, Senate. 1964. "Evaluation Standards for Primary Outdoor Recreation Benefits." Supplement 1 to Senate Doc. 97, 87th Congress, 2nd Session, Policies, Standards, and Procedures in the Formulation, Evaluation and review of Plans for Use and Development of Water and Related Land Resources. Washington, D. C.

Vartia, Y. O. 1983. "Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions." Econometrica 51:1:79-98.

Vaughan, W. J., and C. S. Russell. 1982. "Valuing a Fishing Day: An Application of a Systematic Varying Parameter Model." Land Economics 58:(Nov)4:450-463.

Vaughan, W. J., and C. S. Russell. 1982. "Fresh Water Recreational Fishing: The National Benefits of Water Pollution Control." Washington, D.C.: RFF, Inc.

Vaughan, W. J., E. S. Russell, L. P. Granessi and L. A. Nielsen. 1982. "Measuring and Predicting Water Quality in Recreation Related Terms." Journal of Environmental Economics and Management 15:363-380.

Vaughan, W. J., C. S. Russell and M. Hazilla. 1982. "A Note on the Use of Travel-Cost Models with Unequal Zonal Populations: Comment." Land Economics 58:400-07.

Wales, T. J. 1978. "Labour Supply and Commuting Time, An Empirical Study." Journal of Econometrics 8:215-226.

Wales, T. J., and A. D. Woodland. 1976. "Estimation of Household Utility Functions and Labor Supply Response." International Economics Review 17:2:397-410.

Wales, T. J. and A. D. Woodland. 1977. "Estimation of the Allocation of Time for Work, Leisure and Housework." Econometrica 45:115-32.

Wales, T. J. and A. D. Woodland. 1978. " A Limited Dependent Variables Model for a System of Share Equation, with an Application to Meat Consumption." Department of Economics, University of British Columbia, Vancouver.

Wales, T. J., and A. O. Woodland. 1983. "Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints." Journal of Econometrics 21:263-285.

Walsh, R. G. 1974. "Effects of Improved Research Methods on the Value of Recreational Benefits." Paper presented at the National Symposium on the Economics of Outdoor Recreation, New Orleans.

Walsh, R. G., and Lynde O. Gilliam. 1982. "Benefits of Wildnerness Expansion with Excess Demand for Indian Peaks." Western Journal of Agricultural Economics 7:(Jul)1:1-12.

Walsh, R. G., R. A. Gillman and J. B. Loomis. 1982. Wilderness Resource Economics: Recreation Use and Preservation Values. Report by the Department of Economics, Colorado State University, to the American Wilderness Alliance, Denver.

Walsh, R. G., D. G. Greenley, R. A. Young, J. R. McKean, and A. A. Prato, 1978. Option Values, Preservation Values and Recreational Benefits of Improved Water Quality: A Case Study of the South Platte River Basin, Colorado. EPA-600/5-78-001, U. S. Environmental Protection Agency, Office of Research and Development.

Walsh, R. G., John B. Loomis, and Richard S. Gilman. 1984. "Valuing Option, Existance and Bequest Demands for Wilderness." Land Economics 60:1:14-29.

Walsh, R. G., Nicole P. Mill and Lynde O. Gilliam. 1983. "Congestion and Willingness to Pay for Expansion of Skiing Capacity." Land Economics 59:2:195-210.

Ward, Frank A. 1981. "A Review of Problems and Solution in the Use of the Travel Cost Method for Valuing Recreation Resource." In Outdoor Recreation and the Public Interest. R. L. Gum and L. M. Arthur (ed.), Oregon Agr. Exp. Sta. Spec. Ref. 610.

Ward, Frank A. 1983. "On the Necessity of Simultaneous Recreation Demand Equation Estimation: Comment." Land Economics 59:4:455-58.

Ward, Frank A. 1983. "Measuring the Cost of Time in Recreation Demand Analysis: Comment." American Journal of Agricultural Economics 65:1:167-8.

Water Resources Council. 1973. "Principles and Standards for Planning Water and Related Land Resources. Federal Register 38, No. 174.

Water Resources Council. 1979. "Procedures for Evaluation of National Economic Development (NED) Benefits and Costs in Water Resource Planning (Level C). Federal Register 72950-72965.

Water Resources Council. 1980. "Proposed rules: Principles, Standards and Procedures for Planning Water and Related Land Resources." Federal Register 45, No. 73.

Wennergren, E. B., Herbert H. Fullerton, John E. Keith, and Robert Meale. 1975. "Economic Value of Water Oriented Recreation Quality." Utah Water Research Laboratory Study PRRAE 805-1. Logan, Utah, Utah State University.

Wennergren, E. B., and Darwin B. Neilsen. 1970. "Probability Estimates of Recreation Demand." Journal of Leisure Research 2:112-122.

Wetzel, James N. 1977. "Estimating the Benefits of Recreation Under Conditions of Congestion." Journal of Environmental Economics and Management 4:(Sept):239-46.

Wetzstein, M. E., and Richard D. Green. 1978. "Use of Principal Component Attractiveness Indexes in Recreation Demand." Western Journal of Agricultural Economics 62:11-21.

Wetzstein, M. E., and J. G. McNeely. 1980. "Specification Errors and Inference in Recreation Demand Models." American Journal of Agricultural Economics 62:799-800.

Wetzstein, M. E., and Rod F. Ziemer, 1982. "An Application of A Truncated Regression Model to Recreation Demand." Mimeo 15 pps. Presented at American Agricultural Economics Association meetings.

Willig, R. D. 1976. "Consumer's Surplus Without Apology". American Economic Review 66:4:589-97.

Willig, R. D. 1978. "Incremental Consumer's Surplus and Hedonic Price Adjustment." Journal of Economic Theory 17:227-253.

Wilman, Elizabeth A. 1980. "The Value of Time in Recreation Benefit Studies." Journal of Environmental Economics and Management 7:9:272-86.

Wilson, R. W., and F. W. Bell. 1978. "A Method of Measuring the Value of Improvements in Water Quality for Recreational Fishing." Proceedings of Business and Economic Statistics Section of ASA.

Wilson, R. R. 1970. "Consumer Behavior Models: Time Allocation, Consumer Assembly, and Outdoor Recreation." An Economic Study of the Demand for Outdoor Recreation. Coop. Reg., Res. Tech. Comm. Rep No. 2. Reno, Nev., pp. 16-18.

Wilson, R. R. 1971. "Demand Theory Time Allocation and Outdoor Recreation." Southern Journal of Agricultural Economics 3:103.

Woodland, Alan D. 1979. "Stochastic Specification and the Estimation of Share Equations." Journal of Econometrics 10:361-383.

Yatchew, Adonis. 1983. "Multivariate Distributions Involving Ratios of Normal Variables." Discussion paper, Dept. of Economics, University of Toronto.

Yu, J. K. and M. M. Fogel. 1978. "The Development of a Combined Water Quality Index." Water Resources Bulletin 14:1239-1250.

Zellner, Arnold. 1978. "Estimation of Functions of Population Means and Regression Coefficients including Structural Coefficients: A Minimum Expected Loss (MELO) Approach." Journal of Econometrics 8:127-158.

Zellner, Arnold and Soo-Bin Park. 1979. "Minimum Expected Loss (MELO) Estimators for Functions of Parameters and Structural Coefficients of Econometric Models." Journal of the American Statistical Association 74:365:185-193.

Ziemer, R. F., and W. N. Musser. 1979. "Population Specific Recreation Demand Models and the Consequences of Pooling Sample Data." Western Journal of Agricultural Economics 4:121-128.

Ziemer, R. F., W. N. Musser and R. C. Mill. 1980. "Recreation Demand Equations: Functional Form and Consumer Surplus." American Journal of Agricultural Economics 62:136-141.