

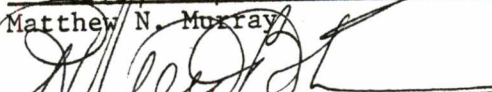
To the Graduate Council:

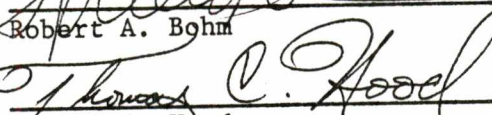
I am submitting herewith a dissertation written by David Michael Fletcher entitled "An Investigation into the Comparability of Responses in Hypothetical and Actual Markets for an Environmental Public Good". I have examined the final copy of the dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy with a major in Economics.


James R. Kahn, Major Professor

We have read this dissertation
and recommend its acceptance:


Matthew N. Murray


Robert A. Bohm


Thomas C. Hood

Accepted for the Council:


Associate Vice Chancellor
and Dean of The Graduate School

AN INVESTIGATION INTO THE COMPARABILITY OF
HYPOTHETICAL AND ACTUAL MARKETS FOR
AN ENVIRONMENTAL PUBLIC GOOD

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

David Michael Fletcher

August 1995

ABSTRACT

Currently two methods are commonly employed to estimate the value of an environmental resource. These are the travel cost method and the contingent valuation method. This dissertation empirically examines whether the hypothetical nature of the contingent valuation method produces a hypothetical bias. The results of this dissertation indicate that hypothetical methods, such as the contingent valuation method, may indeed systematically overestimate the value of environmental public goods due to the presence of a hypothetical bias.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION AND PURPOSE.....	1
The Importance of Further Research.....	2
Natural Resource Damage Assessment.....	3
Environmental Policy Formulation.....	5
A Brief History and Description of Non-Market Valuation Techniques.....	9
The Travel Cost Method.....	13
The Hedonic Price Method.....	18
The Contingent Valuation Method.....	21
Comparability of Hypothetical and Actual Responses.....	23
The Contribution of this Dissertation.....	27
II. HYPOTHETICAL VERSUS ACTUAL RESPONSES.....	32
Potential Sources of Bias in Hypothetical Responses.....	34
Motives to Misrepresent Opinions.....	38
Lack of Incentives to Answer Carefully.....	42
Responses that Reflect Factors Other than Values.....	49
Inadequate Understanding.....	55
Review of the Hypothetical versus Actual Response Literature.....	57
Attitudes and Behavior.....	58

	Comparability Tests Involving Private Goods.....	62
	Comparability Tests Involving Public Goods.....	69
	Importance of Testing the Comparability of Actual and Hypothetical Responses in the Context of the Travel Cost Method.....	74
III.	THEORETICAL DEVELOPMENT OF THE EMPIRICAL MODEL.....	79
	The Traditional Travel Cost Model and Marshallian Consumer Surplus.....	80
	The Household Production Framework.....	85
	Hicksian Welfare Measures.....	88
	Derivation of Hicksian Welfare Measures Directly from Marshallian Demand Curves.....	92
	A Reconsideration of Marshallian Consumer Surplus.....	93
	A Testable Hypothesis.....	96
IV.	THE EMPIRICAL MODEL.....	105
	The Data Set.....	105
	The Choice of Functional Form for Travel Cost Demand Estimation.....	112
	The Opportunity Cost of Time.....	114
	Estimated versus Actual Number of Trips Taken.....	116
	Statistical Tests of the Null Hypothesis.....	118
V.	EMPIRICAL RESULTS.....	123
	Demand Estimation Results.....	124

	Estimates of Marshallian Consumer Surplus.....	132
	Results of Statistical Tests.....	136
VI.	CONCLUSIONS.....	145
	Summary.....	145
	Results and Conclusions.....	149
	Suggestions for Future Research.....	155
	BIBLIOGRAPHY.....	158
	APPENDIX.....	171
	VITA.....	173

LIST OF TABLES

TABLE		PAGE
Chapter 1: Introduction and Purpose		
1. Methods for estimating values for environmental goods.....		10
Chapter 2: Hypothetical versus Actual Responses		
1. Results from the Map Experiments.....		64
Chapter 3: Theoretical Development of the Empirical Model		
1. Formulas for the Derivation of Marshallian Consumer Surplus.....		83
Chapter 4: The Empirical Model		
1. Travel Cost Model Variables.....		108
Chapter 5: Empirical Results		
1. Studentized Residuals.....		127
2. Final Model - Log-linear Form - Actual Responses.....		130
3. Consumer Surplus Estimates.....		135
4. Matrix of Test Results.....		138

Chapter One

Introduction and Purpose

Environmental public policy decisions have become a prominent concern at the global, federal, state, and local government levels. Examples of such public policy choices include deciding whether the old growth forest, home to the spotted owl, should be preserved as habitat conservation areas, or harvested as a source of timber, and deciding whether steps which would require current economic sacrifices should be taken to reduce the potential for global warming. Additionally, litigation seeking to recover damages caused by polluters has also become an important issue. This type of litigation was pursued by the Federal government and the State of Alaska when the Exxon Valdez oil spill occurred in Prince William Sound. These two pursuits require very different courses of action, but they are related in that both require the measurement of the value of non-market goods.

The measurement of the social value of environmental and natural resources requires special valuation techniques, because standard economic valuation techniques cannot be applied to this class of goods. Normal market goods are associated with readily observable price and quantity data from which demand curves can be estimated, but such data are not directly available for environmental resources because they are not normally bought and sold in markets. If any price at all is paid for an environmental good, such as an entrance fee paid for admission to a national park,

then this price is usually very small and not related to the demand for, or the cost of supplying the environmental resource.

The Importance of Further Research

There are several distinct techniques which can be used to value environmental goods, each of which have their own unique strengths and weaknesses. The question of which of these techniques should be employed for environmental valuation tasks has sharply divided environmental economists. This division has been amplified due to the crucial role that environmental valuation estimates play in public policy decisions and natural resource damage assessments.

The combination of the policy relevance and the technical difficulty associated with these techniques implies that they will be controversial. The policy relevance also suggests that there is a need to answer many of the unresolved research issues in a timely fashion. This dissertation will focus on a subset of these issues; the correspondence between hypothetically stated values and actual values in the context of an environmental good. This study will investigate this correspondence by comparing valuations of a public good (deer hunting sites) based on hypothetical responses to travel cost questions to valuations based on reports of actual travel cost behavior.

Natural Resource Damage Assessment

The importance of accurately measuring the social value of environmental and natural resources, and therefore the importance of this research, is not likely to be fully appreciated without an understanding of the magnitude of the actions and decisions which are based on environmental valuations. In order to provide the foundation for this understanding a discussion of the two primary uses of environmental valuation, natural resource damage assessment and environmental policy formulation, will be presented in the following pages.

The Exxon Valdez oil tanker gathered national attention when it ran aground on the Bligh Reef in Prince William Sound off the coast of Alaska on March 24, 1989. This event raised public awareness of the severity of the damages that can result from an oil spill and the associated importance of natural resource damage assessment.

How much should the Exxon Corporation pay for polluting Prince William Sound? Exxon was legally liable for both cleanup costs and damages to the natural resource. Although the cleanup costs were fairly straightforward to measure, as there were accounting records for labor, materials, and equipment being billed to Exxon, the damages to the environment were more difficult to derive. Estimation of these damages required economists to use non-market valuation techniques.

Under the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA) Congress gave federal, state, and local governments

the right as trustees to sue for damages resulting from the release of harmful materials into publicly owned natural resources such as rivers, lakes, oceans, or terrestrial habitats (Kopp and Smith 1993). CERCLA and its reauthorizing amendments, the Superfund Amendments and Reauthorization Act of 1986 (SARA), established two types of liability for releases of hazardous substances.¹ The first is the cost of cleaning up the hazardous substance and restoring the damaged site. The second is defined as the residual liability arising after cleanup. This residual liability allows for compensation for damages to the resource from the time it is injured until it is cleaned up, as well as compensation for any damages to the resource remaining after the cleanup.

The correct monetary valuation of these damages is important due to the role it plays as an incentive mechanism. The compensation payments required for oil spill damages will send price signals to corporations engaged in transporting oil. These corporations can then calculate the expected monetary cost of the potential liability assumed when they transport oil. This expected cost will provide the necessary incentive for firms to take precautionary steps, such as the use of double-hulled tankers, as long as the marginal cost associated with taking precautionary measures does not exceed the marginal benefits of reducing potential liability. In this manner,

¹CERCLA does not classify petroleum products as hazardous substances. Therefore, claims for natural resource damages due to oil spills, such as the Exxon Valdez spill, are filed under state common law, provisions of the Clean Water Act, or provisions of the Oil Pollution Act of 1990 (Kopp and Smith 1993).

the correct measurement of oil spill damages can lead to firms taking the socially optimal amount of precautionary measures.

Litigation driven damage assessments are ex post facto in nature. These damage assessments, as in the Exxon Valdez case, are performed after the damages have already occurred.² In contrast when non-market valuation estimates are used in a public policy analysis setting they are performed before the public policy change is implemented. Public policy decisions are the second major area for the usage of non-market valuation techniques in the field of environmental economics.

Environmental Policy Formulation

Many applications of environmental valuation are intended to aid in the making of good policy decisions. These applications include assessing the desirability of a specific project, choosing among competing projects, or deciding upon the optimal level of pollution for society.

Assessing the worthiness of one or a group of potential projects involves comparing the costs of attaining some level of environmental quality with the benefits

²The State of California is currently attempting to develop a baseline model for evaluation of future oil spills. These models are intended to send price signals to corporations engaged in transporting oil. After receiving these price signals the corporations should increase precautionary measures for protecting the most highly valued resources relative to lesser valued resources. These signalling effects should induce more efficient behavior on the part of corporations, because increased knowledge of the potential litigation damages would more accurately internalize the oil spill externality (Kopp and Smith 1993).

associated with achieving that level of environmental quality. This process is known as benefit-cost analysis.

Consider the following application of benefit-cost analysis to the red drum fishery. In the late 1980's red drum were pursued by both commercial and recreational fishermen in the coastal waters of the Gulf of Mexico. A grass roots movement in Gulf coast states like Florida called for either restricting or prohibiting the commercial fishing of red drum.

In 1989 Trellis Green estimated the increase in net benefits to recreational fishermen of a 10% and 50% increase in expected red drum catch rates to be \$10.34 and \$73.6 million dollars annually. The anticipated loss of producer surplus was not directly estimated for the commercial sector, but based on published studies by other authors Green found that the losses to the commercial industry were likely to be extremely small when compared to the gains of the recreational fishermen. Green concluded that there would be a positive net benefit from restricting or prohibiting the commercial fishing of red drum.

In addition to fisheries management, benefit-cost analysis can also be applied to other important policy questions. Should electric utility rates be increased to pay for improvements in air quality? Should the old-growth forest be preserved as habitat conservation areas? Should a hydroelectric dam reduce water discharges used for power production in order to protect salmon populations? Should a new dam be built for power and irrigation purposes, when pristine recreational and wilderness areas will be destroyed in the process? Benefit-cost analysis can be applied to each of these

questions, but only if the value of cleaner air, the old growth forest, and other natural resources can be accurately estimated. Estimation of these values requires the application of non-market valuation techniques.

In addition to applying benefit-cost analysis to project evaluation, public policy makers must also formulate pollution control strategies. Pollution control strategies in the United States have primarily relied on command and control policies.³ Recently, with the support of most economists, policies based on economic incentives such as marketable pollution permits and per unit pollution taxes are beginning to be implemented. The implementation of either direct controls or economic incentives requires the choice of some target level of pollution. Most economists would agree that this target should be what is commonly referred to as the optimal level of pollution, that is the level of pollution at which the marginal costs of pollution abatement are equal to the marginal damages of pollution.

Achieving the optimal level of pollution will minimize the total social cost of pollution.⁴ In order to achieve the goal of determining an optimal level of pollution,

³Command and control policies, often referred to as direct controls, place rigid constraints on the behavior of firms or households. If a firm is not acting within these constraints it is considered to be acting unlawfully. Command and control constraints usually specify allowable levels of inputs or outputs. Requiring sulfur removing scrubbers on the smokestacks of coal burning utilities is an example of a restriction on inputs, and prohibiting the dumping of toxic substances is an example of a restriction on outputs.

⁴The total social cost of pollution are minimized when the marginal abatement costs of reducing pollution are equated to the marginal damages pollution. At any level of pollution emissions greater than this, the damages of the last unit of pollution are greater than the cost of abating it, and therefore total social cost, which equal the summation of the abatement cost and the damages, can be lowered by abating another

the public policy maker must acquire knowledge of the marginal abatement cost function and the marginal damage function.⁵ Estimating the marginal abatement cost function is the easier of the two tasks. The marginal abatement cost function can be estimated using market prices of such inputs as labor, alternative raw materials, and pollution reducing equipment and information regarding the rates at which these inputs will physically reduce pollution emissions.

Marginal damage function estimation requires two steps. First, information on the physical relationship between additional emissions of pollution and the resulting damages to items we value such as human health, ecosystems, aesthetics, crop yields, and human made structures must be obtained. Second, the dollar values of such damages must be estimated. Monetary values of damages to human health, aesthetics, and ecosystems can only be estimated by using non-market valuation techniques.

Measuring the monetary values of environmental resources is an important task. Without these monetary measures the processes described above of benefit-cost analysis, natural resource damage litigation, and choosing efficient pollution control policies would be made much more difficult. As environmental issues continue to grow in prominence the importance of non-market valuation techniques will also continue to increase.

unit of pollution. A similar analysis holds for the case when the marginal cost of pollution abatement exceed the marginal damages of additional pollution.

⁵Even if a public policy is not intended to reach this optimal level of pollution abatement the correct estimation of the marginal benefits and costs is still important. If the marginal benefits of a policy are greater than the marginal costs then at least the policy will result in a net benefit to society.

A Brief History and Description of Non-market Valuation Techniques

The following discussion of valuation techniques is necessary for several reasons. First, the descriptive classification that follows will help to explain the importance of comparing a "hypothetical indirect" travel cost model to an "actual indirect" travel cost model. Second, without an understanding of the development and shortcomings of the travel cost and hedonic price methods the reader may not develop a full understanding of why valuation methods based on hypothetical answers are needed. Finally, a similar review of the development and shortcomings of the contingent valuation method will provide both additional reasons for the necessity of hypothetical methods and an introduction to the serious reservations that many economists have concerning such valuation methods.

A variety of methods have been used to measure the values of environmental assets. These methods can be classified according to two characteristics (Mitchell and Carson 1989). The first characteristic is whether the data come from observing individuals who are taking actions the consequences of which they must accept, or the data come from individual's responses to hypothetical questions. The second characteristic is whether the method produces monetary values for the environmental good directly or the method requires the use of a model of individual economic behavior to indirectly infer monetary values. Mitchell and Carson (1989) construct four categories based on these two characteristics. Any non-market valuation

technique will fit neatly into one of these categories. The four categories: observed direct, observed indirect, hypothetical indirect, and hypothetical direct, and many of the valuation techniques that fit into these categories are presented in Table 1-1.⁶

Table 1-1 Methods for Estimating Values for Environmental Goods

	OBSERVED BEHAVIOR	HYPOTHETICAL
DIRECT	<u><i>Observed Direct</i></u>	<u><i>Hypothetical Direct</i></u>
	Market prices and quantities	Contingent valuation
	Simulated markets	
INDIRECT	<u><i>Observed Indirect</i></u>	<u><i>Hypothetical Indirect</i></u>
	Travel cost	Hypothetical travel cost
	Hedonic price	Conjoint Analysis
	Averting behavior	Contingent Ranking

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

Observed direct methods contain both valuations based on market prices and quantities, and those derived from simulated markets. Due to the public goods nature of most environmental goods, researchers are usually unable to obtain market quantity and price data. There are also relatively few simulated market studies. The difficulty

⁶Those methods characterized as "indirect" are methods that estimate the value for an environmental good based on price and quantity information of a market good which possesses some type of complementary or substitute relationship with the environmental good. Examples of such market goods are housing prices or trip cost. Conversely, the characteristic "directness" shall refer to valuation processes based on the price and quantity data of the environmental good being valued. The characteristic "direct" should not be taken to imply that it is unnecessary to estimate a demand curve to arrive at an estimate of consumer surplus.

in obtaining data is the primary factor limiting the number of simulated market studies, and although they have not become a prominent method for non-market valuation, simulated market studies have made an important contribution to the literature on hypothetical bias. Observed direct methods base values on observable real world actions taken by individuals and monetary values are direct in the sense that decisions regarding these goods are based on their monetary prices.

Observed indirect methods include among others the travel cost method and the hedonic price method. These two methods have been used in many attempts to value environmental goods. Many economists favor observed indirect methods because they are based on actual market behavior. The use of this class of models requires the assumption of some type of substitute or complementary relationship between the observed market good and the environmental good. For example, the travel cost method assumes a complementary relationship between a recreation site and the costs associated with travel to the recreation site such as fuel purchases.

Hypothetical indirect methods often use many of the same assumptions and operational formats as the observable indirect class. The primary difference is that hypothetical indirect methods utilize answers to hypothetical questions instead of looking to actual market behavior. For instance, the hypothetical travel cost method might ask a respondent how far they would travel to use a site or how many trips they would make to a recreation site given an increase in environmental quality. Conjoint analysis, which is also a hypothetical indirect method, has shown potential in

marketing applications. Unfortunately, conjoint analysis has only been applied to the valuation of environmental goods in a very small number of cases.⁷

Hypothetical direct methods directly ask respondents how much they value an environmental good. For example, an individual might be asked how much she is willing to pay for improved water quality. This question is an example of contingent valuation which is the most prominent of the hypothetical direct methods.

Hypothetical direct methods avoid the many assumptions about the relationship between market goods and the environmental good needed by the observed indirect methods. However, hypothetical direct methods must make the crucial assumption that individual responses to hypothetical questions are completely comparable with responses in actual markets. If this assumption can be accepted then hypothetical direct methods become available which are "unique in their simplicity, theoretical justification, and ability to value the entire range of benefit categories" (Mitchell and Carson 1989, 86) It is the validity of this assumption of comparability between hypothetical and actual responses that is the focus of this dissertation.

⁷The most notable of these applications utilized conjoint analysis to derive willingness to pay measures of value for hunting trips (Mackenzie 1993). Other studies include the ranking of rhododendrons attributes (Gineo 1990), and valuing changes in water quality (Adamowicz, Louviere, and Williams 1992).

The Travel Cost Method

Most value estimates have been produced using either the travel cost method, the hedonic price method, or the contingent valuation method. The travel cost method was first suggested in a letter from Harold Hotelling to the director of the National Park Service in 1947. Trice and Wood (1958) and Clawson (1959), were the first to actually implement the travel cost method to value an environmental good. Later Marion Clawson and Jack Knetsch brought the travel cost method to a broader audience when The Economics of Outdoor Recreations was published by Resources for the Future in 1966. These initial applications were the beginnings of a major line of research in environmental economics.

The travel cost method proposed by Clawson and Knetsch (1966) was intended to measure the recreational demand for a single site. This method essentially uses the cost of travel, both to and from the recreation site, as a proxy for the price of visiting the recreation site.⁸ The Clawson-Knetsch model is often referred to as a zonal travel cost model. The simple zonal travel cost method proceeds as follows. First, the recreation site is identified and the surrounding region is divided into concentric circular zones. The cost of travel from each zone to the site and back is then measured. Next, an on-site sampling of visitors determines their zones of origin, and

⁸As was discussed earlier this proxy is necessary because the entrance fee to most recreation sites is either nominal or zero, and will not vary across individuals or time. This makes estimation of the individual's demand based on the entry price and number of visits impossible.

the number of visitor days per capita is calculated for each zone of origin. These visitation rates are then regressed on zonal travel costs and socioeconomic variables such as average income. The regression results are then used to trace out the demand curve for the recreation site.

The travel cost method has undergone a considerable evolution since the development of the zonal travel cost model. Some of the more important developments in the travel cost model include the following:

1. The use of individual level data as opposed to using zonal aggregates.
2. The use of the household production function framework to establish that if the assumption of weak complementarity holds then the area under the demand curve for the essential marketed good is exactly equal to the area under the demand curve for the public environmental good (Bockstael and McConnell 1983).⁹
3. The development of equations which made the computation of the Hicksian measure of compensating variation a straightforward exercise. Without these equations only the Marshallian measure of consumer surplus can be derived (Hausman 1981).
4. The incorporation of the opportunity cost of time into travel cost demand models. Time spent traveling to a site should be included as a component of

⁹The assumption of weak complementarity requires that individuals have zero value for the environmental good when none of the marketed good is purchased. For example, an individual who takes zero trips to a lake is assumed to place zero value on the existence or characteristics of the lake.

a person's overall travel cost (Cesario and Knetsch 1970; McConnell and Strand 1981).

5. The development of travel cost models which incorporate potential substitute sites and their travel cost (Milon 1988).

6. The inclusion of site quality into the travel cost method. A noteworthy example is the varying parameters model developed by Vaughn and Russell (1982) in which they assume the estimated coefficients of demand systems are a function of site quality characteristics.

The travel cost model is based on the idea that the cost of travel to a site is an important component of the full cost of a visit, and that there will usually be a wide variation in travel costs across a sample of visitors to that site. This variation in travel costs can be used to estimate a travel cost demand function. By utilizing the concept of weak complementarity the value of the site can be assumed to be equal to the area under the compensated demand curve.¹⁰

Although most environmental economists have concluded that the travel cost method has performed well (Smith 1993), a number of potential weaknesses must be considered. The most serious of these potential problems include the following:

¹⁰Utilizing the concept of weak complementarity in the framework of the travel model is basically equivalent to assuming that a person must travel to a recreation site in order to obtain any satisfaction from the existence of an environmental asset such as a lake. Weak complementarity is discussed in more detail in chapter three.

1. Properly measuring the travel cost associated with visiting a site may require choosing some arbitrary method for incorporating the opportunity cost of travel time into the travel cost measurement, and allocating the cost for multi-purpose trips across sites.

2. Travel cost interviews conducted on-site collect data only from individuals who visit the site. These on-site interviews will give rise to a truncation bias (no observations of zero trips) which must be corrected for. The researcher must also account for any endogenous stratification bias (over representation of people who take more trips).

3. Correctly accounting for the effects of substitute sites when estimating the demand for a particular site requires that the researcher properly identify the set of sites that the recreationist regards as potential substitute sites.

4. Estimating the effects of changes in site quality characteristics involves two difficulties. First, the researcher must identify which recreational quality characteristics, such as number of fish caught or clarity of water, matter to the recreationist. Next the relationship between measurable physical environmental quality characteristics, such as dissolved oxygen levels, and the recreational quality characteristics must be identified.

Each of these issues are serious, but the inability of the travel cost method to measure some types of values and entire classes of environmental goods is probably the most important flaw in the travel cost model. If a person makes zero trips to a

site then the travel cost method will estimate her value of the site as zero. It may be very likely that a site such as the Grand Canyon, is valued by individuals who have never taken a trip to it. Values of this type are referred to as non-use values. The entire class of observed indirect methods and most hypothetical indirect methods fail to estimate non-use values.¹¹ This shortcoming is no trivial matter as many environmental economists believe that non-use values can be much larger than use values in many cases.¹²

The travel cost model is further limited as to the types of environmental goods it can properly value. The travel cost model is primarily used to value site specific recreational benefits. It cannot measure, for example, effects on human health, protection of biodiversity, or the consequences of modifying or destroying particular ecosystems.

Although the travel cost method has several unresolved methodological issues and is limited as to the types of environmental goods it can be employed to value, it is regarded highly among many environmental economists. The travel cost method is the most widely employed of the indirect methods for measuring the use value of environmental goods.

¹¹Conjoint analysis, which is a hypothetical indirect technique may be able to measure these non-use values.

¹²The case of non-use values far exceeding use values is probably found in association with the protection of many species of wildlife, such as whales for instance.

The Hedonic Price Method

The hedonic price model is another indirect method for valuing environmental goods. Environmental applications of hedonic price models have focused on changes in either wages or property values, and their relationship to environmental quality. Hedonic property value models have been applied primarily to environmental amenities such as air quality, while hedonic wage models have been applied to both environmental amenities, and risk of death, injury or illness.

The hedonic property value model assumes that some environmental goods have a complementary relationship with residential property. Within a large metropolitan area it is possible to observe variations in air quality. Such a metropolitan setting will be used to provide a simplified example of how the hedonic property value model can be applied, although the model can also be applied among two or more distinct geographical areas (Freeman 1993).

The single large metropolitan area is divided into subregions. Data on sales of residential property and air quality are collected and grouped by subregion. Data are also gathered on other variables that may affect the sale price of residential property such as the number of square feet of living area, number of bathrooms, distance to shopping centers or work areas, number of local parks, and the crime rate of each region. The statistical relationship between each of these variables and the sale price of residential property is estimated. If the coefficient on the air quality variable is positive and statistically significant, then it can be concluded that consumers are

willing to pay more for property located in areas with cleaner air. The increase in the sale price of residential property, that is due to a one unit increase in air quality, can be derived from the coefficient on air quality previously estimated. This increase in sale price is the discounted present value of the present and future benefits associated with the increase in air quality. The annual benefits per household can now be calculated (Freeman 1993).

Hedonic wage models have been used primarily to value changes in the risk of injury or death.¹³ Such values are important for both public policy decisions and litigation. In an economic context these values can be expressed as the amount that an individual is willing to accept for a decrease in their life expectancy.

Hedonic wage models are based on the economic concept of compensating wage differentials. The basic premise is that, *ceteris paribus*, employees will prefer jobs with more favorable working conditions to ones with less favorable conditions. The larger supply of employees for the more pleasant jobs will tend to lower the wage levels for these jobs. In a manner similar to that used in the hedonic property value example the relationship between wages and a number of variables such as education and job risk is estimated. The difference in the wages between two jobs that are the same, except for the difference in working conditions, will reflect the employees monetary valuation of the differences in working conditions.

¹³Hedonic wage models have also been applied to environmental goods. Examples include a study of the benefits associated with changes in freshwater fishing quality (Clark and Kahn 1989) and an evaluation of changes in air quality (Cropper and Arriaga-Salinas 1980).

The primary attraction of the hedonic price method is its' reliance on observable behavior. The hedonic price method, however, has a number of weaknesses. First, it requires a large number of assumptions be made by the researcher. Second, the hedonic wage model may fail if workers are either unaware of, or misperceive the risk associated with a job. Finally, the hedonic property value model may measure environmental effects other than the targeted one. For example, air pollution may reduce visibility as well as cause health problems and damage materials.¹⁴

The hedonic price method can be used to measure the value of increased mortality risk, damages to nonliving systems such as materials, and regional environmental amenities such as air quality or noise levels (Freeman 1993). In relative terms the hedonic price method has been judged to be less reliable than the travel cost method (Smith 1993). It has been applied to far fewer studies, has a greater variation in benefit estimates for similar goods, and requires a more complex set of assumptions than the travel cost method.

¹⁴This is not a problem unless the researcher is attempting to measure only the benefits associated with increased visibility. If as is more commonly the case a comprehensive assessment of the benefits of an improvement in environmental quality is desired then this issue becomes irrelevant. An additional problem arises if other methods are used to value the other categories of change. In this case a potential for double counting exists.

The Contingent Valuation Method

The fastest growing body of literature in the field of non-market valuation centers on the contingent valuation method. The contingent valuation method uses survey questions to discover how much people value an environmental good. This method derives a respondent's willingness to pay in dollar amounts.¹⁵ These willingness to pay dollar amounts are exacted by creating hypothetical markets in which the survey respondents are given the opportunity to make purchases of the environmental good or service.

In 1947 Ciriacy-Wantrup first suggested using survey methods to value environmental goods. Later, in 1963 Robert Davis became the first to implement a contingent valuation study. Davis did not know of Ciriacy-Wantrup's suggestion when he began using surveys to estimate the value of outdoor recreation in a Maine backwoods area. Davis implemented a bidding game in which he interviewed 121 hunters and other recreationists. This research, which was his dissertation at Harvard University, was the first survey attempt at revealing a respondent's maximum willingness to pay.

¹⁵Alternately, individuals may be asked their willingness to accept compensation for loss of an environmental asset, or a decrease in its' quality. Willingness to accept measures are sought in only a small minority of contingent valuation studies. Willingness to accept measures are suspect in many cases, and may be quite different in size when compared to willingness to pay measures for an equivalent increase in environmental quality.

The next use of contingent valuation to value an environmental good was performed by Ronald Ridker in 1967. Ridker employed the contingent valuation method to value the benefits associated with air quality. Several other studies followed these two initial contingent valuation efforts in the early 1970's.¹⁶

Randall, Ives, and Eastman (1974) used the contingent valuation method to value the damages of reduced visibility levels resulting from the use of coal to produce electricity. This work was published in the first edition of the Journal of Environmental Economics and Management, which exposed the use of contingent valuation to a much larger group of economists. This study is regarded as the most theoretically rigorous of the early contingent valuation applications. Following these early studies the contingent valuation method was applied to a wider range of public goods. The most notable extensions have involved attempts to estimate the values people place on environmental goods which they never intend to use or visit (non-use values).

The contingent valuation method may be preferred both to the travel cost and hedonic price methods for its' ability to measure non-use values, its' ability to value a much wider range of public goods, its' avoidance of the many assumptions necessary to implement the travel cost and hedonic price models, and the relative simplicity of using the method. Despite these advantages many economists have been reluctant to

¹⁶These early studies include: Matthews and Brown (1970) salmon fishing; Cicchetti and Smith (1973) congestion reduction associated with wilderness recreation; Darling (1973) urban parks; Acton (1973) reduction in the risk of dying from a heart attack; Hammack and Brown (1974) migratory waterfowl hunting.

accept the contingent valuation method because of its' reliance on hypothetical data.

If the values given by respondents are not valid then the many advantages of using the contingent valuation method become meaningless.

Comparability of Hypothetical and Actual Responses

Economists have traditionally felt that benefit estimates based on hypothetical data would be suspect. In chapter 2 a number of potential sources of bias will be discussed, but much of the concern among economists can be summarized as follows. Expressions of behavior in a hypothetical context are usually costless (or almost costless) because these hypothetical responses are not associated with real economic consequences. Because mistakes are costless, choices made in hypothetical markets may not be considered as carefully as real consumption decisions. Therefore, many economists are reluctant to base valuations, that may be used to allocate real resources, on preferences that are not revealed through real economic choices.

Other economists believe that hypothetical valuation methods can provide accurate estimates, but is this confidence empirically supported? A relatively small number of studies have attempted to address the concern that answers to hypothetical questions may not be comparable to actual observable behavior. Experiments aimed at resolving this issue can be divided into two classes: experiments with private goods and experiments with public goods. An experiment with a private good may ask a group of individuals how much each person would be willing to pay for a painting

and then compare these hypothetical responses to the actions taken by another group who are actually given the opportunity to buy the same painting. Notable private goods experiments include: Neill, Cummings, Ganderton, Harrison, and McGuckin's (1994) experiments involving a watercolor painting by a Navajo Indian artist and a framed print of a sixteenth century map of the world; Dickie, Fisher, and Gerking's (1987) study involving the sale of fresh strawberries; Cummings, Harrison, and Rutstrom's (1994) study utilizing a pocket calculator, an electric juicer, and a box of chocolate truffles; and Kealy, Montgomery, and Dovidio's (1990) experiments utilizing chocolate bars.

An experiment with a public good might estimate the value of a recreation site using a hypothetical method such as the contingent valuation method or hypothetical travel cost method and then compare these results to those obtained using an observable method such as the travel cost method. Noteworthy public goods experiments include: Bishop and Heberlein's (1990) investigations involving deer and goose hunting permits; Seller, Stoll, and Chavas's (1985) inquiry utilizing boat ramp permits; and Smith and Desvouges's (1986) test involving water quality in the Mongahela River basin.

Throughout this dissertation the term "public good" shall refer to non-marketed goods which to some degree exhibit the public good's characteristics of nonrivalry and nonexcludability in consumption. The term public good is not meant

to imply that the good is a pure public good such as national defense or air quality.¹⁷

Public goods experiments face the difficulty of not having as direct a criterion for hypothetical responses as observed market transactions. Public goods experiments are necessary however because individuals may respond differently when asked questions involving private and public goods for a number of reasons.

First, the degree of familiarity is a strong determinant of a respondent's ability to accurately state her preferences regarding a good.¹⁸ Consumers are generally more *familiar* with the characteristics of most private goods and can therefore, more accurately reveal their willingness to pay for these goods.

People may also be familiar with current *market prices* for many private goods such as the strawberries mentioned above. People may find it easier to provide willingness to pay amounts for private goods simply because they are familiar with the market prices of these goods, and may tend to say they are willing to pay whatever the good costs, almost without considering how much they value the good.¹⁹

¹⁷Deer hunting sites will be utilized to perform this study. Although deer hunting sites are not pure public goods, because it is possible to exclude hunters from using them and they may become partially rival due to conjection or overhunting of the deer herd, they will be referred to as simply a public good. This is a common approach taken by many environmental economists.

¹⁸The importance of familiarity is discussed in the theory of reasoned action developed by Fishbein and Ajzen (1975). This will be discussed further in chapter two.

¹⁹This interpretation would be consistent with the findings of Dickie et. al. (1987) who found no statistically significant difference between hypothetically stated and actual willingness to pay amounts for strawberries.

Finally, many people have different principles concerning private and public goods. For example, some people find it morally reprehensible to buy and sell environmental amenities such as clean air.²⁰ Such individuals may tend to provide protest bids that represent an aversion to placing a monetary value on an environmental good.

The likelihood that individuals may respond differently when answering survey questions concerning public and private goods suggests that studies involving both classes of goods are necessary to adequately explore the hypothetical versus actual response issue. Experiments involving private goods represent a "best case" scenario for testing the comparability of hypothetical and actual responses. They represent a best case scenario because individuals are both more familiar with the characteristics of private goods and more accustomed to considering these goods in terms of a monetary valuation. If respondents are unable to give accurate hypothetical responses for these goods then it is unlikely that they can accurately give hypothetical responses for the less familiar public goods.

Although experiments involving private goods can make an important contribution to understanding the hypothetical versus actual response issue any conclusions based solely on these types of studies would be incomplete. The issue of importance after all is not whether hypothetical responses can be used to obtain accurate values for private goods, but rather, can public goods be valued using

²⁰For example, some environmentalists have characterized the trading of marketable pollution permits as the moral equivalent of selling the rights to commit crimes such as murder.

hypothetical questions. Therefore, experiments involving public goods are necessary because people may react differently when confronted with the task of stating hypothetical preferences for public goods for the reasons discussed above.

The Contribution of this Dissertation

The goal of this dissertation is to provide another piece of information which will help to resolve the hypothetical versus actual response debate. This information will help to address both the comparability issue and the emerging issue of hypothetical value calibration. This dissertation will not settle the debate. No single study can provide a resolution. It will however provide another important and unique piece of information that will help economists to understand the relationship between hypothetical and actual economic responses concerning environmental goods.

It has been asserted above that this dissertation is needed in order to provide additional evidence as to whether hypothetical surveys provide responses that are comparable to real economic commitments. It may seem to the reader of this paper that after learning of the importance of non-market valuation for litigation and benefit-cost analysis purposes, that surely hypothetical methods would not be used for these purposes unless there were an empirical basis for accepting the validity of their results. Unfortunately, this is simply not the case (Neill et al. 1994). Other leading researchers in this field such as Jonathan Baron (1994) and Howard Schumann (1994) agree with the proposition that acceptance of the hypothetical methods is premature.

These authors state that additional empirical research is needed in this area. This dissertation will address this need, by comparing valuations for deer hunting sites derived from reports of *actual* travel cost behavior to valuations based on *hypothetical* responses to travel cost questions.

In addition to addressing the comparability issue this dissertation will help to address the related issue of calibration. Although the calibration issue has only arisen recently it has gained prominence due to the actions of the National Oceanic and Atmospheric Administration (NOAA) commission. In January 1994 the draft NOAA rules for damage assessment included a suggestion that contingent valuation responses be calibrated downward by a factor of two (NOAA Jan. 1994).

The calibration issue accepts that the comparability between hypothetical and actual responses is well based empirically. I would assert that this is not the case, and therefore, any actions taken to systematically calibrate responses to hypothetical question are extremely premature. This position is taken by Alan Randall (1994, 4) who states that:

"The paradigm shift being promoted as 'the calibration issue' takes it as demonstrated that contingent valuation is systematically biased toward overestimation of welfare change, and directs the discussion toward the question: by how much should contingent valuation responses be calibrated downward? I would argue vigorously that the essential premise is unproven and the question is therefore premature and presumptuous".

By providing results using a hypothetical travel cost model and an actual behavior travel cost model this dissertation will provide additional empirical results on which to base future calibration decisions. This dissertation, however, is more than just

another study on which to base calibration and comparability conclusions. This empirical comparison of a hypothetical valuation method and a valuation method based on observable behavior is unique in several aspects.

Previous comparison studies involving public goods (Bishop and Heberlein 1990; Seller et al. 1985; Smith and Desvouges 1986) have compared an *indirect* observed method such as the travel cost method to a *direct* hypothetical method such as the contingent valuation method. This introduced a number of problems concerning the choices made by the researcher about a variety of methodological issues. By comparing an *indirect* observable model to an *indirect* hypothetical model many of these problems can be avoided because both methods utilize the same set of methodological choices and assumptions. This minimizes the problem referred to by Randall (1994) as the comparison of apples and oranges which plagued previous comparison studies of public goods. I have found no other public goods study that compares two indirect methods. This feature will be discussed in greater detail in chapter two.

Other aspects that add to the importance of this empirical research are its' application to a very familiar environmental good and the use of a hypothetical travel cost question format. By utilizing such a familiar public good this study presents a best-case scenario for hypothetical responses to accurately reflect real preferences.²¹

²¹The public good being addressed is deer hunting sites across the United States. The sample includes only responses from individuals who are currently deer hunters. Hunters are generally very familiar with both the characteristics of hunting sites and the cost of travel to these sites. The question format and sample of respondents will be discussed further in chapters two and four.

If deer hunters who are intimately familiar with the deer hunting sites are not able to give accurate hypothetical responses, then we should be very skeptical that respondents would be able to provide accurate responses for less familiar environmental goods such as biodiversity preservation. Additionally, as will be discussed in more detail in chapter two, this hypothetical travel cost format should minimize any incentives for strategic behavior. Individuals responding to a hypothetical tradeoff between the trip cost and number of trips taken will be less likely engage in strategic behavior than if they were asked to directly state a maximum willingness to pay for the site in the form of higher license fees.

Currently the empirical evidence needed to draw conclusions about the comparability of hypothetical and actual responses, and to address the calibration issue are lacking. This dissertation will attempt to fulfill this need in a somewhat unique manner. By comparing similar hypothetical indirect and observable indirect methods, the methodological choices and assumptions will remain constant across the two which will minimize the problem of comparing apples and oranges. It also utilizes a very familiar public good which should reduce the potential for biased estimates. Finally, another advantage is obtained by using the hypothetical travel cost question format, which should minimize incentives for strategic behavior. This dissertation will make an important and unique contribution to the issues of comparability and calibration of hypothetical responses.

The use of non-market valuation techniques to derive the social value for environmental and natural resources has become increasingly important. Primary uses

of these non-market valuation techniques have been for natural resource damage assessments and benefit-cost analyses. A number of methods have been developed in an attempt to value environmental goods. Although some of the hypothetical methods may be superior to the indirect observable methods for the reasons discussed above, using such methods is questionable due to the lack of empirical support for the assumption that hypothetical responses are directly comparable to real economic actions. Methods based on hypothetical questions may be biased for a number of reasons which will be discussed in detail in chapter two.

Two types of tests have attempted to address this hypothetical versus actual response issue. These tests have been experiments with private goods and experiments with public goods. Although recently a number of valuable experiments with private goods have been performed, (Neill et al. 1994; Cummings et al. 1994) tests with public goods are also necessary. This dissertation is aimed at filling this need. It addresses the hypothetical versus actual response issue empirically for the case of a public good. Although this dissertation will not resolve the issue it will provide another important piece to a very complex puzzle.

Chapter Two

Hypothetical versus Actual Responses

In chapter 1 strong arguments were cited for the use of hypothetical valuation methods if an accurate revelation of preferences can be obtained. Therefore, an important question to answer is whether valuation methods that utilize hypothetical data can produce unbiased benefit estimates which would accurately reveal preferences. This chapter first addresses the theoretical reasons that hypothetical and actual responses may differ and then reviews a number of studies that have sought to investigate these potential sources of bias. Finally, the specific contribution of this dissertation to the hypothetical versus actual response literature will be discussed.

A framework for classifying the reasons that hypothetical and actual responses differ and a detailed discussions of the possible sources of systematic bias that may affect a hypothetical response valuation method are presented below. This framework is presented in order to build a foundation for the development of a testable hypothesis. The detailed discussion of potential sources of bias is necessary as a source of reference which will make potential explanations of this study's findings much clearer. Additionally, there is a great deal of confusion among natural resource economists concerning the definitions of some of the biases and some disagreement as to what actually constitutes a source of bias (ex. hypothetical bias). The detailed discussion of biases then will clarify what is meant by some of the conclusions presented in chapter six and minimize questions which might be semantic in nature.

Theoretical questions concerning the ability of the hypothetical methods to obtain accurate information about respondent preferences have been classified in a number of ways. One of the most recent and appealing discussions of the concepts of reliability, validity, and sampling error has been presented by Myrick Freeman (1993). The following framework for categorizing the reasons that hypothetical and actual responses may differ is adapted from Freeman's presentation (1993). The current adaptation presented here reaches a significant conclusion that differs with Freeman's work. The current analysis presents a completely different perspective on hypothetical bias.

Let Ra_i be the i th individual's response to a real economic situation, where this individual will have to accept the real consequences of her choice. Let Rh_i be the i th individual's response to a hypothetical economic market, which attempts to mimic a real market that would produce Ra_i , but where there are no real economic consequences of her action.

Ra_i can differ from Rh_i for three reasons. The first reason is a random error process (P_1), with a mean of zero, affecting Rh_i . The second is a systematic error process (P_2) affecting Rh_i . The third reason (P_3) is related to the probability that Rh_i will actually be observed. Combining these three processes we have $Rh_i = g\{Ra_i, P_1, P_2, P_3\}$, where $g\{.\}$ aggregates the three sources of divergence between actual and hypothetical responses.

P_1 is referred to by Freeman (1993) as the random error process. If the only source of divergence between Ra_i and Rh_i is this random error process P_1 then there

is no bias in the measurement of Ra_i and the survey instrument is valid. The only divergence between the means of Ra_i and Rh_i would fall under the title of reliability problems and not a lack of validity.^{22 23}

P_2 represents many types of systematic bias. These include strategic bias, hypothetical bias, embedding, and other sources of bias. Many of these sources of bias may not pertain to the present study, but each will be discussed later in this chapter. Also, it is important to address the controversial concept of "hypothetical bias". This controversy will be discussed within the context of the taxonomy of biases introduced below.

P_3 represents such problems as sampling bias, and item nonresponse bias. Biases resulting from this difference are not the primary focus of this dissertation. Therefore, it will not require a detailed discussion.

Potential Sources of Bias in Hypothetical Responses

A taxonomy of potential biases (P_2) can be developed in a number of ways. The technique used here is to group biases around the major criticisms of hypothetical valuation methods. Steven Shavell (1993) lists several criticisms of the contingent

²²An instrument is deemed reliable if it is consistent across trials of the same data sample. An instrument is valid if it measures what it is supposed to measure (Kirk, 1990).

²³The reliability of the sample mean can be increased by either increasing the sample size or improving the sample design (Freeman, 1993).

valuation method.²⁴ These criticisms include the following four claims: that individuals may have motives to misrepresent their opinions; that they may have poor incentives to answer questions carefully; that their answers may reflect something other than a valuation; and that they may not have an adequate understanding of the good being valued.

These criticisms provide a coherent way of classifying potential sources of bias. A systematic taxonomy of the major biases that may affect hypothetical response valuation methods is presented below in Figure 2-1. While this taxonomy of biases centers around the behavior of the individual, it should be understood that these biases may be heightened or lessened by the questionnaire construction. Also this taxonomy presents the sources of bias that researchers currently find most relevant to hypothetical methods in general, and the contingent valuation method in particular. As further research is conducted this taxonomy of biases will change. These potential sources of bias are discussed in detail in the pages that follow Figure 2-1.

²⁴Although Shavell's (1993) criticisms were directed specifically at the contingent valuation method many of them are also applicable to other hypothetical valuation methods.

1. *Motives to misrepresent opinions:* Individuals may misrepresent their assessments if they believe that benefits (economic or psychic) can be obtained by supplying answers that are different from their actual evaluations.

A. *Strategic bias* arises when respondents deliberately attempt to alter either the provision level of a good, or the amount paid for a good, by intentionally distorting their answers.

B. *Compliance bias* transpires when respondents attempt to gain favor in the eyes of either the interviewer or the survey sponsor.

2. *Individuals may lack incentives to answer carefully:* *Hypothetical bias* develops when respondents fail to adequately search their preferences because it may be difficult to make good decisions, and there are no penalties for making poor decisions, in hypothetical markets.

A. *Implied value cues* manifest themselves when the uncertainty associated with making complex transactions for which the respondents have little or no experience causes them to search for cues as to the good's "correct value".

B. *Budget or time constraint insensitivity* may also arise if a respondent is confronted with a complex hypothetical market purchase. In such a situation he may fail to adequately consider his budget constraint, and therefore, tend to overstate his willingness to pay for the good. An individual might also fail to adequately consider his time constraints, and thereby, overestimate the frequency with which he would utilize a good. For example, an individual might overestimate the number of trips they would make to a beach or a park if they fail to anticipate career or family related obligations.

3. *Individuals may supply answers that reflect factors other than their valuations:*

A. *Insensitivity to Scope* occurs when responses are insensitive to the size or amount of the environmental good being valued.

1. The *embedding effect* materializes when willingness to pay for a good decreases when individuals are first asked to value a more inclusive good. For example, increased air quality in Atlanta, GA and the entire Southeastern United States.

2. A *nibble effect* arises when willingness to pay for a combined good (A+B) is less than the summation of the values for good A only, and for good B assuming A is already provided.

3. A *quantity effect* occurs when respondents are insensitive to numerical quantity measurements of a good such as small reductions in risk.

B. *Symbolic bias* arises when valuation responses represent an environmental good's general symbolic meaning, or respondent's feelings about the valuation context, instead of the good's value itself.

Figure 2-1 Taxonomy of Potential Biases

4. *Individuals sometimes do not have adequate understanding of what they are being asked to evaluate:* Individuals may need a great deal of information and time for consideration in order to properly respond to a complicated hypothetical market.

A. *Time to think* before responding may affect the individual's hypothetical response. This may be troublesome because most hypothetical survey methods require on the spot evaluations.

B. *Information* given to individuals will affect responses. The important consideration is deciding upon the correct type and amount of information to provide.

Adapted from Shavell (1990).

Motives to Misrepresent Opinions

Economists have traditionally been very concerned with motivations for respondents to intentionally misrepresent their willingness to pay for a public good. Such actions are referred to as **strategic behavior**, which within the context of contingent valuation is the deliberate attempt by respondents to influence the provision of a public good by intentionally distorting the amount they say they are willing to pay. When individuals acting strategically report willingness to pay values that do not represent real economic commitments a **strategic bias** is introduced.

Economists have defined two types of strategic behavior known as free riding and over pledging. **Free riding** occurs when an individual pays less for a public good than it is really worth to him in the expectation that others will pay enough to provide it. The free rider problem is the direct result of the non-excludability characteristic of public goods. Incentives to free ride will be present if individuals believe that they will actually have to pay the amount they state as their willingness to pay and that the good is likely to be provided regardless of their understatement of willingness to pay.

Over pledging is the act of stating a falsely inflated willingness to pay in order to increase the probability that the good will be provided (Mitchell and Carson 1989). Respondents have strong incentives to engage in over pledging if they believe that their stated willingness to pay may affect the provision of the good and that there

is no relationship between their stated willingness to pay and the amount they will actually be required to pay.

If an individual behaves strategically then he will only report his true willingness to pay if he believes that provision of the good is contingent upon his stated willingness to pay and that he will have to pay the full amount he offers. Given these two conditions it is in the persons self interest to state his true willingness to pay.^{25 26 27}

Freeman (1993) believes that strategic bias can only be introduced when two conditions hold. First, survey participants must have some reason to believe that the actual outcome will be affected by their responses. This first condition provides the incentive to act strategically. Second, individuals must have the ability to influence the outcome by misleading the interviewer. Freeman (1993) sees incentive

²⁵The absence of private markets makes the strategic behavior of truthfully reporting willingness to pay irrelevant to most contingent valuation studies.

²⁶Conversely, if the individual believes that there will be no relationship between the required payment and his stated willingness to pay, and if he believes that the provision of the good will not be affected by his stated willingness to pay then the optimal strategic behavior would be to minimize the effort involved in searching his preferences. The individual can optimize by simply stating the first willingness to pay amount that he thinks of.

²⁷Peter Bohm (1972) conducted a group of seminal experiments for the presence of overpledging and free riding. These experiments were conducted under a variety of conditions concerning the connection between the individual's response and both the payment requirement and the provision of the good.

compatible mechanisms, such as the direct referendum question format, as a solution to strategic bias because a respondent's best strategy is to provide truthful answers.²⁸

Freeman's (1993) conclusion that the dichotomous choice question format provides respondents with incentives to answer hypothetical questions truthfully may be incorrect. Although it is reasonable to assume that the dichotomous choice framework is incentive compatible for an actual choice decision involving real economic consequences, this assumption does not hold true for hypothetical choices. A flaw in the rationale on which the assumption of incentive compatibility of the dichotomous choice framework for hypothetical choices will be demonstrated with a simple counter example.

Suppose first that an individual has a real (not hypothetical) opportunity to purchase a good. Assume that she believes that both provision of the good and the price she must pay is dependent upon her answer to a referendum question. If her true willingness to pay for the good is \$25 and she is asked whether or not she is willing to pay \$75 for the good, then it is in her best interest to truthfully answer "no".

Now consider that she is asked a similar question in a hypothetical setting about a good for which she has a true willingness to pay of \$25. Assume that the individual believes that her response may affect the provision of the good but that her

²⁸The direct referendum question is often referred to as the dichotomous choice question method. This method requires a respondent to answer "yes" or "no" to a question that asks whether or not she would be willing to pay a specific amount for the environmental good of interest.

cost is not dependent upon her answer. If then she is faced with a question asking her whether or not she is willing to pay \$75 for the good, she has a typical "over pledging" incentive to act strategically and answer "yes" even though her true willingness to pay is only \$25. In this case, which may be typical of many contingent valuation questions, the respondent still has incentives to over pledge when faced with a dichotomous choice question.

Compliance bias is related to strategic bias in that both involve the intentional misrepresentation of preferences by the respondent. The primary difference between the two is the goal which motivates the false responses. Strategic bias is an attempt to influence either the provision of, or the individual's payment for, an environmental good. Compliance bias on the other hand is an attempt to gain favor in the eyes of either the interviewer or the survey sponsor (Mitchell and Carson 1989).

In-person interviews, and to a lesser degree telephone interviews, are potentially susceptible to respondents attempting to please the interviewer. As the degree of familiarity with the valuation question decreases participants become more likely to look to the interviewer for assurances that their responses are "normal".

Compliance bias may also be introduced by the sponsor of the survey. If, for example, Ducks Unlimited made it known that they were the sponsor for a survey in which respondents were asked to report their willingness to pay for waterfowl or wetlands protection this might pressure respondents to report high values for such goods. The potential for such a compliance bias is one reason that surveys conducted

by reputable and neutral survey organizations are given more credence than those conducted by interest groups such as environmental groups or corporations.

The literature investigating compliance bias is almost nonexistent when compared to the number of studies investigating strategic bias. The few studies that have looked into compliance bias indicate that compliance bias may not be a serious problem.²⁹

Lack of Incentives to Answer Carefully

Expressions of behavior in hypothetical markets are costless because these hypothetical responses are not associated with real economic consequences. Hypothetical mistakes are costless, therefore, choices made in hypothetical markets may not be considered as carefully as real market decisions. Hypothetical markets for environmental goods may also present respondents with unfamiliar goods for which they may have difficulty defining their preferences. The combination of the difficulty of searching preferences for these goods and the lack of incentives to do so carefully in hypothetical markets may introduce a systematic error into survey responses. This form of systematic error is referred to as **hypothetical bias**.

²⁹ Studies that have investigated some form of compliance bias have usually done so only indirectly. For example a valuation study by Desvougues, Smith, and McGivney (1983) investigated interviewer effects. They found no evidence that interviewer effects bias valuation results.

Hypothetical bias is a controversial topic among environmental economists. Two basic positions have developed on this subject. One group of economists believes that hypothetical bias represents a systematic source of error. This group's position has been described by Bishop and Heberlein (1990, 92) who state that hypothetical bias, "stems from the inability of respondents to accurately predict how they would behave if a market were created."

Another group of environmental economists do not accept that the hypothetical nature of survey methods represents a systematic source of error. Both Mitchell and Carson (1989), and Freeman (1993) take this position. Mitchell and Carson (1989, 216) have said, "...the concept of 'hypothetical bias' which has been defined as 'the potential error induced by not confronting any individual with the real situation' is a misnomer. The only unique effect of a scenario's lack of realism is not bias, but random, directionless error."

Freeman (1993, 168) agrees with Mitchell and Carson's (1989) position and discounts the legitimacy of the hypothetical bias concept by saying, "The process captured by P_1 reflects what I had termed the problem of accuracy in my earlier book and what some others have called hypothetical 'bias'. In the framework presented here this would not be called a bias, since by assumption the random inaccuracies in responses have a zero mean."³⁰

³⁰I have substituted the term P_1 for Freeman's term $f_1(.)$. This substitution was made in order to preserve consistency with my adaptations of Freeman's classification framework and it in no way alters the substance of his objections to the concept of hypothetical bias.

Both Freeman's (1993) and Mitchell and Carson's (1989) positions rest on the assumption that the error attributable to a survey's hypothetical nature is random and has a zero mean. In effect they are arguing that if all sources of deviation between Ra_i and Rh_i were eliminated except for this hypothetical bias then there would be no systematic bias in the hypothetical responses. In this case $E(Ra_i - Rh_i) = 0$.

If their position is correct, then in the absence of any other sources of systematic bias, we should observe approximately an equal number of individual hypothetical responses that fall above and below the real market response in any particular study. In reality a number of studies seem to be systematically biased upward or downward. The important point is that although the direction of the bias may change from study to study, the error within a particular study seems to be systematically biased upward or downward. Given that in actual experience it has been observed that $(Ra_i - Rh_i)$ is not equal to zero within individual studies, then we must conclude that either the studies which report such results are exhibiting a hypothetical bias, or they must be plagued by some other type of systematic bias. An example of such a study is Neill et al. (1994) which will be discussed below in the review of private and public goods experiments.

The table which appears above presents a systematic hypothetical bias, but how can methods that seem to overestimate benefits in some studies and underestimate benefits in others be subject to a systematic hypothetical bias? The answer is that hypothetical bias can take a number of forms depending upon the type of survey question asked.

Hypothetical questions are by their very nature less realistic than actual situations. This lack of realism may make it much more difficult for the respondent to search his preferences for an answer. This lack of realism may be especially pronounced for contingent valuation questions in which the individual is asked to value a public good for which he has no experience in making monetary tradeoff decisions (ex. biodiversity, cleaner air). Such an individual, who is confused by a hypothetical valuation question, may look to the interviewer or to the survey questions for clues as to the good's "correct" value. These clues are referred to as **implied value cues**. The type of hypothetical biases induced by various types of implied value cues include starting point bias, range bias relational bias, importance bias, and position bias. (Mitchell and Carson 1989)

The implied value cue which has received the most attention in the contingent valuation literature is the starting point. **Starting point bias** arises when the respondent's willingness to pay amount is influenced by a value introduced in the survey question. The bidding game format is most susceptible to starting point bias. The bidding format asks an individual if they would be willing to pay some specific dollar amount for the good. If the response is affirmative then the individual is asked about higher dollar amounts until a negative response is given.³¹ At this point an iterative bidding process may more precisely identify the individuals maximum willingness to pay.

³¹Conversely, if the initial bid is rejected then lower willingness to pay amounts are offered until a positive response is received.

Faced with the difficulty of a hypothetical situation characterized by a high degree of uncertainty regarding the good's value, an individual may perceive the initial bid as conveying the approximately correct value for it. Several studies have found that relatively high initial bid starting points tend to bias willingness to pay estimates upward, and low initial bids induce a downward bias (Cummings, Brookshire, and Schulze 1986; Rowe, d'Arge, and Brookshire 1980).

Another form of contingent valuation question is the payment card technique. When asked to value a good the individual is shown a card with a large number of potential values from which to choose. A **range bias** is introduced if the amounts presented on the payment card affect the willingness to pay amount chosen. This may occur if the respondent considers the payment card as reflecting the researchers range of expectations about "normal" valuation behavior.

Relational bias occurs when a relationship between the good being valued and some other public good implies some type of value not intended by the researcher. If benchmarks, such as the cost of other public goods, appear on a payment card to remind respondents that they are already paying for other public goods this may cause an obvious relational bias. A more serious threat of relational bias might occur in connection with recreational activities such as taking a hunting trip or visiting a park. Existing hunting and fishing license fees and entrance fees to parks may be taken by respondents as representing a fair value for the environmental good.

Another type of hypothetical bias, **importance bias**, occurs simply because an individual is asked to participate in the survey. It is reasonable for the respondent to

assume that the interviewer would not be conducting a complicated and expensive survey for a good that was not important or valuable. Researchers should take care to make respondents feel comfortable with indicating that they do not view the good as being valuable. Researchers should also avoid using any language or symbols which may inappropriately convey value.

The final type of hypothetical bias associated with implied value cues is **position bias**. If survey participants are asked a series of valuation questions they may assume that the order of these questions implies a relative value. This question sequencing is often used to value different levels of a public good such as variations in risk reduction levels.

Each of these types of hypothetical bias are born out of respondent uncertainty. If the respondent is faced with a question about a good for which he is unfamiliar and does not have well defined preferences he may have great difficulty arriving at a valuation estimate. Given this difficulty he may look for clues as to a proper value for the good.

If a respondent does not have any easily available implied value cues, as for example might occur in a carefully worded contingent valuation question, hypothetical bias will probably take the form of an overestimation of willingness to pay. One possible explanation of this form of behavior is respondents' **insensitivity to their budget or time constraints**. Because the markets are hypothetical, and therefore purchases are costless, respondents may not carefully consider the hypothetical

tradeoff. In particular, they may fail to give adequate consideration to either their budget or time constraints.

An individual who does not give adequate consideration to his budget constraint would be expected to overestimate his willingness to pay for a good, or for a given purchase price, overestimate the quantity of the good he will purchase. So in the absence of any implied value cues we might expect an overestimation of willingness to pay. The results of the Neill et al. (1994) study seem to be consistent with this hypothesis. Neill et al. (1994, 153) provide experimental evidence that suggests that, "... hypothetical willingness to pay is consistently and significantly higher than the willingness to pay that reflects real economic commitments."

A respondent who does not give adequate consideration to her time constraint would be expected to overestimate the frequency with which she would utilize an environmental good. If, for example, she is asked how many trips to a scenic park or beach she would take, given some hypothetical level of environmental quality or travel cost, and she fails to fully consider or anticipate time constraints such as family or career obligations, then she would be expected to overestimate the number of trips she would actually take. In the context of a hypothetical travel cost model this type of mistake would lead to an overestimation of the respondents willingness to pay for the environmental good.

Each of these forms of hypothetical bias is attributable to differences between hypothetical and real markets. Specifically, the condition that mistakes are costless may lead respondents to give answers that are not carefully considered. If implied

value cues are present they may bias estimates either upward or downward, but in the absence of such cues the hypothetical nature of constructed markets should induce an upward bias to responses.

Responses that Reflect Factors Other than Values

Insensitivity to scope includes three types of biases. This scope insensitivity terminology has been adapted by Jonathan Baron (1994) to include the well know embedding effect, as well as a nibble effect, and a quantity effect. In many past studies this embedding effect has been defined so as to refer to each of the three effects discussed by Baron.

In the **embedding effect** willingness to pay for a good depends on the context in which it is embedded. Willingness to pay to clean up one lake is smaller if respondents are first asked their willingness to pay to clean up all lakes in a region (Baron 1994). The **nibble effect** occurs when respondents are asked to state their willingness to pay for two goods, such as individual lakes, first as a sequence of goods and then as one combined good. First they are asked their willingness to pay for good A alone, then good B assuming good A has already been provided, and finally goods A and B in combination. If a nibbling effect bias is present the willingness to pay for the combined good will be less than the summation of the values stated for the individual goods, and in the extreme case will be no greater than the willingness to pay for either good A or B alone. Baron (1994) proposes a third

type of insensitivity to scope bias which he calls the quantity effect. He defines the quantity effect as occurring when respondents are insensitive to numerical quantity measurements of a good such as the amount of risk reduced by a safety device.

The insensitivity to scope problem is unlike many of the other biases discussed because it has been observed in a non-hypothetical experimental format. Hayes, Shogren, and Kliebenstein (1993) conducted such an experiment that tests willingness to pay for food safety. Participants were asked their willingness to pay for a reduction in risk from each of five food borne pathogens and from all five combined. This laboratory experiment involved real money and a real good. Hayes et al. (1993) found no significant difference between the average willingness to pay for the five separate sources of risk reduction and the willingness to pay for the combined reduction. This separates insensitivity to scope problems from those that arise due to the hypothetical nature of some valuation methods.

Consider an insensitivity scope study carried out by Desvouges, Johnson, Dunford, Hudson, Wilson, and Boyle (1993) in which respondents are asked their willingness to pay to save a certain percentage of a bird population with a program to put covers on waste oil ponds that would otherwise entrap birds. In this study the stated willingness to pay for a program that saved 2% of the birds was not significantly different from the willingness to pay for a program that saved only 1% of the birds. These percentages refer to programs that increase the population size from 98% to 100% and from 99% to 100%.

Economic theory suggest that the willingness to pay for the 2% change should be at least twice as large as the willingness to pay for a 1% increase.³² Survey participants however reported willingness to pay values that were basically the same for each program. Such behavior seems irrational from an economic perspective.

How might behavior of this nature be explained? Three suggestions for respondent behavior may explain these results. The first, a combination of existence values and question misperception is suggested by Fisher (1994). The second, the budget account constraint concept is proposed by Baron (1994) and Thaler (1993). The third is the insensitivity to numerical quantity measurements defined above.

Survey participants who value the existence of a species may have misinterpreted the Desvougues et al. (1993) survey question to mean that the programs would just save 1% to 2% of the total population of birds. Additionally, they may have believed that 1% of the present bird population would be large enough to be sustainable in the sense of avoiding extinction. These respondents then could maintain the species by funding the program that saved just 1% of the total population. If the respondents valued only the existence of the species and placed zero value on the survival of additional birds, beyond a minimum sustainable population, then we should expect the willingness to pay for each of the two scenarios

³²If the program that increases the population from 98% to 100% is called X_2 and the increase from 99% to 100% is called X_1 , then the willingness to pay for X_2 should be greater than or equal to two times the willingness to pay of X_1 . This occurs because the concept of diminishing marginal utility implies that individuals should receive less utility from saving the last one percent of birds than the percentage that precedes it. Therefore, the increase in willingness to pay for 2% of birds saved should be greater than a proportional doubling.

described by Desvougues et al. (1993) to be equal. Existence values then may explain some apparent scope insensitivity findings. However, regarding the Desvougues et al. (1993) study it may be unreasonable to believe that the majority of the respondents misinterpreted the survey question. To further explain these results it is necessary to consider a second explanation of respondent behavior, the concept of the budget constraint account.

The concept of a budget constraint account suggests that people think of money in separate accounts. Each account can be applied to certain goods or types of goods. Therefore, individuals may set a limit to the amount they are willing to spend on environmental goods or public goods in general. When asked their willingness to pay for an environmental good they may state this amount. If they are then asked to value a more inclusive environmental good we would expect their stated willingness to pay to remain unchanged because they have already exhausted their environmental funds account. If the budget constraint account hypothesis accurately reflected respondents behavior it would explain the Desvougues et al. (1993) findings. It appears that a combination of the of the budget constraint account hypothesis and the presence of existence values can explain a significant portion of these findings, but that a non-trivial amount of respondent behavior probably represents an insensitivity to numerical quantity measurement. This numerical quantity insensitivity may arise if individuals do not perceive a difference in 98% and 99% of the bird population. That is, some individuals may perceive the survival of these quantities of birds as virtually

the same environmental good, and therefore, should not be expected to reveal different willingness to pay amounts for the two programs.

Baron (1994) believes that if budget constraint accounts reflect respondents' behavior then scope insensitivity would only be found in expressions of willingness to pay, not expressions of willingness to accept. This possibility is currently being tested by Baron.

If either budget accounts or existence values accurately reflect respondents' behavior, then scope insensitivity does not represent a bias but rather actual economic behavior. In this case then the interpretation that stated preference methods are flawed if they yield results that imply "more is not preferred to less" is incorrect. It appears that current results on insensitivity to scope problems may be inconclusive. Further examination of this issue is needed before any definitive conclusions can be drawn concerning the seriousness of these potential biases for stated preference approaches.

Symbolic bias is another important potential source of systematic error. Symbolic bias occurs when survey participants mistakenly value a good's general symbolic meaning instead of the specific good intended by the researcher (Mitchell and Carson 1989). Environmental goods may be especially susceptible to symbolic bias due to the relatively high level of emotions many people have concerning environmental issues and the difficulty inherent in attempting to define some types of environmental goods.

Consider the issue of nuclear power. If respondents are asked to state their willingness to pay for specific reductions in risk from a nuclear power plant they may be unable to separate their emotional feelings about nuclear power from the valuation exercise. The emotionally threatening nature of nuclear power may cause them to state falsely inflated willingness to pay values for risk reductions, which reflect their overall negative feelings about nuclear power.

Positive feelings can also bias responses. Suppose a survey asks individuals their willingness to pay to preserve a particular species of minnow from extinction. Some individuals might associate this action with protecting the environment in general, which they believe is a worthy goal. Their responses then might reflect general positive feelings about protecting the environment and not be closely tied to their desire to protect the specific minnow. Steven Shavell (1993) states that such results may reflect a "warm glow effect" where individuals who supply a positive answer experience a feeling of virtue, similar to that derived from giving to charity.

In addition to emotionally charged issues, symbolic bias may also be a problem when valuing goods that are less familiar to respondents. Goods with significant nonuse values may invoke a particularly strong propensity for individuals to respond to a symbol rather than the actual good (Cummings and Harrison 1992). As with many other types of bias issues properly defining the environmental good in a way that is clearly understood by respondents is the best course of action for reducing symbolic bias.

Hypothetical market responses can also represent feelings concerning aspects of the valuation context, other than the good itself, such as the choice of a **payment vehicle**. The choice of a payment vehicle can influence responses to valuation questions (Brookshire, Randall, and Stoll 1980). Payment vehicles used in hypothetical valuation studies include entrance fees, increased travel costs, license fees, utility rates and a variety of taxes. If respondents are asked their willingness to pay higher taxes for an improvement in an environmental amenity their responses may reflect opposition to the tax as opposed to their preferences regarding the amenity. If however, the researcher understands the impact of the payment vehicle then no real bias exist. The researcher may want to evaluate an overall policy that includes increased taxes. The choice of a payment vehicle is a simple but important decision to be made by the researcher. Whenever possible most researchers attempt to use a relatively neutral payment vehicle such as increased utility bills.

Inadequate Understanding

In addition to the three major criticisms of hypothetical methods discussed above there is another major source of concern. This concern centers around the respondents' inadequate understanding of the good they are attempting to value. Specific issues of concern include the amount of information provided to respondents and the time period they are allowed for answering. Although these potential

problems have not received the amount of attention that those previously discussed have, they still deserve the researcher's attention.

The "amount of time" that subjects have to think about their responses is an issue that is just emerging. Cummings and Harrison (1992, 69) suggest that willingness to pay responses "may be sensitive to the amount of time that subjects are given to formulate their values". Most stated preference methods require "on-the-spot" valuations by respondents which may not give them sufficient time to search their preferences. Not enough evidence is currently available to clarify the effects of varying the amount of "time to think". This leaves the researcher with very little guidance as to the correct course of action in this area.

Another emerging issue is that changes in the amount and type of "information" given to subjects will affect their responses. Bergstrom, Stoll, and Randall (1990) tested the effects of providing "service information", which is information concerning the uses of a commodity, on valuation responses. The authors used a mail survey to elicit values for three scenarios related to wetlands protection along the Louisiana coast. The initial scenarios concentrated on the effects of wetlands protection on fish catch levels and hunting bag levels. One set of survey participants was given service information not given to the other set. This service information concerned other positive services provided by the wetlands, such as scenery improvement. Bergstrom et al. (1990) found that the provision of service information significantly increases respondents' valuation estimates.

How should results of this type be interpreted? A reasonable supposition is that changes in the information provided regarding a good should cause valuation responses to change. New information, after all, has the effect of presenting respondents with a new good to value. The issue for researchers then is how much and what type of information should be provided. The correct information set should include a complete description of the change in the environmental good and all of the benefits and costs associated with this change that can reasonably be presented during the course of the survey.³³

The preceding pages of this chapter have identified several reasons that responses to hypothetical questions of actual behavior may differ. A number of potential sources of bias have been identified and placed into the following four categories: motives to misrepresent opinions; lack of incentives to answer carefully; answers that reflect something other than a valuation; and inadequate understanding of the good being valued.

Review of the Hypothetical versus Actual Response Literature

Differences between hypothetical and actual responses have been investigated in several ways. In the following pages the most pertinent of this literature will be

³³Costs in this situation refer to the consequences of the change in environmental quality for which the respondent has a negative valuation. In the wetlands preservation example discussed above this might include increases in mosquito or poisonous snake populations.

reviewed. First, a discussion of the literature developed by social psychologists which inquires into the ability of attitudes to predict actual behavior will be presented. Next, the work conducted primarily by economists which investigates the comparability of hypothetical and actual responses within the context of environmental valuation will be reviewed. This body of research will be divided into studies that are applied to private goods and those pertaining to public goods. This review of the literature will serve to provide a historical perspective as to how and why this hypothetical versus actual response debate has arisen. More importantly this review of the literature will highlight the specific contributions that this dissertation will make to the current body of knowledge.

Attitudes and Behavior

The field of social psychology has a history of investigating the relationship between attitudes and behaviors. Although the research into attitudes and attitude change has been extensive only a limited review pertaining specifically to the history of the ability of attitudes to predict behavior will be presented here.³⁴

One of the earliest and best know studies in the field of social psychology was conducted by LaPiere in 1934. LaPiere spent several summers traveling across the United States with a married couple both of Chinese ancestry. While traveling they

³⁴For a comprehensive review of the research concerning attitudes and attitude change refer to Olson and Zanna (1993).

attempted to eat at various restaurants and obtain lodging at hotels. LaPiere, a caucasian university professor, and the Chinese couple were accepted in all but one instance. Six months after the trip LaPiere sent questionnaires to the same establishments asking if they would accept Chinese patrons as guests. Ninety-two percent of the establishments reported that they would not accept members of the Chinese race as guests. LaPiere's (1934) rather compelling results seemed to indicate that hypothetical responses from surveys cannot predict actual behavior.

The LaPiere study may be flawed because the hypothetical guests were not described as a *college educated* Chinese couple that would be *accompanied by a caucasian university professor*. Had this additional information been included it is possible that the hypothetical rejection rate would have been lower.

Since the 1934 study by LaPiere a large number of studies have attempted to predict individuals' behavior based on prior survey measures of their attitudes. Most of this research has shown a positive correlation between stated attitudes and subsequent behavior. Schuman and Johnson (1976) in a review of this literature concluded that this positive correlation was strong enough to indicate that a causal relationship between attitudes and behavior exists.

The most widely accepted model of the relationship between attitudes and behavior is the theory of reasoned action. This model was developed by Fishbein and Ajzen (1975). This theory proposes that both attitudes and subjective norms combine to determine behavioral intentions which in turn cause volitional behaviors (Olson and Zanna 1993). Although there have been a number of criticisms and proposed

alternatives to the theory of reasoned action it has remained the standard in the field and in several recent applications has been used to successfully predict behavior.³⁵

Fishbein and Ajzen's (1975) theory can serve economists who seek to value environmental goods using hypothetical methods in two ways. First, it provides a framework for understanding the relationship between attitudes and behavioral intentions, and the connection between behavioral intentions and actual behavior. This is important because hypothetical valuation studies attempt to measure behavioral intentions. The second contribution of Fishbein and Ajzen's work is the proposal of three factors that affect the ability of surveys to predict behavior.

The three factors proposed by Fishbein and Ajzen are correspondence, proximity, and familiarity. Correspondence is defined as identity among the action, the target at which the action is directed, the context in which the action is performed, and the time when the action is performed. The hypothetical travel cost survey question used in this study, meets the correspondence criteria relatively well. The hypothetical travel cost survey accomplishes this by describing in detail the recreation trip and the costs of travel. As with the hypothetical travel cost model, most types of contingent valuation survey questions meet the correspondence criteria with their detailed description of a hypothetical market and their introduction of a payment obligation.

³⁵For an excellent review of these criticisms, alternative theories, and recent applications refer to Olson and Zanna 1993.

The second factor, proximity, is defined as the number of intervening stages in the reasoned action model (Fishbein and Ajzen 1975). The fewer the stages between a component and the behavior, the greater the predictive power that component is likely to have. This criteria proposes that intentions will provide a stronger measure of behavior than either attitudes or beliefs because of its closer proximity (assuming equal correspondence). The hypothetical travel cost question meets the proximity criteria by asking individuals about their behavioral intentions to take recreation trips. Contingent valuation questions also meet this proximity criteria relatively well as they seek to measure behavioral intentions to pay for an environmental good.

Finally, Fishbein and Ajzen (1975) hypothesize that the more familiar the behavior is to the respondents, the more likely their attitudes and intentions will be able to predict actual behavior. The familiarity criteria is the primary weakness of many contingent valuation studies. Individuals will be relatively unfamiliar with many environmental problems such as ozone depletion, global warming, acid rain, and biodiversity issues. In contrast, the hypothetical travel cost question used in this dissertation will meet the familiarity criteria extremely well. Therefore, this dissertation, *ceteris paribus*, will present a best case scenario for hypothetical data's ability to predict actual behavior. This sample includes only those who actually have taken deer hunting trips. These hunters are very familiar with the costs associated with taking a deer hunting trip as well as the recreational benefits of this activity.

The LaPiere study (1934) initially cast doubt on the ability of attitudes and intentions to predict behavior. The literature as a whole, however, is more positive.

In particular, Fishbein and Ajzen (1975) developed three criteria for evaluating the strength of the attitude-behavior relationship which were discussed above. Most hypothetical valuation studies meet the correspondence and proximity criteria relatively well. The degree to which the familiarity criteria is met, however, will depend upon the type of good being valued.

Although the attitude-behavior relationship literature is insightful, it cannot adequately answer the following question. Are hypothetical responses comparable to actual responses within the context of environmental valuation? A body of literature has developed which tests for differences between environmental values derived using actual and hypothetical data.

Tests of the equivalence between actual and hypothetical responses to valuation questions can be usefully divided into those that address public and private goods. This distinction is a necessary one, because individuals may react differently when asked hypothetical valuation questions concerning public versus private goods. This may occur because individuals may: be more familiar with private goods; be aware of the market prices for private goods; and have different principles concerning the valuation of public and private goods.

Comparability Tests Involving Private Goods

In this section a brief review of the private goods research literature will be presented. Although no single comparability study can be definitive considerable time

will be devoted to a pair of private goods experiments recently conducted by Neill et al. (1994). This study is representative of the best experimental research that attempts to determine whether or not hypothetical surveys accurately elicit real economic commitments. Their research tests the hypothesis that there is no significant difference between values reported by subjects when payment is real or hypothetical. Specifically, two types of experiments are conducted. The first experiment compares responses elicited using an open ended contingent valuation question to responses obtained utilizing a real Vickrey (second price) auction.³⁶ The Vickrey auction is real in the sense that the winner of the auction has to actually pay for the good and then gets to take it home. The second experiment is designed to determine whether it is the hypothetical nature, or the open-ended question format, of the contingent valuation method that might bias responses. This second experiment compares not only responses of an open-ended contingent valuation question to a Vickrey auction, but also compares these real Vickrey auction responses to a hypothetical Vickrey auction.

The private goods used for these tests were a watercolor painting by a Navajo Indian artist in the first experiment, and in the second experiment the private good was a framed print of a sixteenth century map of the world. The results and conclusions of the experiments involving each of the two goods are similar, and

³⁶A Vickrey auction takes sealed bids from participants. The winner of the good is the person bidding the highest amount, but the price that they pay is equal to the second highest bid. The dominant strategy is to bid truthfully because regardless of how others bid the individual can do no better than to bid truthfully (Neill et al. 1994).

therefore, only the results of the more comprehensive test involving the map will be presented here. The results of the map experiment are presented in Table 2-1 below.

Table 2-1. Results from the Map Experiments

Values Deleted (Truncation)	Sample Size	Mean (\$)	Median (\$)	Standard Deviation (\$)
	<i>Generic</i>	<i>Contingent</i>	<i>Valuation</i>	
None	48	329	75	844
≥ \$1000	44	109	65	127
≥ \$100	27	40	40	25
	<i>Hypothetical</i>	<i>Vickrey</i>	<i>Auction</i>	
None	51	301	60	591
≥ \$1000	45	110	50	153
≥ \$100	28	31	30	22
	<i>Real</i>	<i>Vickrey</i>	<i>Auction</i>	
None	60	12	5	21
≥ \$1000	60	12	5	21
≥ \$100	59	10	5	15

Source: Neill et al. 1994 "Hypothetical surveys and Economic Commitments". *Land Economics*. 70(2): 152.

The map experiment utilized three valuation institutions. These institutions were a real Vickrey auction, a hypothetical Vickrey auction, and an open ended contingent valuation question. The participants in this study were undergraduates at the University of New Mexico. The undergraduate subjects who participated in the real Vickrey auction were required to include a payment equal to their bid in an envelope. The winner of the auction would have the difference between their bid and the second place bid refunded if necessary. All other bidders would have their envelopes and payments returned to them. Participants in the hypothetical Vickrey auction were given similar questionnaires, but were told that the auction was

hypothetical and were not required to include payment with their bids. The open ended contingent valuation question participants were shown the map and asked, "If the map were to be made available for you to purchase here and now what is the maximum amount you would pay for it?" (Neill et al. 1994, 148).

Before reviewing the implications of this table some explanation of what each column represents is needed. The far left column titled "Values Deleted (Truncation)" represents the values which were deleted from the sample. The column just to the right of the "Values Deleted" column presents the various sample sizes. The value 48, which appears at the top of the sample size column, represents the sample size for the generic contingent valuation experiment when none of the responses are deleted. The Neill et al. (1994) study compares the actions of three different groups of individuals which each responded to one of the three valuation formats. This explains why the sample sizes of 48, 51, and 60, vary across the valuation methods even when none of the responses are deleted from the samples. The sample sizes presented also vary within a valuation method because the authors presented three types of comparisons. The first comparison includes all responses in the sample, while the second and third delete responses that were greater than or equal to \$1,000 and greater than or equal to \$100 respectively.

By simply reviewing the table it becomes apparent that there is a large difference between bids associated with real payments and bids associated with only hypothetical payments. Both hypothetical methods, the hypothetical Vickrey auction and the contingent valuation method, produced willingness to pay responses that

greatly exceeded those elicited using the real Vickrey auction. Both the mean and median responses elicited using the hypothetical methods were at least three times as high as were the responses to the real payment Vickrey auction. These results hold regardless of the treatment of extremely high bids.³⁷

Neill et al. (1994) employed the non-parametric Kolmogorov-Smirnov test to statistically analyze these results. Employing this test they found that both the hypothetical Vickrey auction results and the contingent valuation results were significantly higher than those based upon the real Vickrey auction. This finding holds at a significance level of less than 0.1%. They also found that the Kolmogorov-Smirnov test cannot reject the hypothesis that the hypothetical Vickrey auction's willingness to pay is equal to the contingent valuation method's willingness to pay.

Given these statistical results Neill et al. (1994) reached two conclusions. First, hypothetical willingness to pay is consistently and significantly higher than the willingness to pay amounts associated with real economic commitments. Secondly, they conclude that this difference is not due to question format, but is instead attributable solely to the differences between hypothetical and real payments.

Another private goods study that needs to be mentioned briefly is the often cited work of Dickie, Fisher, and Gerking (1987). These authors conducted an experiment involving the sale of fresh strawberries. Two samples of individuals were contacted at home. One was given the actual opportunity to purchase strawberries, where strawberries would actually be delivered and real payment was to be made by

³⁷Please refer to column one of Table 2-2 for the various treatment of outliers.

the consumer. The second sample was only given a hypothetical opportunity to state their maximum willingness to pay for strawberries. Dickie et al. (1987) found that the results from the actual and hypothetical strawberry sales were not statistically different. In a recent review of this study Smith (1993) stated that their findings were very sensitive to the treatment of outliers and interviewer effects. Smith (1993) noted that the conclusions reached were not incorrect but just very sensitive to the researchers judgement concerning outliers and interviewer effects.

The Neill et al. (1994) study provided empirical evidence that hypothetically stated willingness to pay amounts may exceed responses associated with real economic payments by a large amount. In contrast, the Dickie et al. (1987) strawberry study provided empirical evidence that hypothetical and actual willingness to pay amounts may not be significantly different. How can these apparently conflicting results be explained?

One reasonable explanation is that a larger number of respondents would be familiar with the current prevailing market price for strawberries than would be familiar with the current market price for a framed print of a sixteenth century map of the world. We might then expect respondents to state willingness to pay values that were very close to the well known market price for strawberries, if their responses represented what they believed was a fair value for the good, or how much they thought the strawberries normally cost. If respondents find it easier to provide willingness to pay amounts for private goods simply because they know the current market prices, a problem arises. This market cost may not reflect their own

personal valuation of the good. Therefore, the Dickie et al. results could represent a simple clustering of responses around a familiar market price, which would not be evidence of an accurate revelation of maximum willingness to pay amounts.

Individuals might also have been more familiar with the characteristics of strawberries than the characteristics of the framed print. Strawberries are consumed on a regular basis by a larger segment of the population than a framed print of a sixteenth century map. As was discussed above, the theory of reasoned action developed by Fishbein and Ajzen (1975) implies that respondents can more accurately state their preferences for a good when they are more familiar with the characteristics of the good. Therefore, it should be expected that respondents in the strawberry study could more accurately state their willingness to pay in a hypothetical setting because they were more familiar with the characteristics of strawberries than were the respondents in the map experiment.

Although the private goods studies reviewed above can provide insights into the ways in which individuals respond to hypothetical markets, there are several factors that may prevent the generalization of these private goods results to public goods. First, private goods are associated with the presence of market prices that may be widely known and may represent a type of fair price to the respondent. A related problem is that many private goods are more familiar to individuals than most environmental public goods.

Another reason that individuals may respond differently to hypothetical valuation surveys involving public goods are the different principles they may have

regarding private and public goods. This difference in principles may take two forms. First, some individuals find it unethical to think of the environment in terms of monetary values. These respondents may tend to give protest answers that reflect their moral conflict with the act of placing a monetary value on the environment.

Another possible effect of the difference in principles concerning private and public goods is the following. Individuals may possibly have a higher degree of civic mindedness when answering survey questions involving public goods. If this is the case they might be more likely to provide truthful answers to hypothetical questions because they are less likely to engage in strategic behavior. For each of the reasons discussed above individuals may respond differently to hypothetical valuation questions involving public and private goods. Given this strong possibility, conclusions regarding the ability of hypothetical valuation methods to accurately elicit willingness to pay amounts for public goods cannot be based on private goods studies alone. Rather, experiments involving public goods are also necessary.

Comparability Tests Involving Public Goods

One of the most noteworthy public goods studies was conducted by Smith, Desvouges, and Fisher (1986). This study focuses on changes in water quality in the Mongahela River Basin. The Smith et al. (1986) study attempts to *directly* test the comparability of individual consumer surplus estimates derived using the hypothetical contingent valuation method and an observable travel cost method. The travel cost

method, which is based on actual behavior, is seen by the authors as a criteria by which the hypothetical contingent valuation method can be judged.

This study estimates three types of travel cost models and four types of contingent valuation formats. The three travel cost models consist of two generalized models and one simple pooled model. These generalized models are a two step version of Vaughn and Russell's (1982) varying parameters model. In the first step a separate travel cost demand model is estimated for each of the 22 sites in their Mongahela River basin sample. In the second step each of the demand function's parameters are regressed on the site's characteristics. The only variation in the two generalized travel cost models is the type of estimator used in the first stage of estimation. One model uses ordinary least squares and the other uses maximum likelihood estimation. The third model estimated was a simple pooled travel cost model which estimated a single travel cost demand model utilizing pooled observations from all 22 river basin sites. The contingent valuation survey was designed to elicit estimates of use values, for a change in water quality, that would be *directly comparable* to the estimates derived using the travel cost models. Water quality was described to respondents using a water quality ladder which illustrated changes in water quality such as an improvement from boatable to fishable, fishable to swimmable, and swimmable to drinkable. Four approaches were used to ask the willingness to pay question. These included two iterative bidding questions with starting points of \$25 and \$125, an open ended question, and a direct question which used a payment card. The sample included 69 respondents.

As a first step in their comparison the authors presented average benefit estimates for each of the estimated models. They note that the contingent valuation estimates seem to be very sensitive to the question format used. The largest estimated contingent valuation mean was approximately six times as large as the smallest.

Next, the comparison of contingent valuation and travel cost estimates were made using individual data. First the hypothetical bids were regressed on each of the travel cost benefit estimates for the corresponding water quality change. If the travel cost and contingent valuation methods yield similar estimates then the intercept should be very close to zero and the slope approximately equal to one. These regression equations include variables for the type of contingent valuation question used. Inclusion of these qualitative (dummy) variables made it possible to examine the effect that each question format had on the relationship between the hypothetical and actual methods (Smith et al. 1986).

Both the hypotheses that the intercept was equal to zero, and that the slope was equal to one, were tested using a t-test. Except for the generalized travel cost model, and only then for a loss of area, the intercepts were not found to be statistically different from zero. However, the hypothesis that the slope is equal to one was rejected in all six cases for the generalized travel cost model. Based on the rejection of the slope hypothesis Smith et al. (1986) concluded that the generalized travel cost and contingent valuation models do not provide comparable estimates.

In contrast, the simple travel cost model comparison rejected neither the hypothesis that the intercept was equal to zero, nor the hypothesis that the slope was equal to one. Therefore, Smith et al. (1986) concluded that the contingent valuation and simple travel cost models provide benefit estimates that are comparable.

Although the Smith et al. (1986) study provided important insights into the comparability of hypothetical and actual responses, it was hampered by several weaknesses. The first deficiency of this study centers around the relatively weak power of its' statistical tests which occurs due to the comparatively small sample size ($N=69$). It should be noted that the intercept and slopes that were not found to be statistically different from zero and one, respectively, were quite large. For a water quality change from boatable to fishable, an intercept of -18.193 was not found to be statistically different from zero, and a slope of 8.411 was not found to be statistically different from one. This weakness is the result of a relatively small sample size for this type of comparative test (Mitchell and Carson, 1989). A major strength of this dissertation will be the ability to perform this type of test with a much larger sample size ($N > 1000$). The increased sample size will strengthen the power of the statistical test employed in this dissertation.

A second weakness of the Smith et al. (1986) research is its' method for measuring quality changes. This study compares changes in consumer surplus, resulting from a change in water quality, as estimated by the contingent valuation and travel cost methods. The problem is that the two valuation methods measure water quality in different ways. The travel cost model used dissolved oxygen levels as a

quality measure, while the contingent valuation model used the water quality ladder described above. Therefore, the quality changes measured in the two types of models may not have been equivalent. If this is the case, then we should expect the two models to provide different estimates of consumer surplus.

Finally the comparison of the two methods may have been less direct than the authors had intended. Although the authors were able to design a hypothetical method that captured only use values, and thereby increased the level of comparability, there is another problem that distinctly reduces the comparability of their test. This problem is the fundamental difference in methodology necessary for implementing their hypothetical method and their observable behavior method. This dissertation accomplishes one of the positive aspects of the Smith et al. (1986) research by comparing only use values to use values, but also goes an important step further by comparing an observable *indirect* method to a hypothetical *indirect* method. Since these two indirect travel cost methods are identical except for their reliance on hypothetical and actual behavior there will not be any differences in methodological assumptions and choices in this research.

Despite the shortcomings, two elements make the Smith et al. (1986) research extremely notable. First, is its' format of comparing benefit estimates for each of the 69 individuals. The estimates from each travel cost model were matched with the contingent valuation bids for each individual. Previous comparative studies had been restricted to comparing means for groups of individuals. The research by Smith et al.

(1986) was an important step forward for comparative studies due to its' introduction of analytical techniques based on individual level data.

The Smith et al. (1986) study is also notable for its' design of a contingent valuation question that would capture only use values. This is important because it made for a more direct comparison of a hypothetical and an actual method than in previous studies. Previous research had compared observable indirect methods that captured *only use* values to hypothetical direct methods that captured *both use and non-use* values. It should be expected that benefit estimates that capture both use and non-use values would differ from those that measure only use values, even if they both accurately measure consumer surplus.

Importance of Testing the Comparability of Actual and Hypothetical Responses in the Context of the Travel Cost Model

Research involving public and private goods, such as that described above, has aided in our understanding of how individuals respond to hypothetical valuation questions. However, additional empirical evidence is needed to address both the issues of comparability and calibration. This study will be an important addition to this growing body of research.

There are several aspects of this test which will make its' empirical results important and unique. This dissertation employs the seldom used **hypothetical travel cost method** to investigate the comparability issue in the context of a **public good**.³⁸

One of the strengths of this study is that the public good is very *familiar* to the survey participants. The public good is deer hunting sites, and the sample of respondents includes only individuals who were active hunters at the time of the survey. Active deer hunters are likely to be very familiar with the characteristics of the hunting areas, the cost associated with taking a hunting trip, and the recreational benefits derived from taking the trip.

The high degree of familiarity should present a best case scenario for respondents to accurately state their preferences regarding this public good. If respondents cannot accurately answer hypothetical valuation questions regarding such a familiar public good, then it is unlikely that they can answer accurately when asked about less familiar public goods.

Another meaningful feature of this test is its' large sample size. The sample includes over 1,000 hunters. This increase in sample size will strengthen the power of the statistical test. Relatively small sample sizes have plagued many previous studies and reduced the strength of their conclusions (Mitchell and Carson 1989).

³⁸Deer hunting sites are not a pure public good, but as was discussed in chapter 1 do have public good attributes. Some economists refer to this type of good as a quasi-public good. This study shall refer to deer hunting sites as simply a public good.

A unique facet of this dissertation is that it utilizes the rarely employed hypothetical travel cost method. Benefit estimates calculated using the hypothetical travel cost method will be compared to estimates derived using an actual behavior travel cost model. Since both the hypothetical and actual valuation methods are travel cost models, there will not be any *differences in assumptions and methodological choices* between the two methods. This is unique to this dissertation. Previous public goods studies have most often compared a direct hypothetical method with and indirect actual behavior method (Bishop and Heberlein 1990; Seller et al. 1985; Smith et al. 1986). If differences were found between the benefit estimates elicited using for example, a hypothetical contingent valuation method and an actual behavior travel cost model, the researcher could not determine whether such differences were due to the hypothetical nature of the responses or the methodological differences in the two valuation techniques.

The travel cost question format also minimizes the possible impacts of *strategic bias* and *insensitivity to scope*. As a result this test will primarily investigate hypothetical bias. The question format does not involve any embedding of the public good or hard to comprehend numerical quantity measurements. Therefore, differences in the hypothetical and actual valuation methods should not represent any insensitivity to scope problems. Additionally, hunters who are asked to state the number of trips they would take, given some hypothetical trip cost, are less likely to act strategically than if they were asked about their willingness to pay increased taxes or license fees. Respondents are unlikely to act strategically because they are not

liable to make a connection between the number of trips they report that they would take and the possibility of changes in either the provision level of hunting sites or their cost for using these sites. Since the impacts of insensitivity to scope and strategic bias should not cause the benefit estimates produced using the two methods to diverge, any differences should be attributable to the hypothetical nature of responses.

Another important aspect of my study is that although the good is familiar to respondents there are *not any well known "prices"* for their answers to cluster around. As was discussed in the Dickie et al. (1987) strawberry study, such a clustering of responses may not represent the accurate revelation of maximum willingness to pay amounts. This clustering effect is not limited to private goods. If the deer hunters had been asked about their maximum willingness to pay in the form of licensing fees, then their responses might have clustered around the fees they are currently required to pay. Such responses probably would not represent an accurate measure of consumer surplus.

There are also no easily available implied value cues. Therefore differences in responses are apt to represent a hypothetical bias resulting from respondents' failure to accurately consider their budget or time constraints. If, due to the costless nature of hypothetical purchases, respondents are insensitive to their budget or time constraints we would expect their hypothetical responses to overestimate their willingness to pay for the good. More specifically in the case of the hypothetical travel cost model used here we would expect that, at a given hypothetical trip cost,

respondents might overestimate the quantity of trips they would purchase. The Neill et al. (1994) map and painting results seem to be consistent with this hypothesis.

This research will extend the testing of this hypothesis to a public good.

In summary, there are a number of reasons that responses in hypothetical markets may differ from actual economic decisions. The primary reasons for this possible divergence are: motives to misrepresent opinions, a lack of incentives to answer carefully, answers that reflect something other than a valuation, and an inadequate understanding what being evaluated. Research that has investigated these sources of response error were grouped into the attitude-behavior relationship literature, private goods experiments, and public goods studies. This test involving a public good will make an important contribution to our understanding of a very complex issue.

Chapter Three

Theoretical Development of the Empirical Model

This chapter will achieve several important objectives. It will set forth a testable hypothesis for assessing the comparability of hypothetical responses to reports of actual behavior. It will also provide justification for several methodological choices which are crucial to the successful execution of this comparison. This chapter will also build a foundation for assessing both the legitimacy of using Marshallian Consumer Surplus for implementing the central comparison of responses, and the strengths and weakness of this dissertation relative to other studies in this area.

Building this foundation requires that this chapter begin by developing a basic framework for discussing the microeconomic foundations of the traditional travel cost method, and its' relationship to Marshallian consumer surplus. A thorough understanding of Marshallian consumer surplus, especially as it compares to the theoretically preferable Hicksian measure of welfare, is essential to fully appreciate the importance of using the former in this study.

Next, two theoretical advances made in the early 1980's will be introduced to further build the foundation for the arguments concerning the choice between the Hicksian and Marshallian measures of welfare. These two advances are the application of the household production function framework to the travel cost method, and the development of formulas which allow the Hicksian compensating measures of

consumer surplus to be directly calculated from an ordinary (observable) demand curve.

The Traditional Travel Cost Model and Marshallian Consumer Surplus

A framework for the discussion of the travel cost method can be specified simply. This framework will first be presented in the context of private goods only, then after using this scheme to derive measures of value for a private good the analysis will be extended to the case of a public environmental good.

Let an individual's utility function be represented as

$$u(x,z) \tag{3.1}$$

where z is some particular market good and x is a composite market good. Assume that the consumer maximizes her utility by choosing among the entire set of market goods, and in the process expends her entire income. This problem can be represented as

$$\max u(x,z) \tag{3.2a}$$

$$\text{s.t. } x + p \cdot z \leq Y \tag{3.2b}$$

$$x, z \geq 0$$

where p is the price of z , the price of the composite good x is normalized to one, and Y denotes the consumer's income. The consumer solves equation 3.2 and attains a maximum utility level, U^* , by consuming the optimal bundle of goods (x^*, z^*) . This constrained optimization yields the ordinary, or Marshallian demand function for z . The Marshallian demand for z , $z^* = m(p, Y)$, can be observed because it is a function of the observable variables p and Y .

Marshallian consumer surplus is the difference between what the consumer has to pay for a good, and the maximum amount that he would have been willing to pay for the good as revealed by the ordinary demand curve. Marshallian consumer surplus can be represented as the area below the ordinary demand curve and above the price line. Marshallian consumer surplus is equal to area ABC in Figure 3-1.

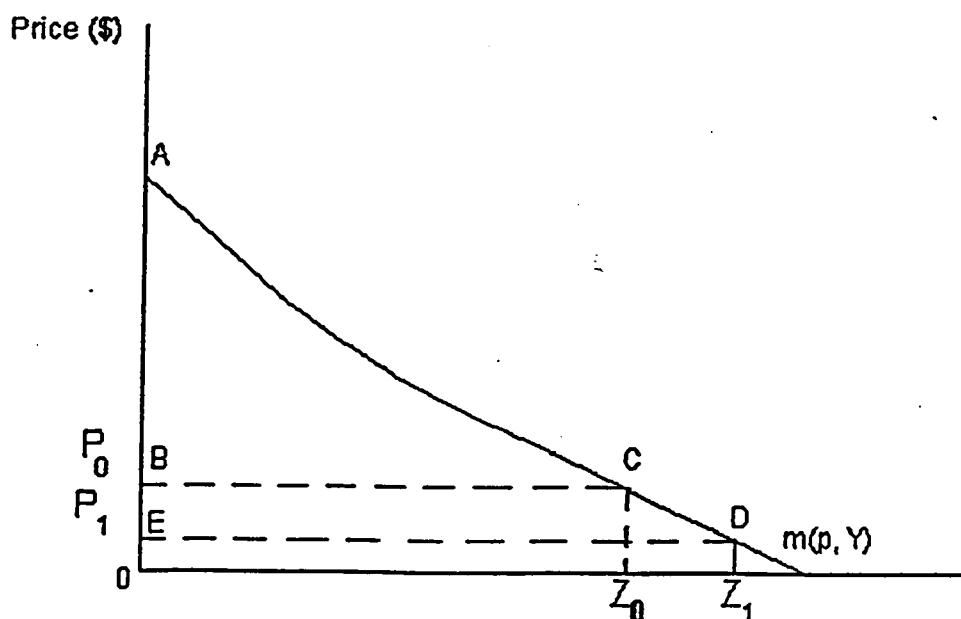


Figure 3-1 Graphical Representation of Marshallian Consumer Surplus

This area can be measured by taking the integral of the Marshallian demand curve from the origin to the quantity of z actually consumed (area ACZ_0O), and then subtracting the integral representing the area below the price line (area BCZ_0O).

Suppose the ordinary demand curve is estimated using a linear functional form such as equation 3.3,

$$z = \alpha + \beta p + \gamma w \quad (3.3)$$

where w represents socioeconomic variables such as income, age, and family size. In this case the integral representing a change in Marshallian consumer surplus, due to a decrease in price from P_0 to P_1 , is equal to equation 3.4, and is represented above in Figure 3-1 as area BCDE.

$$[(z_1)^2 - (z_0)^2] / 2\beta \quad (3.4)$$

where z_0 refers to the quantity of z actually consumed before a price change and z_1 is the quantity of z anticipated to be consumed after the price change. The formulas for the derivation of Marshallian consumer surplus associated with several common functional forms are presented in Table 3-1 below.

Table 3-1. Formulas for the Derivation of Marshallian Consumer Surplus

Functional Form	Marshallian Consumer Surplus
Linear: $z = \alpha + \beta p + \gamma Y$	$[(z_1)^2 - (z_0)^2] / 2\beta$
Semilog: $\ln z = \alpha + \beta p + \gamma Y$	$(z_1 - z_0) / \beta$
Log Linear: $\ln z = \alpha + \beta \ln p + \gamma \ln Y$	$(z_1 p_1 - z_0 p_0) / (\beta + 1)$
Quadratic $z = \alpha + \beta_1 p + \beta_2 p^2 + \gamma_1 Y + \gamma_2 Y^2$	$p_1 [2z_1 + (4\alpha + \beta_1 p_1)] / 6$ $- p_0 [2z_0 + (4\alpha + \beta_1 p_0)] / 6$

Source: Adapted from Bockstael, Hanemann, and Strand 1989. "Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models" report to the Economic Analysis Division, U.S. Environmental Protection Agency, Washington, DC; and from Ziemer, Musser, and Hill 1980. "Recreation Demand Equations: Functional Form and Consumer Surplus." American Journal of Agricultural Economics :136-141.

This introductory discussion demonstrated that for the *private marketed good* z , the information revealed in the ordinary demand curve can be used to calculate Marshallian consumer surplus. This dissertation however, is focusing on the ability of hypothetical methods, such as the hypothetical travel cost method, to accurately derive values for an *environmental public good*. Specifically, the environmental good being utilized in this dissertation is public hunting sites. The framework presented above which focused on the private good z , can easily be adapted to accommodate the environmental good of recreational hunting areas.

Unlike some other public goods the consumer has direct control over the number of times he utilizes the hunting area.³⁹ Therefore, to extend the analysis requires the simple replacement of the private good z , with an environmental good (hunting areas). In this situation involving a recreation site formulas 3.2a and 3.2b become

$$\max u(x,q) \quad (3.5a)$$

$$\text{s.t. } x + p \cdot q \leq Y \quad (3.5b)$$

$$x, q \geq 0$$

where q represents access to a hunting area, and p now refers to the price of access to the area.

The primary difficulty that environmental economists face is the absence of prices (p) for access to the hunting areas. Early travel cost models solved this problem by using travel cost as a proxy for the price of admission to the hunting area or park. This can be done because of the unique features of this type of public good. First, the benefits of entrance to the site are the same regardless of the distance an individual travels to a site. The second feature is that the cost of travel will vary across individuals that live different distances from the site because unlike other goods, recreational sites are immobile.

³⁹However, the number of times the consumer utilizes the hunting area may be constrained by hunting regulations which specify season length and other constraints on access and use.

By assuming that travel costs were a proxy for the price of the environmental good Marshallian consumer surplus could be estimated for the site by assuming that the demand for a private good, trips to the site, approximated demand for the environmental good which was the use of the recreation area. Economists reasoned intuitively that this would be a fairly close approximation and then based measures of value on the demand for the private good, but did so without a clear theoretical framework on which to base this procedure. This process of using the cost of travel to a recreation site as an implicit price for the environmental good continued until the early 1980's.

The Household Production Framework

In 1983 Bockstael and McConnell provided the needed framework in an influential article published in the American Economic Review. They presented a new theoretical foundation that justified the use of the travel cost method as an exact measurement of welfare based on the household production function framework. A review of their findings as well as the development of methods for calculating the Hicksian measures of welfare directly from ordinary demand curves is necessary because it will provide the foundation for assessing the merits of using estimates of Marshallian consumer surplus instead of the Hicksian measure of welfare commonly referred to as compensating variation.

Bockstael and McConnell (1983) supplied a theoretical basis for using the travel cost method to derive an exact measurement of welfare based on the household production function. The household production function framework is not a new approach to demand estimation, but rather a general model of individual decision making. This paradigm can be used to model household and economic behavior, and can thereby provide insights into measuring the value of environmental goods.

Becker (1965) originated the household production framework. This approach is based on the observation that households frequently purchase market goods that do not directly yield utility, but instead are combined with nonmarket goods and services to produce commodities which they value.

Bockstael and McConnell (1983) applied this framework to the valuation of public goods. They demonstrated that if two conditions hold then changes in the area under the demand curve for a private good can serve as a welfare measure for changes in the public good. These two requirements are that one input (a private good) must be *essential* for the production of the commodity, and that the public good is also *complementary* to the commodity.

The key requirement, that the private good is essential to produce the commodity, is equivalent to assuming weak complementarity. Weak complementarity entails two conditions. The first is that the private good z be nonessential in the sense that there be a price for z , known as the choke price, at which the demand for z falls to zero. Graphically, this implies that the demand curve has a vertical intercept. In terms of the consumer's utility function, weak complementarity implies that there is

some level of expenditure on goods other than z that will sustain a reference level of utility, U_0 , when the quantity of z (and therefore q) consumed falls to zero. The second condition is that when p , the price of z , is equal to or exceeds the choke price, the maximum willingness to pay for the environmental good is zero. In other words, weak complementarity requires that the marginal utility derived from q (hunting areas) be zero when the quantity demanded of the complementary private good z (travel to the site) is zero.⁴⁰

Given the conditions for weak complementarity, an estimate of Marshallian consumer surplus can be derived by taking the integral of the ordinary demand curve as described above.⁴¹ Achieving the ability to calculate measures of Marshallian consumer surplus was an important step forward in the management of natural resources. Economists however, have known since the work of Hicks (1939) that Marshallian consumer surplus is not the theoretically correct measure of willingness to pay because it fails to hold the individual's utility level constant. The theoretically correct measure is based on the appropriate Hicksian compensated demand curve, but unlike ordinary demand curves, Hicksian curves are not directly observable.

⁴⁰By assuming weak complementarity the researcher assumes that people do not have any non-use values for the environmental good of concern. For example, a person who never takes a trip to Prince William Sound, Alaska, would not value the prevention of oil spills in this area.

⁴¹In this study z_1 will equal zero as p_1 is set equal to the choke price (the price at which no trips are taken). This method is used to evaluate the loss in welfare that would be associated with losing access to the hunting areas.

Hicksian Welfare Measures

A second major theoretical advancement in travel cost methodology also occurred in the early 1980's. During this time period, economists developed methods for calculating the Hicksian measures of welfare directly from ordinary demand curves.

The solution to either equation 3.2 or 3.5 presented above, is an ordinary (Marshallian) demand function. Marshallian demand functions represent the quantity demanded of a good as a function of price and income. A Marshallian demand function is uncompensated in the sense that as price changes income is not adjusted to compensate for any changes in utility.

Duality theory can be used to clarify the differences between compensated and uncompensated demand functions. Problems 3.2 and 3.5 involve utility maximization subject to a maximum expenditure level. Equation 3.6 which is the minimization of expenditures subject to the attainment of some minimum level of utility, is the dual to this problem.

$$\min x + p \bullet q \quad (3.6a)$$

$$\text{s.t. } u(x, q) \geq U \quad (3.6b)$$

The solution to minimization problem 3.6 is a compensated (Hicksian) demand function, $q^* = h(p, U)$. The quantity demanded of the environmental good q is now a

function of price and utility. Along the Hicksian demand curve, as price changes the individual's level of expenditures (on both x and q) changes in order to maintain a constant utility level. By substituting the Hicksian demand function, $h(p,U)$, into the minimization equation (3.6a), the expenditure function, $e(p,U) = Y$, can be derived (Deaton and Muellbauer 1988). The expenditure function represents the minimum level of expenditures necessary to attain a particular level of utility U , given some level of prices.

The expenditure function can be used to define the Hicksian welfare measures compensating variation and equivalent variation. Compensating variation is the quantity of income that compensates the individual for a price change. That is, given the *new price* it is the change in income necessary to return her to the *original level of utility*. Equivalent variation is the income change that could be given in conjunction with the *original price* to obtain the *new level of utility* that would have been achieved if the price had actually changed. To define these measures in terms of the expenditure function, assumes that the price of a good increases from p_0 to p_1 , and as a result utility falls from U_0 to U_1 . Compensating variation (CV) and equivalent variation (EV) are now defined as

$$CV(p_0, p_1) = e(p_1, U_0) - e(p_0, U_0) \quad (3.7a)$$

$$EV(p_0, p_1) = e(p_1, U_1) - e(p_0, U_1) \quad (3.7b)$$

In a case involving a price increase, compensating variation provides enough income to maintain the individual's original utility level of U_0 . This amount of income compensation is just equal to the difference in the two expenditure functions depicted in equation 3.7a. Equivalent variation can be defined in a similar manner with reference to equation 3.7b as the reduction in income necessary to reduce the individual's utility level to U_1 .

The derivative of the expenditure function with respect to price is the Hicksian, or income compensated, demand function. Therefore, compensating variation (eq. 3.7a) and equivalent variation (eq. 3.7b) are simply the integrals of the compensated demand curves evaluated from p_0 to p_1 . These measures are depicted in Figure 3-2.

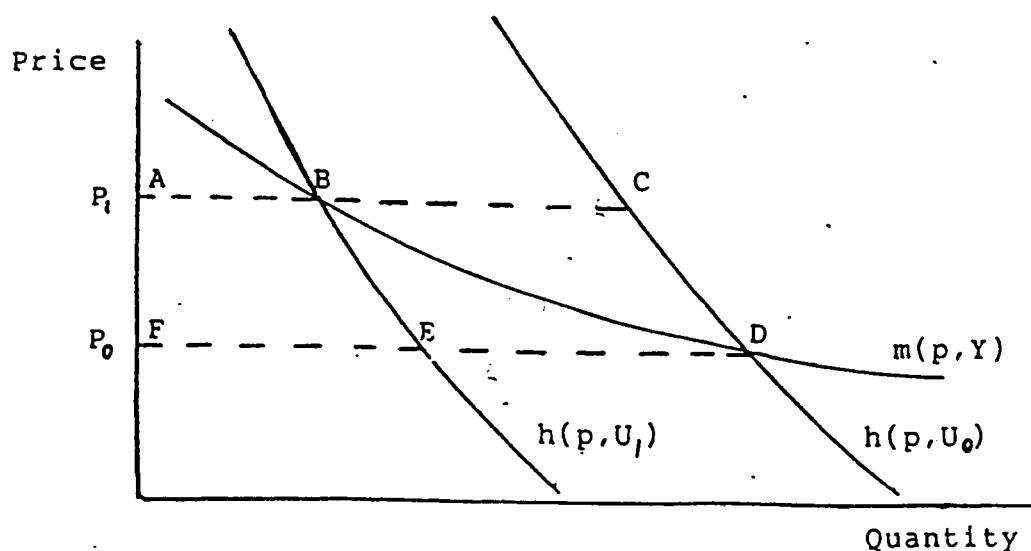


Figure 3-2 Hicksian and Marshallian Welfare Measures

The area to the left of $h(p, U_1)$, between prices p_0 and p_1 , represents Hicksian equivalent variation (ABEF). Compensating variation is the area to the left of $h(p, U_0)$, from prices p_0 to p_1 (ACDF).

Marshallian consumer surplus is the area (ABDF) to the left of the ordinary demand curve $m(p, Y)$, between prices p_0 and p_1 . This ordinary demand curve will always intersect the compensated curve $h(p, U_0)$ at price p_0 and intersect the compensated curve associated with U_1 at p_1 .

The only difference between the Marshallian and Hicksian curves is the income effect. Therefore, when the income effect is positive, as will occur with a normal good, compensating variation will exceed Marshallian consumer surplus (MCS) which in turn will be greater than equivalent variation ($CV > MCS > EV$). When the income effect is negative (an inferior good) this inequality relationship is reversed ($CV < MCS < EV$). In the case that the income effect is zero the three curves would be identical and so the three measurements of value would be equivalent ($CV = MCS = EV$).

Two types of demand functions, the ordinary Marshallian and Hicksian compensated and their corresponding welfare measures have been defined. Ordinary demand curves have the disadvantage that price and income effects are bundled together. Therefore, the Marshallian measure of consumer surplus which is derived from an ordinary demand curve will include both price (substitution) and income effects. In contrast, a theoretically exact measure of willingness to pay would focus on price effects alone.

The Hicksian measure compensating variation captures only the change in welfare associated with price effects. It is therefore a superior theoretical measure of value. Unfortunately, economists cannot observe utility, which in turn precludes them from directly observing compensated demand functions. Because compensated demand functions cannot be observed, economists have traditionally estimated ordinary demand functions and used the associated welfare measure Marshallian consumer surplus as an approximation to the Hicksian measures.

Derivation of Hicksian Welfare Measures

Directly from Marshallian Demand Curves

In the early 1980's formulas were developed which enabled researchers to calculate Hicksian welfare measures directly from the empirically observable ordinary demand curves. These types of formulas have been developed for several popular functional forms used to estimate demand equations, including the linear, semilog, quadratic, and log-linear functional forms (Hausman 1981; Bockstael et al. 1989).

Although this development was an important improvement in recreational demand theory, its' applicability is limited. The formulas can be applied to only a small number of popular functional forms, and given certain circumstances they cannot be used at all. Additionally, even if circumstances are such that these formulas can be calculated, and one of the popular functional forms is the appropriate choice for demand estimation, the variance of the calculated Hicksian welfare measure

may be unacceptably large (Adamowicz, Fletcher, Graham-Tomasi 1989). Therefore, given certain circumstances Marshallian consumer surplus may provide a better estimate of willingness to pay than the more theoretically exact Hicksian welfare measure.

A Reconsideration of Marshallian Consumer Surplus

Marshallian consumer surplus may be a better measure of willingness to pay than Hicksian compensating variation in this study for a number of reasons. This dissertation utilizes a sample of deer hunters to perform both a hypothetical and an actual travel cost study. Previous deer hunting studies have reported findings concerning the income coefficient that indicate the Marshallian measure to be a more appropriate choice (Loomis, Creel, and Park 1991; Balkan and Kahn 1988). Additionally, the imprecise manner in which the data concerning respondents' income levels was collected and measured for the U.S.F.W.S survey also suggests the usage of the Marshallian measure of consumer surplus. Five reasons for selecting the Marshallian measure of consumer surplus for use in this study will be discussed below.

1. Previous travel cost studies involving deer hunters have estimated **negative income coefficients** (Loomis et al. 1991; Balkan and Kahn 1988). Given a negative income coefficient, use of the formulas for directly calculating compensating

variation simply will not work. Although this restriction is not widely known, it is extremely important. If compensating variation is calculated with a negative income parameter estimate, the formulas will provide negative estimates of compensating variation. Negative estimates of compensating variation are economically meaningless.⁴²

2. The income variable is often poorly measured in recreation demand studies (Loomis et al. 1991). This may be particularly troublesome because estimates of compensating variation are strongly impacted by the income coefficient. This concern is noted by Bockstael, McConnell, and Strand (1991, 248-249) who state, "Perhaps no variable is measured with as much error as the income variable. ... The method of integrating back places more weight on the income coefficient than is warranted by the accuracy of income measures."

The survey utilized for this dissertation measures income very imprecisely. Survey respondents are asked which of seven very broad household income categories they fit into, and furthermore, no accounting was made of the impacts of federal, state, and local taxes. The vague manner in which the income variable is measured in this survey suggests that any attempt to measure compensating variation may be very imprecise.

⁴²Negative measures of compensating variation are economically meaningless in the following sense. The travel cost data is collected for individuals who purchase a positive number of recreation experiences. If these individuals actually had a negative willingness to pay for the recreation experience then they would purchase zero trips.

3. The estimate of deer hunting demand is likely to produce a statistically insignificant income coefficient (Loomis et al. 1991). This sample is limited to those who actually participated in the recreational activity by taking at least one hunting trip. It may be that income levels are more likely to distinguish hunters from non-hunters than they are to affect the number of trips taken in a season. If this hypothesis of respondent behavior is accurate then the statistical estimate of the income coefficient will probably be insignificant due to this limitation (no observations on individuals who took zero trips).

Given the expectation of a statistically insignificant income parameter estimate and the imprecise manner in which the income variable is measured it is expected that the calculation of compensating variation will not be very accurate. Conversely, the calculation of Marshallian consumer surplus does not incorporate the income parameter estimate, and therefore it is likely to be a more precise measurement of welfare.

4. The researcher's choice among various functional forms, which can be used for estimating the ordinary demand curve, will have a much greater impact on estimates of compensating variation than on estimates of Marshallian consumer surplus. This may occur because different functional forms which appear to fit the data equally well can integrate back to very different expenditure functions. The degree of exactness of the compensating variation measure depends upon the choice of the appropriate functional form (Bockstael et al. 1991).

5. Finally, there may be only a **relatively small difference** between the Marshallian and Hicksian welfare measures. The error induced by using Marshallian consumer surplus as an approximation of willingness to pay may be insignificant, in cases where the income coefficient is small, when compared to other potential sources of error (Willig 1976; Randall and Stoll 1983). In particular, this error may be very small in relation to measurement error in the data.

The purpose of this dissertation is to explore the comparability of hypothetical responses and actual economic behavior. The estimates of Marshallian consumer surplus derived using the hypothetical and actual travel cost methods will be used as a basis for this comparison. These Marshallian consumer surplus estimates will not be used to evaluate public policy based on hunters' willingness to pay for hunting areas. Therefore, it is less important that they represent exact Hicksian measures of willingness to pay as long as the error is consistent across both the hypothetical and actual methods.

A Testable Hypothesis

This dissertation attempts to provide evidence as to whether hypothetical responses to valuation questions correspond closely to real economic decisions. Furthermore, it attempts to investigate this hypothetical versus actual response issue in the context of a real environmental public good. This final section of chapter 3 will

first review the reasons that hypothetical and actual responses may differ and then present a testable hypothesis.

In chapter 2 a framework for classifying the reasons that responses in hypothetical and actual scenarios may differ was presented. The relationship between actual and hypothetical responses and the three reasons that they may diverge is formalized as $Rh_i = g\{Ra_i, P_1, P_2, P_3\}$. To briefly review, recall that Ra_i represents the i th individual's response to a real economic situation, where this response will entail real economic benefits and costs. Rh_i is the i th individual's response in a hypothetical (costless) economic market which attempts to mimic the real economic market that produced Ra_i . The potential sources of divergence between actual and hypothetical responses, a random error process with a mean of zero, a systematic error process, and the possibility that Rh_i will not actually be observed, are represented by P_1 , P_2 , and P_3 , respectively.

Given the relationship, $Rh_i = g\{Ra_i, P_1, P_2, P_3\}$, if it is also assumed that the hypothetical and actual response survey questions utilized in this study are equivalent, then it is clear that if P_1 , P_2 , and P_3 , do not affect the hypothetical survey responses Rh_i will be equivalent to Ra_i . If on the other hand either P_1 , P_2 , or P_3 , do affect the hypothetical response, then Rh_i will not be equivalent to Ra_i . The focus of this dissertation is to investigate whether P_2 affects Rh_i . If neither P_2 nor P_3 exert an effect on Rh_i then $E[Rh_i - Ra_i] = 0$. Such a result implies that the only source of divergence between Rh_i and Ra_i is random directionless error. This study will not directly compare Ra_i and Rh_i , but will instead compare measures of Marshallian

consumer surplus based on sets of these two types of responses. The hypothesis of primary interest in this dissertation is that there is no significant difference between values derived using hypothetical travel cost data and values which are based on actual economic decisions.

This hypothesis will be evaluated by comparing estimates of economic value (MCS) which are based on the actual and hypothetical responses, Ra_i and Rh_i , instead of comparing the survey responses directly. This is done for two reasons. The first reason is that the reports of actual and hypothetical behavior are difficult to compare directly. The actual behavior travel cost question asks hunters to report the number of deer hunting trips that they took in 1985 and the cost of a typical trip. The hunters are then presented with a hypothetical travel cost question which asks them how many trips they would have taken if the cost of a typical trip had been twice the amount they had reported actually incurring.⁴³ It is expected then that respondents will report that they would take fewer trips at the higher hypothetical cost of travel. Therefore, the initial responses to these questions are not directly comparable, but the measures of economic value which are based on these values are comparable. That is, if the respondents make a consistent tradeoff between the cost of a trip (price) and the number of trips taken (quantity), then Ra_i and Rh_i should lie on the same demand curve, and therefore, should produce equivalent measures of consumer surplus.

⁴³The survey questions and responses will be discussed in more detail in chapter 4.

The second reason for comparing consumer surplus estimates is that public decisions are based on these estimates and not on the number of trips a person would take at a given travel cost. By comparing the estimates of economic value the magnitude of the impact of any error inherent in utilizing hypothetical methods will be more readily apparent. Restated, it is likely to be more important for policy makers to know that hypothetical methods are likely to overestimate benefits by 75%, than to know that a hypothetical travel cost survey will be likely to overestimate the number of trips the typical hunter would take by approximately two trips.

The hypothesis this study seeks to investigate then can be restated as whether $E[MCS_{i \text{ hypo}} - MCS_{i \text{ actual}}] = 0$. Variations of this general null hypothesis will be tested in several ways. The hypothesis that these values are the same will be tested using a matched-pairs Wilcoxon test.⁴⁴ The nonparametric Kolmogorov-Smirnov procedure will be employed to test the hypothesis that there is no significant difference between the distribution of values derived using the hypothetical responses and the distribution of values based on actual behavior. Finally, another version of the general null hypothesis that will be tested is that there is no difference between the coefficients estimated using the hypothetical data and the coefficients generated using pooled data consisting of both the hypothetical responses and actual behavior. This hypothesis will be tested using a "Chow" test (F-test), which is a standard test for a structural break. The results of these and other tests will provide strong evidence as

⁴⁴More detail concerning the Kolmogorov-Smirnov test and the Wilcoxon test will be presented in chapter five along with all relevant citations.

to whether or not responses in hypothetical markets correspond closely to actual economic decisions.

Before proceeding on to chapter 4 it would be useful to review the reasons that hypothetical and actual responses may differ and to discuss which of these are most likely to be encountered in this study. P_2 which is the systematic error process with a mean of zero is the central concern of this dissertation. This systematic error processes was discussed in chapter 2 as a general taxonomy of potential hypothetical response biases. This taxonomy presented four major potential sources of bias. These four potential sources of bias include the possibility that individuals have motives to misrepresent their opinions, that answers given by the individuals may reflect factors other than their valuations, that individuals have an inadequate understanding of the good they are being asked to value, and that the hypothetical nature of a constructed market may bias responses. This last source of systematic error is termed hypothetical bias and is the most likely source of bias in this study. Hypothetical bias may occur when, due to the hypothetical nature of surveys, individuals lack incentives to answer carefully. Mistakes made in hypothetical markets are costless, while in contrast poor decisions made in real economic markets entail real economic penalties. These penalties, which are missing in surveys, provide the incentives for individuals to carefully consider their responses.

Hypothetical bias can take two forms. If respondents are unsure of their preferences they may search for "implied value cues" as to the good's true value. A second form of hypothetical bias, budget or time constraint insensitivity, also arises

due to the costless nature of hypothetical responses. Budget constraint insensitivity occurs when an individual fails to adequately consider his budget constraint when making hypothetical purchases. Similarly, time constraint insensitivity occurs when an individual fails to adequately consider constraints on leisure activities due to his limited amount of leisure time.

The probability that each of these various sources of potential bias will affect a particular survey may depend upon the type of good being valued, the type of survey question employed, and the characteristics of the respondents. This dissertation utilizes the U.S.F.W.S. survey of deer hunters. This survey asks both hypothetical and actual behavior travel cost questions of active deer hunters.⁴⁵

The sources of error mentioned above do not each have the same potential to cause a divergence between the actual and hypothetical travel cost responses. Each of the types of bias mentioned above will now be briefly evaluated for its' potential to cause such a divergence in this particular study.

Incentives to misrepresent opinions should be minimized by the travel cost question format. Before individuals will act strategically they must first perceive a connection between the misrepresentation of their opinions and the potential for some type of personal gain. The hypothetical travel cost question asks hunters the number of trips they would take at a given hypothetical travel cost. By using this indirect method to ask about a hunter's preferences, as opposed to directly asking their

⁴⁵The travel cost questions asked in the U.S.F.W.S survey are presented in Appendix A.

maximum willingness to pay in the form of higher taxes or license fees, the question format makes it more difficult for the respondent to connect the responses he supplies to possible changes in either the provision level of the sites or the amount he will have to pay for access to these sites. Additionally, the nature of the good may act to minimize the potential for compliance bias. This potential bias is minimized because active hunters are not likely to feel that they will gain favor in the eyes of a Census Bureau employee by falsely stating the number of trips they would take.

There is also a very low probability that the survey participants will give answers that reflect factors other than their valuations. Hunters are familiar with the act of making tradeoffs between the cost of a trip and the number of trips they take. This familiarity with the quantities involved in the tradeoff should reduce any insensitivity to scope problems to a negligible level. Moreover, the type of environmental good should also diminish any possibility of symbolic bias. Respondents are less likely to react to a symbolic meaning when valuing hunting sites than when they are valuing a good such as protection of an endangered species like a panda bear. Additionally, by avoiding the introduction of taxes, license fees, and other possibly objectionable payment vehicles the hypothetical travel cost method diminishes the likelihood of a significant context bias.

The responses are also unlikely to be biased due to an inadequate understanding of the environmental good being valued. The U.S.F.W.S. sample includes only active hunters. It is reasonable to assume that active hunters are very familiar with the hunting sites and therefore, have a solid understanding of the

characteristics of the hunting sites, the tradeoff of travel cost for hunting trips, and the utility derived from the recreational experience.

The final major source of systematic response error is hypothetical bias which occurs when respondents lack incentives to adequately search their preferences. Hypothetical bias can take two forms, implied value cues and an insensitivity to the budget or time constraints. In reference to implied value cues, the survey participants are first asked the number of actual deer hunting trips they took in the previous hunting season and the actual cost of a typical trip. Respondents are then asked how many deer hunting trips they would have taken if the average trip cost were twice as much as the actual amount they reported. Since the hypothetical question is asked immediately after the actual behavior question, the respondents may use their past behavior as a type of implied value cue. If the report of past behavior acts as an implied value cue, then this should improve the degree of correspondence between the actual and hypothetical responses.

Finally, hypothetical bias may also occur if respondents are insensitive to their budget or time constraints. The U.S.F.W.S. survey presented deer hunters with a hypothetical scenario in which the price of the environmental good was given and they were asked how many trips they would purchase. If a respondent did not give adequate consideration to either his budget or time constraint he might tend to overestimate the quantity of recreational hunting experiences he would hypothetically purchase. This seems to be a reasonable hypothesis concerning respondent (deer

hunter) behavior. If respondents are insensitive to their budget or time constraints their hypothetical responses may be systematically biased upward.

In summary, the hypothetical travel cost method as applied here to active deer hunters, should minimize the potential biases resulting from motives to misrepresent opinions, answers that reflect factors other than a valuation, and an inadequate understanding of the good being valued, but hypothetical bias may still be important. Hypothetical bias may have two types of effects. First, the possibility of an implied value cue may tend to increase the degree of correspondence between the hypothetical responses and actual behavior. Second, respondents who fail to adequately consider either their budget or time constraint may tend to over estimate the number of trips they would take. The insensitivity to a budget or time constraint hypothesis suggests that valuation estimates based on hypothetical behavior may be systematically biased upward.

This study is aimed at investigating the relationship between hypothetically stated preferences and real economic commitments. To achieve this a testable hypothesis was formulated. A number of tests were proposed for evaluating this hypothesis and a more detailed explanation of these tests will follow in chapter 4.

Chapter Four

The Empirical Model

This chapter will first discuss the **data set** which will be utilized to estimate the ordinary demand curves. Next, the choice of an appropriate **functional form** to use in the process of demand estimation, and the potential impact of this choice will be reviewed. Following this, the theoretical issues of how to handle the monetization of the **opportunity cost of time**, and whether to base consumer surplus estimates on the **actual or predicted number of trips** will be addressed. Finally, this chapter will review the **statistical tests** which will be employed to test the null hypothesis. Each of these issues will have a significant impact on the test results and conclusions that follow in chapters 5 and 6. Therefore, each of these topics requires a thorough review

The Data Set

The data set utilized for this study is the 1985 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (NSFWAR) which was a project of the United States Fish and Wildlife Service (USFWS). This survey is the seventh in a series conducted since 1955.

The 1985 NSFWAR conducted in-person surveys of 28,011 fishers, hunters, and non-consumptive recreationists across the United States.⁴⁶ Individuals were asked to recall their outdoor recreation activities for the previous calendar year (1985), and report the number of recreation trips they took, the distance they traveled, the cost of a typical trip, and the ways in which they used the recreation site. Survey participants were also asked about a number of their socioeconomic characteristics such as age, income level, and marital status.

This dissertation will utilize only the responses from the individuals identified as active deer hunters, because this group was asked both hypothetical and actual behavior travel cost valuations questions. This group consisted of 1,245 active deer hunters who each took at least one deer hunting trip in 1985. The economic valuation (travel cost) questions they were asked are presented in the appendix.

Each of the NSFWAR variables used in this study will be briefly described. The dependent variable used in the actual behavior travel cost model is ATRIPS. This variable represents the number of trips that the individual reported that he *actually took* in 1985. The corresponding dependent variable for the hypothetical response travel cost model is H2TRIPS, which represents the number of trips the individual reports that he would have *hypothetically taken* at some suggested hypothetical price. APRICE and H2PRICE are the primary independent variables for the actual and hypothetical cases respectively which represent the costs of taking a

⁴⁶Non-consumptive recreationists consist of bird watchers, wildlife photographers, and others that utilize wildlife without harvesting or catching it.

deer hunting trip.⁴⁷ These trip costs are equal to the summation of the monetary costs of taking a trip, ACOST or H2COST, and the opportunity cost of time travel, TIMECOST. INCY represents the respondent's household income level. This variable is discussed in detail following Table 4-1. CITYCATG is a categorical variable which refers to the description of the town or city in which the individual resides. A value of unity represents a big city or urban area, a value of 2 represents a small city or town, and 3 refers to a rural area. SUBACTVY is the activity that the respondent would have substituted for deer hunting had deer hunting been so expensive that they did not go at all. These potential substitute activities include, other big game hunting, other hunting, fishing, other outdoor recreation, work, don't know, and other. The independent variable REGN represents eleven regions of the United States. The variables REGN and SUBACTVY are treated as class (dummy) variables.⁴⁸ AGE refers to the hunter's age and ranges from 16 to 82. EDUCATDZ represents the highest level or grade of education attended by the respondent, where 1-8 are elementary, 9-12 are high school, and 13-19 are years of college.⁴⁹ MARRIEDZ is a yes/no dummy variable for whether or not the hunter is

⁴⁷The values presented in table 4-1 set the opportunity cost of time to be equal to the wage rate. An alternative method for this will be discussed below.

⁴⁸Treating regional variation as a "class variable" has the same statistical impact as using individual "dummy variables" for each of the different regions. Using the class variable approach simply produces the additional output of an overall significance level statistic (F statistic). This treatment is further discussed in chapter 5.

⁴⁹In this study, the number of years of education (1-19 yrs) were treated as a quantitative variable. An alternative method would be to treat education as a categorical (qualitative) variable by breaking this string of 19 years into several

married. BAG is another dummy variable which refers to whether the hunter harvested a deer during 1985. The mean, minimum, and maximum values, as well as the standard deviation for each of the variables discussed above are presented below in table 4-1.

Table 4-1. Travel Cost Model Variables

Variable	N	Mean	Std Dev	Minimum	Maximum
ATRIPS	1245	8.45	11.88	1.00	103.00
H2TRIPS	1245	4.98	8.55	1.00	91.00
APRICE	1245	98.64	195.20	1.06	3313.31
H2PRICE	1245	162.92	330.70	2.06	6313.31
ACOST	1245	64.29	141.44	1.00	3000.00
H2COST	1245	128.58	282.89	2.00	6000.00
TIMECOST	1245	34.35	82.78	0.03	1577.63
INCY	1245	29230.92	17797.08	5000.00	87500.00
EDUCATDZ	1245	12.68	2.50	2.00	19.00

The 1985 NSFWAR has several strengths which make it the best data set available for achieving the goals of this study. These advantages were discussed in chapter 2, and therefore will only be briefly summarized here.

The first strength is that by employing the seldom used hypothetical travel cost question format, it becomes possible to compare an observable *indirect* method to a hypothetical *indirect* method. This will eliminate the confusion which characterized

previous studies where values derived using indirect actual behavior methods were compared to those derived using direct hypothetical methods. These previous comparisons suffered because any differences found in the derived welfare values could have been attributed to either the differences in assumptions and methodological choices between the two methods, or the hypothetical versus actual nature of the responses.

Another strength of this data set is also attributable to the travel cost question format. The travel cost question format should reduce any incentives for strategic behavior as well as minimize any insensitivity to scope effects. This question format is also void of any implied value cues that would cause the hypothetical and actual responses to diverge. Therefore, any differences found in the valuation estimates derived using actual behavior, and those based on hypothetical responses, may be attributable to a failure on the part of respondents to adequately consider their budget or time constraints.⁵⁰

Additionally, since this dissertation will focus only on the responses of the 1,245 active hunters, the public good (deer hunting sites) is likely to be a very familiar commodity to this group of respondents. This high degree of familiarity should present a best case scenario for respondents to accurately state their hypothetical preferences regarding an environmental public good.

⁵⁰The presence of an insensitivity to respondent's budget or time constraints is only one of many possible sources bias. Although this may represent a likely source of divergence between hypothetical and actual responses, the reader should be aware that this hypothesis is not empirically testable given the data set utilized in this study.

Finally, the sample size of 1,245 deer hunters is much larger than the samples used in most previous studies which compared values from hypothetical and actual data. Previous studies frequently employed sample sizes well under 100 (Smith et al. 1986; Carson and Mitchell 1989). The much larger sample size will increase the power of the statistical tests, which will enable these tests to reject the null hypothesis more easily.

Although the NSFHWAR generates a rich data set, its method for measuring respondents' income levels is a weakness. The method employed for measuring income gives rise to several problems.

First, this method provides only an imprecise measurement of respondents' income levels. Survey participants are asked to state which of seven very broad categories represented their household income level for 1985. The seven income categories are: under \$10,000; \$10,000 to \$20,000; \$20,000 to \$25,000; \$25,000 to \$30,000; \$30,000 to \$50,000; \$50,000 to \$75,000; and above \$75,000. The vagueness inherent in these rather broad income categories will increase the likelihood that the income variable will be statistically insignificant in terms of explaining the number of deer hunting trips taken.

Another problem is that the income ranges cannot be directly used as a regression variable in this study.⁵¹ Before income can be used as an explanatory

⁵¹An alternative and more defensible method for handling a variable that is measured in intervals or categories, such as the income ranges encountered here, is to treat it as a categorical (dummy) variable. Unfortunately, this method is not appropriate for this study because here the respondents' income levels are used both directly to explain variation in the dependent variable trips, and indirectly in the

variable some method must be employed for choosing a point to represent each income category. The method chosen for the first six categories is to simply take the midpoint of each income range. This may not represent either the true median or mean income distribution for deer hunters across the United States.

This method of calculating midpoints is obviously not applicable to the last range, above \$75,000. After reviewing the available data on income distribution, a point estimate of \$87,500 was chosen to represent this income range.⁵² Although this midpoint choice (\$87,500) involves some researcher judgement, its overall impact is likely to be minimal because less than three percent of the hunters sampled (37 of 1,245) reported that their income was greater than \$75,000.

The two methods for transforming the income categories into specific numerical points used here may not be as troubling as they might first appear. As was discussed in chapter three, the income variable is usually a poor predictor of recreation demand (Loomis et al. 1991). This is especially likely in this study because only deer hunters who took at least one trip in 1985 were surveyed, and income levels are more apt to distinguish hunters from non-hunters than they are to predict the number of trips taken (Bockstael et al. 1991). Also the impact of the

monetization of the opportunity cost of time which will be discussed later in this chapter.

⁵²None of the available sources of information on income distribution within the United States provided detailed information concerning the distribution of income for individuals who earn \$75,000 or more.

income variable on the calculated measure of Marshallian consumer surplus will probably be very small (Randall and Stoll 1983).

The Choice of Functional Form for Travel Cost Demand Estimation

Before Marshallian consumer surplus can be calculated, estimates of the demand for travel to the hunting sites must be generated. The choice among functional forms can have a significant impact on measures of Marshallian consumer surplus, which are derived from the estimated recreation demand equations. Past studies have found a great deal of variation in consumer surplus estimates depending upon the functional form employed.⁵³

The most popular forms chosen in past travel cost studies include the linear, semi-log, and log-linear equations. Economic theory is of little help in determining the exact functional form of the quantitative relationships in recreational demand studies (Ziemer, Musser, and Hill 1980), therefore, this determination has usually been made empirically. This is the approach that will be taken in this dissertation.

⁵³Studies comparing estimates of consumer surplus derived using various functional forms have found differences which are usually much greater than the differences between Hicksian and Marshallian measures of welfare. Bockstael et al.(1989) found a difference of 16% between welfare estimates generated using the linear and semi-log functional forms, while Ziemer et al.(1980) found differences ranging from 29% to 286% in a comparison of the linear, quadratic, and semi-log forms.

There is a small group of studies which have compared the merits of various functional forms used for recreational demand estimation. These studies provide a basis for evaluating the functional forms that might be utilized in this study.

Bockstael et al. (1989) evaluated the linear and semi-log functional forms using the flexible Box-Cox transformation as a criteria. They found that the semi-log form performed better than the linear model. Ziemer et al. (1980) conducted a similar experiment in which they compared consumer surplus estimates derived using linear, quadratic, and semi-log functional forms. These authors, who also utilized a Box-Cox transformation as a criteria, found the semi-log functional form to be superior to the quadratic and linear forms.

Several other studies have found that the semi-log and log-linear functional forms may be preferred to either the quadratic or linear forms for several reasons. Strong (1989) notes that, in general, taking the log of the dependent variable (trips) has been found to reduce heteroskedasticity.⁵⁴ In another relevant study Adamowicz et al. (1989) found the log-linear model to be a variance-minimizing functional form.

This dissertation will determine the appropriate functional form by investigating the nature of the estimated model's variance. First a linear model will be estimated. The residuals of this model will then be inspected for normality, and if they are found to be non-normal other functional forms will be used in an attempt to correct this weakness. The functional forms will also be evaluated by comparing their

⁵⁴Similar comments implying that taking the log of trips will reduce heteroskedasticity were made by Vaughan, Russell, and Hazilla (1982).

goodness of fit measures. As the discussion above suggests it is expected that the semi-log and log-linear functional forms will provide superior estimation results.

The Opportunity Cost of Time

There has been widespread agreement among environmental economists that the time spent traveling to a site should be included as a component of travel cost when estimating recreation demand (Knetsch 1963; Cesario and Knetsch 1970). There is however, no generally agreed upon method for converting the time cost of travel into its monetary equivalent.

The method which is most widely used to convert the time cost of travel into a monetary measure is to assume that the wage rate, or some fraction thereof, represents the appropriate monetary price of travel time. Cesario (1976) concluded that the appropriate scarcity value of time was approximately one third of the wage rate. Based on Cesario's work some economists have assumed that the correct monetary cost of an hour of travel time is approximately one quarter to one half the wage rate.

This dissertation will employ a more theoretically palatable method for determining the appropriate fraction of the wage rate to represent the opportunity cost of time. This alternative, which was developed by McConnell and Strand (1981), allows for the estimation of the relationship between the wage rate and the opportunity

cost of time from the sample data. Suppose that the demand for visits can be specified as follows:

$$V = \alpha + \beta p \quad (4.1)$$

where V represents visits to the hunting site, and p represents the price or travel cost of a visit, α is the estimated intercept and β represents the estimated price coefficient. This can be rewritten as

$$V = \alpha + \beta_1(c + k \cdot t \cdot w) \quad (4.2)$$

where c represents the explicit dollar trip cost⁵⁵, t is travel time, w is the wage rate, and k is the fraction of the wage rate which represents the monetary value of the opportunity cost of time. In order to estimate k , McConnell and Strand first estimate equation 4.3 which includes a separate coefficient for the time costs and the explicit trip costs.

$$V = \alpha + \beta_1(c) + \beta_2(t \cdot w) \quad (4.3)$$

⁵⁵The "explicit" trip costs, c , refers to all costs associated with taking a hunting trip for which a monetary payment may be made. For example, these explicit costs might include automobile fuel purchases and lodging expenses. The total cost of a trip then would equal these explicit costs plus the opportunity cost of time spent traveling to the site.

By making the reasonable assumption that a one dollar change in either the explicit trip costs or the opportunity cost of time spent traveling should have an equivalent effect on the number of trips taken, k can be estimated as $k = \beta_2 / \beta_1$.

Estimated Versus Actual Number of Trips Taken

The hypothesis to be evaluated in this dissertation is that the measure of consumer surplus based on hypothetically stated preferences is equal to a corresponding measure based on actual economic actions.⁵⁶ The correct evaluation of this hypothesis requires the accurate calculation of Marshallian consumer surplus. The calculation of this measure may be sensitive as to whether it is based on the actual number of recreation trips taken or on the number of trips predicted to be taken by the regression estimate. This decision should be based on what the researcher believes to be the major source of error in the recreation demand estimate.

There are several sources of error that may be present in a recreation demand study. These potential sources of estimation error include the following (Bockstael and Strand 1987).

1. **Omitted Variables** - This type of error most often refers to unmeasured socioeconomic variables. Omitted variables are the most widely recognized source of error in recreation demand studies.

⁵⁶Refer to chapter three of this study.

2. **Random Behavior** - This type of error refers to the possibility of a fundamental randomness in human behavior. This is not regarded as a major source of concern in most recreation demand studies.

3. **Measurement Error** - Measurement error can take the form of either the inaccurate recall of the number of recreational trips taken annually or the poor recall of an independent variable such as travel expenses. Although not as widely recognized, measurement error may constitute as large a problem as omitted variables. Measurement error may be an especially big problem when respondents are asked to recall information over a lengthy period of time such as the previous year's activity.

Bockstael and Strand (1987) investigate the relationship between these sources of error and the estimated benefit measure. By referring to the consumer surplus formula presented in chapter three, $CS = z^2/2B$, it becomes apparent that the number of trips, z , will have a dramatic impact on the estimate of consumer surplus. Therefore, any significant error in the estimation of the demand for trips, z , will have a tremendous effect on the estimate of consumer surplus.

Bockstael and Strand's (1987) investigation led them to the following conclusion. If the researcher believes that all of his error is due to omitted variables then she should use the stated (actual) number of trips taken for welfare calculations. However, if the researcher uses the actual number of trips, when in fact some of the error is due to measurement error, then a substantial overestimation of benefits can occur. If on the other hand the researcher uses the estimated number of trips,

because she believes that any error is due to measurement problems, then underestimation of consumer surplus is likely since omitted variables are a likely source of error in any recreation demand study.

Both omitted variables and measurement error are likely to be present in the estimation of these functions. Omitted variables are a concern in virtually all recreation demand studies, and the current study is also likely to be impacted by measurement error because the recall period was one calendar year (relatively long). Based on the work of Bockstael and Strand's (1987) findings it is reasonable to assume that either available option, using the actual or predicted trips, is likely to be associated with a significant degree of error in this study. Therefore, consumer surplus will be calculated using both the actual and estimated number of trips. The statistical tests will then be conducted on each set of benefit estimates so that the results can be shown to be independent of the choice between actual and predicted trips for use in calculating consumer surplus.

Statistical Tests of the Null Hypothesis

Four statistical tests will be employed in this dissertation. These procedures include the Wilcoxon signed-rank test, the sign test, the Kolmogorov-Smirnov test, and an F-test. An outline of each of these procedures will be presented here, and the results of these tests will be presented in chapter 5.

The Wilcoxon signed-rank test can be used to test the hypothesis that the median values of two distributions are equal. This test is referred to as a non-parametric test because it does not make any strict distributional assumptions. In contrast, traditional test statistics, such as a t-test or F-test, are called parametric test because they depend on the specification of a specific probability distribution except for a set of free parameters. The matched pairs Wilcoxon signed-rank procedure, and other non-parametric tests may be preferred to the more traditional parametric tests, because non-parametric methods perform the tests without requiring any strict distributional assumptions, and even in the case of a normal distribution non-parametric methods are almost as powerful as their parametric counterparts.

The matched pairs Wilcoxon signed rank test is implemented by first taking the difference of each of the matched observations, which in this case are the consumer surplus values based on hypothetical and actual data ($MCS_{\text{actual}} - MCS_{\text{hypo}}$). The hypothesis is now stated in terms of testing whether this difference is equal to zero. The absolute values of the differences are ranked from one to n , where n equals 1,244 the sample size.⁵⁷ Each rank is then reassigned the algebraic sign of the difference that generated the rank. The test statistic is the lessor of W_+ or W_- , where W_+ equals the sum of all of the positive ranks and W_- is the sum of the negative ranks. The test statistic W , is the smaller of W_+ and W_- (Milton and Arnold 1990).

⁵⁷Many non-parametric tests are based on taking the ranks of a variable and then analyzing these ranks.

A second very similar non-parametric test utilized in this study is the sign test. This procedure begins by calculating the difference between each matched pair of consumer surplus estimates just as the Wilcoxon signed-rank method did. Next, the sign test sums a net total of the positive and negative algebraic signs of the differences. The sign test may be preferable in cases where the differences are neither normally distributed nor symmetric.

Another non-parametric method employed in this dissertation is the asymptotic Kolmogorov-Smirnov test. The Kolmogorov-Smirnov procedure will test the null hypothesis that there is no significant difference between the distribution of consumer surplus values derived using hypothetical responses and the distribution of values based on reports of actual behavior.

The Kolmogorov-Smirnov statistic is used to evaluate whether the distribution of a variable is the same across different groups. If the asymptotic Kolmogorov-Smirnov statistic is found to be sufficiently large it can be concluded that the two distributions do not arise from the same population.

In addition to the three non-parametric tests described above, a parametric test will also be utilized. This procedure is a type of F-test. This F-test will evaluate the hypothesis that the coefficients estimated using either the actual or hypothetical demand data are equal to the coefficients estimated using a pooled data set consisting of both the hypothetical and actual data.

This test is performed as follows. First, the travel cost demand model is estimated for both the actual and hypothetical data. Next, the data is pooled and a model is estimated for this new pooled data set. The F statistic is now calculated as

$$F_{k,N+M-2k} = [(ESS_R - ESS_{UR})/k] / [ESS_{UR}/(N+M-2k)] \quad (4.4)$$

where ESS_R refers to the restricted sum of the error sum of squares, ESS_{UR} refers to the unrestricted sum of the error sum of squares, k is the degrees of freedom, N is the size of the sample of actual responses, and M refers to the sample size for the hypothetical responses (Pindyck and Rubinfeld 1991).⁵⁸ In this case the unrestricted sum of error sum of squares, ESS_{UR} , is found by summing the sum of squares for the two non-pooled sample regressions. ESS_R , is simply the error sum of squares associated with the pooled sample regression.

If the F statistic is sufficiently large then the hypothesis of equivalence of the estimated coefficients can be rejected. This would imply that there is some significant difference in the way individuals respond in hypothetical and real markets for environmental public goods.

In sum, the tests to be utilized in this study include three non-parametric tests and one parametric test. Both the matched pairs Wilcoxon signed-rank test and the sign test will investigate whether there is a statistically significant difference in the

⁵⁸In the present study both N and M are equal to 1,244. The sample sizes must be equal because they are matched pairs responses on the same set of respondents.

median values of the consumer surplus estimates which are based on the hypothetical and actual response samples. The Kolmogorov-Smirnov procedure will test whether the two distributions of consumer surplus arise from the same population. Finally, the equivalence of the estimated coefficients for the actual and hypothetical demand models will be evaluated using an F-test. This group of tests, when applied to a number of models and estimates of consumer surplus should provide a clear picture of the degree to which the hypothetical and actual responses correspond.

Chapter Five

Empirical Results

Chapter 5 will present three types of empirical results. First, an estimate of the demand for hunting recreation experiences based on reports of actual behavior will be developed. Next, calculations of Marshallian consumer surplus based on both hypothetical responses and actual behavior will be displayed. Finally, this chapter will present the results of several statistical tests which are intended to provide additional evidence about the relationship between hypothetically stated preferences and real economic actions. More specifically, these tests seek to provide additional information concerning the relationship between responses in actual and hypothetical markets for an environmental public good.

Chapter 5 begins by presenting a summary of the development of the statistical model used to estimate the demand for hunting recreation. The first step to be taken in this process is to estimate a linear demand function which will in turn be used to evaluate the assumption of a normally distributed error term. If this assumption is found to be violated then procedures will be undertaken which will correct for this weakness. Next, the data will be inspected for outliers, and if any such data points are discovered they will be dropped from the sample. Following this, the process of model selection, including the determination of which independent variables will be included in the final estimated model, will be undertaken.

After a model has been selected and estimated using both the actual and hypothetical trip data, it will be used to calculate measures of Marshallian consumer surplus. These measures will be calculated under a number of different assumptions such as the decision as to whether the actual or predicted number of trips will be used. Finally, in order to determine whether there is a difference between hypothetical and actual responses, the estimates of demand and consumer surplus will be evaluated with several statistical tests. These tests include a matched pairs Wilcoxon signed rank test, a sign test, a Kolmogorov-Smirnov test, and an F-test.

Demand Estimation Results

The estimation of the demand for hunting trips was conducted utilizing the pooled travel cost method.⁵⁹ The number of hunting trips taken was estimated as a function of the price of the trip, whether or not the hunter harvested a deer, the region of the country in which the hunter resides, and a number of other socioeconomic variables such as age and educational level.

The initial step undertaken in the overall process of developing a statistical model suitable for the demand estimation task at hand was to investigate the error term. Initially a linear model was estimated and used to evaluate the assumption of a

⁵⁹The simple pooled travel cost method estimates recreation demand as a function of the costs of travel and a number of other (mostly socioeconomic) variables. It is referred to as a pooled method because it pools observations across sites and estimates one overall regression. Other well known versions of the travel cost method include the hedonic, discrete choice, and varying parameters methods.

normally distributed error term. A boxplot, histogram, and Shapiro-Wilk statistic were derived from this linear regression model.

The distribution of the residuals of this linear model were investigated both visually and statistically. A visual inspection of the histogram and boxplot revealed that the distribution appears to have a lengthy upper tail. This visual inspection of the histogram and boxplot implies that the residuals are not normally distributed. In addition to this visual evidence, a Shapiro-Wilk test gives a test statistic of 0.67532 with the resulting p-value of 0.0001. The null hypothesis being tested by the Shapiro-Wilk test is that the sample is distributed normally. In this case the p-value of 0.0001 is the probability that any difference from a normal distribution in the sample has occurred by chance. In this instance the Shapiro-Wilk test results mean that it is very likely that the error term is not normally distributed.⁶⁰ The same procedure, of inspecting the histogram, boxplot, and Shapiro-Wilk statistic, was repeated for the hypothetical response model and similar results were obtained. The error term was found to be abnormal in both cases.

In an attempt to correct for this weakness the data was transformed logarithmically. A log-linear model, $\ln z = \alpha + \beta \ln p$, and a semilog model, $\ln z = \alpha + \beta p$, were then estimated and inspected in the manner used above for the linear model. The two transformations each provided an improvement in the

⁶⁰Values for the Shapiro-Wilk statistic can range from zero to one. As the distribution approaches a normal distribution the value for W approaches one. As the Shapiro-Wilk test statistic and p-value increase it becomes more difficult to reject the null hypothesis of a normal distribution. When the null hypothesis cannot be rejected by this test it is assumed that the data is normally distributed.

distribution of the error term.⁶¹ The results of the log-linear model were superior to that of the semilog model, and therefore, only the results pertaining to the log-linear transformation will be discussed below.

The histogram and boxplot for the log-linear transformation looked quite different than those based on the untransformed data. Specifically, the upper tail of the distribution, associated with the linear model, was much longer than the corresponding tail for the logarithmically transformed data. A visual inspection of the box plots and histograms revealed that the transformed data more closely resembles a normal distribution. Furthermore, the Shapiro-Wilk statistic value of 0.985699 and the associated p-value 0.2192, which were derived from the log linear model, provide strong additional evidence that the error terms are more normally distributed after the data was logarithmically transformed.⁶² The log-linear transformation then provides an improvement over the linear model because this transformation leads to a more normally distributed error term.

Now that a transformation has been chosen which will improve the distribution of the residuals, the next step is to determine whether any of the data points should be dropped because they are outliers. The method employed is to inspect the studentized

⁶¹An improvement in this case refers to a finding of a more normally distributed error term.

⁶²Values for the Shapiro-Wilk statistic, W , can range from zero to one. Notice that $W=0.985699$ is much closer to one than the statistic value $W=0.67532$, which is associated with the linear model. This strongly suggests that the log-linear transformation led to a more normally distributed error term. To be precise the Shapiro-Wilk test fails to reject the null hypothesis that the residuals are distributed normally at a 20% significance level ($p\text{-value} = 0.2192$)

residuals. "Studentized residuals" are a standardized version of residuals. They are termed "studentized" because the residuals approximately follow a t distribution. Since, studentized residuals are scale free, they are not affected by the size of the regressors (Myers 1990). By utilizing studentized residuals it is possible to more accurately determine the relative impact of various possible outliers.

Studentized residuals were calculated for each of the 1,245 observations in the NSFWAR data set. The largest studentized residuals and their corresponding data points are presented below in Table 5-1.

Table 5-1. Studentized Residuals

<u>OBSERV.</u>	<u>LNATRIPS</u>	<u>PREDICT</u>	<u>RESIDUAL</u>	<u>RSTUDENT</u>
92	2.99573	0.74180	2.25393	2.49299
142	3.91202	1.89053	2.02149	2.24687
302	4.09434	1.46061	2.63373	2.91399
306	3.73767	1.31615	2.42152	2.67609
329	3.29584	0.82681	2.46903	2.72394
352	4.09434	1.77543	2.31892	2.55993
440	3.55535	0.85554	2.69981	2.98020
569	3.68888	1.47468	2.21420	2.45038
648	3.21888	1.20824	2.01063	2.21748
656	4.49981	2.14240	2.35741	2.60430
694	4.09434	1.87885	2.21550	2.44574
701	4.60517	2.21425	2.39092	2.66371
764	4.31749	2.15168	2.16581	2.39472
799	3.40120	1.06420	2.33700	2.58224
919	4.49981	2.39904	2.10077	2.31882
1018	4.55388	0.68657	3.86730	4.29134
1019	4.09434	0.79720	3.29714	3.64758
1025	4.63473	1.89966	2.73507	3.02366
1075	4.49981	1.71420	2.78561	3.07189
1217	3.21888	1.15635	2.06252	2.27026

Table 5-1 contains all of the observations for which the associated studentized residuals were greater than two (20 of 1,245 observations). The column labeled OBSERV refers to the observation number within the sample of 1,245 hunters. The

second column, LNATRIPS, is the natural log of the number of hunting trips each individual reports that they actually reported taking in 1985. This is the dependent variable in the log-linear travel cost model. PREDICT is the predicted value for LNATRIPS which is derived from the estimated travel cost model. The fourth column is titled RESIDUAL, and contains the residual which was calculated for that observation by subtracting column three from column two (actual-predicted). The final column contains the studentized residuals which are titled RSTUDENT.

These studentized residuals are then evaluated by referring to critical values for the R-student statistic, which can be computed by using the Bonferroni inequality. The critical values calculated using the Bonferroni inequality are similar to the familiar t-statistics. T-statistics cannot be used, however, because they are inappropriate for conducting multiple simultaneous tests, and in this case all 1,245 observations are being evaluated simultaneously.

Given the sample size of 1,245, and a chosen critical value of five percent, employing the critical values described above implies that any observation whose studentized residual is greater than four is identified as an outlier (Myers 1990, 481). Observation number 1,018 has a studentized residual of 4.29, and therefore, should be dropped from the sample because it is an outlier (Myers 1990).

After this outlier was dropped the log-linear model was re-estimated using the data which is based on reports of actual behavior. These estimation results are used as a starting point in the process of choosing the best set of explanatory variables. The set of explanatory variables were evaluated using a number of selection methods

such as the stepwise, forward, and backward procedures for fitting a model.⁶³ In addition an interaction variable was created to account for the interplay between the independent variables education and income levels.

A number of experiments were conducted with a variety of variables being excluded. The final set of explanatory variables, and the results of this regression, appear below in Table 5-2. Initially the actual response data were used in this process of choosing a final set of explanatory variables. The results which appear in Table 5-2 are based on this "actual" response data. The same process was repeated for the "hypothetical" response data and the results were extremely similar. The same independent variables were found to be statistically significant in the hypothetical and actual response experiments. Each of the variables retained are statistically significant at the 10% level in both the hypothetical and actual response models. CITYCATG, the variable referring to the size of the city in which the respondent resides, was dropped due to a lack of significance in explaining variation in the dependent variable. CITYCATG never reached a significance level ($Pr > F$) which exceeded .25.⁶⁴

⁶³Myers (1990, 190) warns of a potential shortcoming in these sequential selection methods when he notes, "The final result is very much dependent on the sequential method chosen and the F_{IN} and F_{OUT} values". The terms " F_{IN} " and " F_{OUT} values" refer to the chosen F-test critical values. To avoid this problem each of the three commonly employed sequential methods, the stepwise, forward, and backward procedures, and several critical (F_{IN} and F_{OUT}) values were utilized. The set of independent variables suggested by these methods was very consistent across these variations in selection methodology.

⁶⁴Smaller $Pr > F$ values imply a greater statistical significance for the variable. In contrast to the CITYCATG variable, the PRICELN variable $Pr > F$ approached .0001.

Table 5-2. Final Model - Log-linear Form - Actual Responses

General Linear Models Procedure					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	22	439.313	19.969	24.26	.0001
Error	1221	1004.989	0.823		
Corrected Total	1243	1444.302			
LNTRIPS					
	R-Square	C.V.	Root MSE	Mean	
	0.304	59.683	0.9072	1.520	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
PRICELN1	1	283.764	283.764	344.76	.0001
BAG	1	23.051	23.051	28.10	.0001
SUBACTVY	6	30.591	5.099	6.19	.0001
INCYLN	1	6.253	6.253	7.60	.0059
AGELN	1	2.971	2.971	3.61	.0876
REGN	9	82.268	9.141	11.11	.0001
MARRIEDZ	1	6.375	6.374	7.75	.0055
EDUCATLN	1	0.574	0.574	0.70	.4037
INCYLN*EDUCAT	1	3.464	3.464	4.21	.0404
T for HO:					
Parameter	Estimate	Parameter=0	Pr > T	Std Error of Estimate	
INTERCEPT	-4.018	-1.11	0.2674	3.621	
PRICELN1	-0.290	-15.45	0.0001	0.019	
BAG	-0.366	-6.63	0.0001	0.055	
INCYLN	0.833	2.29	0.0220	0.101	
AGELN	-0.228	-2.78	0.0054	0.363	
MARRIEDZ	0.174	2.64	0.0084	0.066	
EDUCATLN	2.786	1.97	0.0491	1.415	
INCYLN*EDUCATLN	-0.294	-2.05	0.0404	0.117	

In order to isolate potential variation between different regions and between different substitute activities, the statistical design of "blocking" was used. Treating regions and substitute activities as blocks holds the differences between regions and activities constant so that these differences don't affect prediction in the model.⁶⁵

This set of explanatory variables provide the best estimate of the demand for hunting recreation trips. These variables will be used for the remainder of this study.

The results of the transformations discussed earlier in this chapter demonstrated that the log-linear functional form is the best choice for this study because it normalized the residuals. These results confirm what was suggested by the studies cited in chapter four, the log-linear and semilog functional forms are superior to the linear form for this estimation of recreation demand. Although the linear model provides an inferior estimate of hunting demand, it will still be presented in this study because it is familiar to many environmental economists. This presentation is necessary in order to address the potential concern that the results and conclusions reached in this study are a result merely of the functional form chosen. The results can be extended then, by simply estimating and testing each of these popular functional forms.

⁶⁵This process of "blocking" or using a "class variable" to represent regional variation has the same statistical impact as using individual "dummy variables" for each of the different regions. Using the "class variable" approach simply produces the additional output of an overall significance level statistic (F statistic) for regions and substitute activities. In contrast, simply using individual dummy variables for each region may indicate that some regional dummy variables are significant while others are not. This picture is less clear concerning the overall significance of regions than the picture provided by the block variable approach.

Estimates of Marshallian Consumer Surplus

Estimates of Marshallian consumer surplus were calculated utilizing the travel cost demand estimates presented above. These estimates of consumer surplus are important because they form the basis for three of the four statistical tests performed. These estimates were calculated for each of the three functional forms, for a full and restricted sample, and using the actual and predicted number of trips.

Estimates of consumer surplus are presented below in Table 5-3. The titles "entire sample" and "restricted sample" appear in the far left column of Table 5-3. The entire sample refers to the group of 1,244 observations on which the regression results reported in Table 5-2 are based.⁶⁶ It is important to note, however, that the mean values of consumer surplus are impacted heavily by a small number of very active hunters who took a relatively large number of trips. In order to alleviate any concerns that findings of a difference in consumer surplus estimates based on actual and hypothetical data are due merely to these few individual observations associated with large consumer surplus estimates, welfare estimates were also derived using a sample with the number of trips taken restricted to thirty-two.⁶⁷

⁶⁶Although it is labeled the "entire sample", it does not include the one outlier identified using the studentized residuals procedure.

⁶⁷Individuals who took more than 32 trips were dropped from this restricted sample. This number, 32, is equal to the mean number of trips taken plus two standard deviations (see Table 4-1). The sample size, n , was reduced from 1,244 to 1,198.

In addition to consumer surplus estimates based on the entire and restricted samples, Table 5-2 also presents benefit estimates for each of the functional forms estimated in this study. As was discussed previously, the choice of a particular functional form has been found to have an important effect on estimates of consumer surplus. Therefore, in order to make the results found here as general as possible, welfare estimates were calculated for the log-linear, semilog, and linear functional forms. The terms "log-linear", "semilog", and "linear", appearing on the left side of Table 5-3 refer to these three functional forms.

The final variation in the comparison of the estimates of consumer surplus involve the distinction between basing the estimates on the actual number of trips reported as taken by the hunter or basing them on the number of trips predicted by the regression. As was discussed in chapter four Bockstael and Strand (1987) have found that, although using the actual number of trips is more appropriate than using predicted trips in many cases, there is likely to be some degree of systematic error associated with either choice.⁶⁸ This is true of the current study which is likely to be affected not only by omitted variables, but by recall error as well because the recall period is rather long (one year). Again the benefit estimates will be calculated based on both the "actual" and "predicted" number of trips in order to make the results of this study as general as possible.

⁶⁸Bockstael and Strand (1987) noted, that if the researcher chooses to base estimates of consumer surplus on the actual reported number of trips taken, and some of the error is due to measurement error, then benefits will be overestimated. Conversely, if the researcher bases welfare estimates on the predicted number of trips, then due to the effects of omitted variables benefits will be underestimated.

Table 5-3 presents mean estimates of the individual's total Marshallian consumer surplus associated with one year of hunting activity. This welfare measure, Marshallian consumer surplus, basically represents the value that individuals place on retaining access to the set of hunting sites they visited in 1985. Table 5-3 presents these benefit estimates in the following manner. The two columns of estimates are titled "Actual Consumer Surplus" and "Hypothetical Consumer Surplus". Actual consumer surplus refers to the welfare estimate which is based on the number of trips the survey participants report that they actually took in 1985 and the actual typical cost of those trips. The estimates of hypothetical consumer surplus are derived using the hypothetical market responses, where the survey participants state the number of trips that they would have taken at some hypothetical price provided to them in the survey.

Each line of Table 5-3 contains estimates of these actual and hypothetical benefit measures which differ in some crucial respect. Within this table all possible combinations of estimates based on the: entire and restricted sample sizes, three functional forms, and actual versus predicted trips are presented.

Table 5-3. Consumer Surplus Estimates

Sample Size	Functional Form/ Actual vs. Predicted Trips	Actual Consumer Surplus	Hypothetical Consumer Surplus
ENTIRE SAMPLE	Log-linear Actual Trips	642.65	737.19
	Predicted Trips	134.73	154.55
	Semilog Actual Trips	6670.04	13120.88
	Predicted Trips	1209.32	2378.89
	Linear Actual Trips	11711.52	35846.08
	Predicted Trips	4962.10	15187.78
RESTRICTED SAMPLE	Log-linear Actual Trips	527.95	591.73
	Predicted Trips	125.79	140.99
	Semilog Actual Trips	5824.85	11538.84
	Predicted Trips	1247.38	2471.02
	Linear Actual Trips	7145.57	21446.85
	Predicted Trips	4107.55	12328.47

Statistical Tests

The corresponding estimates of actual and hypothetical consumer surplus were compared using three procedures; the Wilcoxon signed-rank test, the sign test, and the Kolmogorov-Smirnov test. These three procedures test slight variations of the general hypothesis that the measures of consumer surplus based on hypothetical responses are equivalent to the corresponding measures derived from the actual behavior of respondents. In addition to these tests, an F-test which was not dependent upon the estimates of consumer surplus was conducted.⁶⁹ The F-test tests the hypothesis that the coefficient estimates derived using a pooled sample consisting of the hypothetical and actual data are not significantly different from the coefficient estimates derived by estimating a regression equation for each of the two samples independently. The results of these tests are presented in Table 5-4.

Table 5-4 consists of a matrix of test statistics. In total Table 5-4 contains the results of forty-two independent tests of the general hypothesis discussed above. These results are presented in the form of a test statistic, with one test statistic being reported in each cell of Table 5-4. The matrix is formed as follows. Across the top of this matrix are three broad columns, one for each of the three functional forms (Log-linear, Semilog, and Linear) used to estimate the demand for hunting recreation. Each of these three categories of functional form are then further subdivided into

⁶⁹For a review of each of these tests please see the section titled "Statistical Tests of the Null Hypothesis" in chapter four, or refer to Conover (1980), and Pindyck and Rubinfeld (1991).

estimates of consumer surplus based on the actual or predicted number of trips, which are represented as "AT" and "PT" respectively. The hypothetical and actual estimates of consumer surplus are then compared utilizing the signed-rank, sign, and Kolmogorov-Smirnov tests. These three tests are listed in the column titled "Type of Test", where "K-S" refers to the Kolmogorov-Smirnov test. Each of these comparisons of consumer surplus estimates produces a test statistic, which is reported in Table 5-4. These test results are not reported in the form of a significance value (p-value), because all but four of the forty-two test statistics reject the null hypothesis at the one percent significance level. Significance levels are reported for the four cases that failed to reject the null hypothesis at the one percent level. These four less significant results occur in the case of the Kolmogorov-Smirnov test in conjunction with the Log-linear form. These four cases report significance levels in bold inside of parentheses below the associated test statistic.

Table 5-4: Matrix of Test Results

Sample Size	Type of Test	Log-linear		Semilog		Linear	
ENTIRE SAMPLE		AT	PT	AT	PT	AT	PT
	Signed-rank	774390	770902	774390	767924	774390	774390
	Sign	1244	1240	1244	1230	1244	1244
	K-S	1.46 (.0276)	1.82 (.0026)	6.42	18.56	6.42	12.73
	F-Test	4.13		9.04		4.25	
RESTRICTED SAMPLE	Signed-Rank	718202	717354	718202	712216	718202	718202
	Sign	1198	1194	1198	1184	1198	1198
	K-S	1.29 (.0728)	1.39 (.0421)	6.54	18.71	6.54	13.59
	F-Test	4.80		9.41		7.55	

Table 5-4 presents a set of test results that point to a consistent rejection of the null hypothesis. Thirty-two of the forty-two tests reject the null hypothesis at the .01 percent level (.0001 level). The number of tests that can reject the null hypothesis grows to thirty-nine when the significance level is lowered to one percent, and all forty-two tests can reject at the ten percent significance level. This high rate of rejection of the null hypothesis provides evidence that individuals do indeed behave differently in actual and hypothetical markets for environmental public goods.

The test results will be discussed in detail for each of the four types of tests (Wilcoxon, F-test, etc.) conducted. The test statistics in the column referring to the Log-linear functional form and the actual number of trips (as opposed to predicted) will first be discussed for each test. This "example" statistic for each test will then be

compared to the other results garnered using that same test. The group of results for each test will be used to make some general conclusions concerning the null hypothesis.

The signed-rank test statistic appearing in the upper left corner of Table 5-4 has a value of 774,390. This is the matched pairs Wilcoxon signed rank-test statistic for the comparison of the corresponding "actual response" and "hypothetical response" estimates of consumer surplus, where these estimates are based on the log-linear form of demand estimation and the actual number of trips taken. This test statistic was calculated by taking the difference in the matched measures of consumer surplus.⁷⁰ If the two measures were equal then the difference between them would be zero. A positive sign would occur for an observation where the "hypothetical" measure of consumer surplus is greater than the "actual" measure. Next, the ranked-sign test ranks these differences according to absolute value, reassigns them their positive or negative algebraic sign, and then calculates the sum of all of the positive ranks and the sum of all of the negative ranks. The Wilcoxon test statistic, W , is the larger (in absolute value terms) of the sums of the positive and negative ranks.

A large Wilcoxon signed-rank statistic, W , leads to the rejection of the null hypothesis that there is no statistically significant difference in the median values of the corresponding hypothetical and actual consumer surplus estimates. The Wilcoxon signed-rank statistic ranges from a high of 774,390 to a low of 767,924, for the entire

⁷⁰They are "matched" in the sense that an "actual" and a "hypothetical" benefit estimate is calculated for each survey respondent.

sample ($N=1244$), and when the sample size is restricted to 1,198 these statistics range from 718,202 to 712,216. The test statistic value, 774,390, reveals that all of the differences had a positive algebraic sign, where a positive sign is associated with a hypothetical measure of consumer surplus which is greater than the actual consumer surplus estimate. The sum of a sequence of integers from one to n is equal to $N(N+1)/2$. This sum equals 774,390 when $N=1,244$, which implies that all of the ranks must have a positive sign for the test statistic to also equal 774,390. All twelve of the matched pairs Wilcoxon signed-rank test statistics reject the null hypothesis at the .01 percent significance level. The signed-rank results alone imply that it is very unlikely that individual responses in hypothetical and actual markets for hunting experiences are equivalent.

The sign test results are displayed on the row immediately below the Wilcoxon signed-rank results. The sign test statistic, for the comparison involving benefit estimates based on the log-linear form and the actual number of trips, appears in the left hand column on the second row and has a value of 1,244. The calculation of the sign test is similar to that of the signed-rank test. The sign test is calculated by taking the difference of each corresponding pair of actual and hypothetical consumer surplus estimates. In contrast to the Wilcoxon signed-rank test, the sign test ignores the relative sizes of the differences and simply assigns them the algebraic sign of the difference. The effect on the test statistic of a small or large difference is the same,

and therefore, the sign test is more appropriate in cases where the differences are neither normally distributed or symmetric.⁷¹

Another way in which the two tests are similar is that, as in the case of the Wilcoxon test, large (in absolute value terms) test statistics lead to the rejection of the null hypothesis that the median values of the corresponding hypothetical and actual consumer surplus estimates are equal. The sign test statistics range from 1,230 to 1,244 for the entire sample of 1,244 hunters, and the test statistics range from 1,194 to 1,198 for the restricted sample of 1,198 hunters.⁷² Each of these twelve test results reject the null hypothesis at the .01 percent level of significance. The sign test results lead to the consistent conclusion that the null hypothesis of equal median consumer surplus values should be rejected.

The Kolmogorov-Smirnov test statistics appear immediately below the sign test statistics. The test statistic of 1.46 applies to the comparison of consumer surplus estimates based on the log-linear form and actual trips. This cell that contains this test statistic also contains a significance level which appears below the test statistic, in parentheses, and in bold. This significance value of **(.0276)** means that the null hypothesis can be rejected at the 2.76 percent significance level. These parenthetical significance levels are listed only for the Kolmogorov-Smirnov test results, where the

⁷¹Both the Wilcoxon signed-rank test and the sign test are appropriate when the pairs of observations are matched, but the differences in them are not normally distributed. The sign test may be preferred to the Wilcoxon test however, in cases where the distribution of the differences is also not symmetrical.

⁷²The test statistic of 1,244 implies that for that particular comparison all of the differences had a positive algebraic sign.

consumer surplus estimates are based on the log-linear form and the actual number of trips, because the other eight Kolmogorov-Smirnov test results reject the null hypothesis at the .01 percent level (.0001 level).

The Kolmogorov-Smirnov test is quite different from the two tests discussed above. The calculation of the Kolmogorov-Smirnov test statistic is not based on "corresponding" pairs. Since the Kolmogorov-Smirnov test does not account for this correspondence, this test is less sensitive.^{73 74} The Kolmogorov-Smirnov test statistic calculation is also much more complex than the two tests discussed above. This calculation will not be reviewed here, but the reader can refer to Conover (1980) for a detailed description.

The Kolmogorov-Smirnov test statistic investigates whether the distribution of a variable is the same across different groups. The null hypothesis being tested in this case is whether there is a statistically significant difference between the distribution of consumer surplus values based on hypothetical responses and the distribution of values based on reports of actual behavior. If there is a large difference, between the sample of hypothetical consumer surplus estimates and the sample of actual estimates, then the Kolmogorov-Smirnov statistic will be relatively large, leading to a rejection of the null hypothesis. Eight of the twelve Kolmogorov-

⁷³This "less sensitive" test is not as likely to reject the null hypothesis for small differences between the actual and hypothetical consumer surplus estimates.

⁷⁴Since the Kolmogorov-Smirnov test does not account for the correspondence between actual and hypothetical benefit estimates, it may be less appropriate for this study.

Smirnov test results reject the null hypothesis at the .01 percent significance level. The remaining four Kolmogorov-Smirnov test results can reject at the 0.26, 2.76, 4.21, and 7.28 percent levels. The ability to reject at the ten percent level in all twelve cases leads to the reasonable conclusion that the two distributions of consumer surplus estimates probably do not arise from the same population.

In contrast to the three non-parametric tests described above, the F-test is a parametric test. The type of F-test used here is often referred to as a "Chow" test by economists and is sometimes employed to test for a structural break in a data set. Classic applications of this test involve time series data before and after some important event such as a major war. This test is employed here primarily because it provides a test which does not depend on the estimates of consumers surplus, but instead tests directly from the regression results.

The F-test statistics appear immediately below the Kolmogorov-Smirnov test statistics. Although each of the other three tests produced twelve test statistics, the F-test provides only six. This occurs because the other tests rely on the estimates of consumer surplus, which are calculated for both the predicted and actual number of trips, and therefore produce twice as many test results.

The F-test statistics range from 4.13 to 9.41. All six of these statistics are large enough to reject the null hypothesis at the one percent significance level.⁷⁵

⁷⁵The significance levels of one percent associated with the F-test, and of .01 percent associated with the other three tests do not necessarily represent a difference in the ability of these tests to reject the null hypothesis. The F-test significance level is restricted to one percent because these values were taken from tables that did not have a more stringent rejection level. The other three test significance levels could

The null hypothesis being rejected is that there is no significant difference between the (restricted) coefficients which are estimated utilizing a pooled data set consisting of both the hypothetical and actual responses and the (unrestricted) coefficients which are estimated using the separate hypothetical and actual data sets. The consistent rejection of this null hypothesis implies that there is a significant difference in the way in which people respond in actual and hypothetical markets.

The overall picture painted by this set of test results is one of consistent rejection of the null hypothesis. Thirty-nine of the forty two test conducted rejected the null hypothesis at the one percent significance level. The remaining three tests were able to reject the null hypothesis at the ten percent level. Given this consistent rejection of the null hypothesis it seems highly improbable that the responses to hypothetical survey questions are equivalent to reports of actual behavior. These results clearly provide substantial evidence that consumers behave differently in hypothetical and actual markets for environmental public goods.

reach .01 percent because they were calculated up to that level of preciseness by the SAS statistical software package.

Chapter Six

Conclusions

Chapter 6 will accomplish three objectives. First, this chapter will briefly summarize the first four chapters of this dissertation. The primary purpose of this summary is to refresh the reader with the context and motivation for this study. Second, this chapter will review the empirical results presented in chapter 5, determine what can be learned from these results, and discuss why these insights are important. A clear presentation of these insights and conclusions is the most important objective pursued in chapter 6. The conclusions are important because they will aid researchers in addressing both the comparability and calibration issues. Third, chapter 6 will conclude with a brief discussion of the questions which remain unanswered and suggestions for future research.

Summary

The contingent valuation method has become an increasingly popular method for determining the monetary value of environmental goods and natural resources. Along with the increase in usage of this hypothetical direct method has come a sharp increase in the degree of controversy concerning the performance of contingent valuation.

The increase in popularity of the contingent valuation method is due primarily to the following four strengths. First, by utilizing contingent valuation, the researcher can measure not only use values but non-use values as well. The ability to measure non-use values may be the most important advantage of this method. Second, the contingent valuation method is preferred by many environmental and resource economists because it has the ability to measure types of goods which cannot be measured by the more traditionally accepted travel cost and hedonic price methods (observable indirect methods). Examples of the types of goods which cannot be valued using the traditional observable indirect methods include protection of biodiversity and damages to structures and materials. The third notable strength of the contingent valuation method is its avoidance of the many assumptions which must be made in order to utilize the observable indirect methods. Finally, the fourth major strength of the method is the relative simplicity of using it to produce monetary values for public goods. This simplicity is a function of the direct nature in which monetary values are produced, they are simply stated by the respondent.

The controversy surrounding the contingent valuation method has arisen primarily because of its reliance on hypothetical data. The data is "hypothetical" in that the contingent valuation method asks individuals to make economic decisions in *hypothetical* markets. Decisions (mistakes) made in these hypothetical markets are costless because they are not associated with real economic consequences. Since these mistakes are costless, the respondent may not be as careful when making decisions in hypothetical markets as they would be in a corresponding real market.

This concern with hypothetical data has given rise to two issues which have sharply divided environmental economists, the comparability and calibration issues.

Simply stated, the comparability issue focuses on the question, are responses in hypothetical markets for environmental public goods comparable to real economic decisions. The calibration issue is closely related to the comparability issue, and it asks the following question. If hypothetical and actual valuation responses are not directly equivalent, is there some process of adjustment which could make them equivalent, and if this is possible by how much and in what direction should the hypothetical responses be adjusted? Recent publications by the NOAA commission on contingent valuation have recommended that hypothetical statements of value be adjusted downward by a factor of two. The NOAA calibration recommendations have been criticized by some economists as being premature due to a lack of empirical support (Randall 1994).

Most recent empirical studies which have addressed these two issues have utilized strictly private goods, and therefore, findings based on these studies may not be wholly transferable to the problem at hand which is whether hypothetical methods can accurately estimate the value of public environmental goods. This dissertation addresses an important need by providing additional empirical evidence concerning the comparability of hypothetical and actual decisions in the context of a public environmental good.

A number of potential sources of bias may affect hypothetical valuation methods by causing responses in hypothetical markets to diverge from actual

economic actions. The primary sources of bias include motives to misrepresent opinions, a lack of incentives to answer carefully (hypothetical bias), answers that reflect something other than a valuation, and an inadequate understanding of what they are being asked to evaluate.

One of the major strengths of this study is that it focuses its analysis on a particular type of hypothetical bias, which occurs when individuals fail to adequately consider either their budget or time constraints. This study is able to focus on the existence of an insensitivity to budget or time constraints bias, because it uses a hypothetical travel cost model which minimizes the effects of other potential sources of bias.

This study also has two other important strengths. First, the good being valued, deer hunting sites, is likely to be very familiar to the survey participants who actively hunted deer during 1985. These active deer hunters should be familiar with the costs of taking a trip, the characteristics of a site, and the benefits derived from taking a trip. This familiar good presents a best case scenario for the hypothetical method to produce valuation estimates that correspond closely to those derived using an actual behavior model. The second important strength of this study is that it utilizes a relatively large sample size of 1,244 respondents. Most other previous comparative studies have utilized relatively small sample sizes ($N < 100$) which reduced the power of their statistical tests.

This dissertation investigates an important issue, the comparability of hypothetical and actual responses in the context of a market for an environmental

public good. It addresses this issue in a rather unique manner which has several important strengths. Therefore, the conclusions which follow and the empirical results on which they are based should provide a significant addition to the understanding of this complex issue.

Results and Conclusions

The empirical results presented in chapter 5 consistently rejected the hypothesis that hypothetical and actual responses to valuation questions are equivalent. Four types of tests were used to test this hypothesis. These procedures included the matched pairs Wilcoxon signed-rank test, the sign test, the Kolmogorov-Smirnov test, and an F-test. In total there were forty-two individual comparative tests performed, which consisted of twelve signed-rank, sign, and Kolmogorov-Smirnov tests, and six F-tests. In all but three of the forty-two cases the null hypothesis was rejected at the one percent significance level, and in those three cases the null hypothesis was rejected at the ten percent level.^{76 77}

The *first conclusion* that can be drawn from these results is that responses in actual and hypothetical markets for a public environmental good are not directly

⁷⁶These results appear in Table 5-4.

⁷⁷The three results which were not significant at the one percent level were arrived at using the Kolmogorov-Smirnov test. It is no surprise that the Kolmogorov-Smirnov test, which does not account for the paired nature of this study's responses, would be somewhat less sensitive.

comparable. This is the central null hypothesis that was tested and rejected in this study. Furthermore, an examination of the estimates of consumer surplus, which were presented in Table 5-3, indicates that individuals' hypothetical market responses are systematically biased in an upward direction. Twelve pairs of the mean hypothetical and actual consumer surplus measures were calculated. In each of the twelve cases the hypothetical measure of consumer surplus exceeded the measure of consumer surplus which was based on actual economic choices.

The *second conclusion* that can be drawn in this study is that **the difference in hypothetical and actual responses is likely to be due to the respondents' insensitivity to their budget or time constraints.** The hypothetical and actual travel cost survey questions utilized in this study minimized the possibility that several other (widely cited) potential sources of bias would cause the hypothetical and actual results to differ.⁷⁸ As a result, the findings of responses in hypothetical markets which consistently overstate the actual willingness to purchase hunting trips, as represented in the reports of actual costs incurred and number of trips taken, are most likely attributable to a failure on the part of respondents to adequately consider their budget and time constraints when making hypothetical purchases. In the current study this insensitivity manifests itself in the form of an overestimation of the number of hunting trips that a hunter would take, if higher trip costs were actually incurred. Two

⁷⁸Please see chapter 2 for a detailed discussion of potential sources of bias. Additionally, refer to both chapters 2 and 3 for an analysis of the reasons that the hypothetical travel cost model used in this study is not likely to be affected by potential sources of bias other than an insensitivity to budget or time constraints.

qualifications need to be made concerning this conclusion, that individuals tend to overestimate their true willingness to purchase a good in a hypothetical market *because* they are insensitive to their budget and time constraints. First, this is not a testable hypothesis, but rather is simply an explanation which is assumed to be the most realistic explanation of respondent behavior. Restated, the empirical evidence strongly points to a divergence between hypothetical and actual responses, while an insensitivity to budget and time constraints is merely an assumed explanation for this observed phenomenon. Second, this study observed individuals overstating the quantity of trips that they would purchase at a given price. Although it seems quite reasonable, it is unknown whether similar behavior would occur in a situation where respondents were asked the maximum price that they would be willing to pay for a fixed quantity of a good. This is important because most hypothetical valuation methods require respondents to state their maximum willingness to pay for a good, as opposed to stating the quantity of a good (hunting trips) that they would purchase at some stated price.

The *third conclusion* that can be drawn in this study is that the current NOAA calibration recommendations may be premature and overly conservative.⁷⁹ It is important to note that this conclusion is extremely tentative in nature. One study does not provide enough information to concretely draw such a conclusion. This findings of this study do imply that care should be taken in this area and that more research is

⁷⁹This criticism of the NOAA recommendation may not apply to valuation studies involving non-use values. This study is limited to use values.

required concerning appropriate calibration factors. In January 1994 the draft NOAA rules for damage assessment included a suggestion that contingent valuation responses be calibrated downward by a factor of two. This recommendation implicitly assumes that contingent valuation responses tend to produce benefit estimates that are twice as large as the true measure of benefits. The findings of this study provide mixed results in reference to the validity of such an assumption. The NOAA assumption is not strongly supported by the findings of this study.⁸⁰ The estimates of Marshallian consumer surplus derived using the hypothetical travel cost model were greater than those produced using the actual behavior travel cost model in all cases, but the differences ranged from 12 to 206 percent (see Table 5-3). Additionally, the smallest differences in estimated benefit measures (12% and 15%) are associated with the log-linear functional form, which provided the best prediction of hunter behavior. The next best fit was found using the semilog functional form which produced hypothetical benefit estimates which were 97 and 98 percent greater than their actual counterparts. The linear model, which seemed to provide the poorest prediction of hunter behavior, was associated with differences of 200 and 206 percent. These mixed results indicate that more empirical evidence should be gathered before any strong conclusions regarding a calibration factor can be made. If the less accurate linear results are discounted, then it seems likely that the current NOAA recommendation of calibrating

⁸⁰Another empirical study that calls into question the NOAA recommendations is the recent working paper, "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods" by Carson, Flores, Martin, and Wright (1994). The comparisons in this study are also primarily use values.

hypothetical benefit estimates down by a factor of two may be too conservative. It seems reasonable to assume that a benefit estimate which is too small may be as damaging to a public policy decision as an estimate which is too large. Therefore, it might be best to consider hypothetical benefit estimates which have been adjusted downward by a factor of two to be a lower bound for an accurate benefit estimate, while unadjusted hypothetical benefit estimates could be considered an upper bound on the true measure of consumer surplus. A calibration procedure of this type, which would provide a reasonable upper and lower bound for the recreationists' true consumer surplus measure, may be more useful to a public policy decision maker than a specific point estimate of benefits.

The results presented in chapter five also reinforce the concern that environmental economists have concerning the travel cost method and the proper choice of a functional form for use in demand estimation. While the differences in actual and hypothetical benefit estimates ranged from 12 to 206 percent, the differences between actual travel cost benefit estimates which were based on different functional forms were much larger ranging from 23 to 1,722 percent.⁸¹ These differences reinforce the concern that environmental economists have concerning the choice among functional form when using a travel cost model, and in particular point to the weakness of the linear functional form for use in modeling recreation demand.

⁸¹The percentage differences between actual and hypothetical consumer surplus estimates were 15, 12, 97, 98, 200, and 206. The percentage differences among the three functional forms were 23, 76, 938, 1,003, 1,253, and 1,722.

In summary, the major conclusion of this study is that hypothetical benefit estimates are not directly comparable to estimates which are based on actual behavior. Respondents tend to overstate their willingness to purchase environmental public goods in hypothetical markets. Furthermore, this study has put forth a reasonable argument that this overestimation of purchases may be due to a failure on the part of respondents to adequately consider their budget and time constraints when they are asked to make a decision in a hypothetical market. Respondents may fail to adequately consider their budget and time constraints in a hypothetical setting because the penalties for mistakes which exist in real markets are absent. If this is an accurate description of respondent behavior then it seems likely that all types of hypothetical valuation methods, including the often used contingent valuation method, will provide estimates of consumer surplus which overstate the benefits associated with an environmental public good.

In addition to the primary conclusion of this dissertation that hypothetical and actual valuation estimates are not directly comparable, the findings of this dissertation also imply that further research is needed concerning the NOAA calibration suggestions which may be premature and overly conservative, and that the proper choice among various functional forms is a crucial step in implementing the travel cost method to estimate recreation benefits. The importance of the conclusions presented above regarding both the comparability and calibration issues underscores the need for additional research in the area of hypothetical valuation methodology.

Suggestions for Future Research

The conclusion that hypothetical valuation methods tend to overestimate the value of environmental goods is an extremely important finding. The importance of this finding provides a strong impetus for further research in this area. The types of research which might be most usefully pursued are discussed below.

This study has been referred to as being somewhat unique because the seldom employed hypothetical travel cost method was compared to an actual behavior travel cost model. The findings of this study cannot be easily replicated however, because there are very few hypothetical travel cost surveys which have been conducted. The first suggestion for further research then is to add hypothetical travel cost questions to future actual behavior travel cost surveys. This would allow researchers to further explore this study's findings. In particular, these comparative hypothetical travel cost questions should be employed in surveys which incorporate less familiar public goods or survey respondents which may differ from active hunters.

The second suggestion for further research is to investigate whether respondents in hypothetical markets are insensitive to their budget or time constraints, when they are asked to state the maximum price they are willing to pay, as opposed to being asked to state the quantity of a good they would purchase at some fixed price. This dissertation utilized the latter of the two by asking deer hunters how many trips they would take (purchase) given some hypothetical trip cost. The majority of valuation studies however, ask respondents to state their maximum willingness to pay

for an environmental good, such as a specified improvement in water or air quality. Although it seems reasonable that since respondents were insensitive to their budget and time constraints when confronted with a hypothetical quantity decision that they would also display this type of budget insensitivity when asked to state their maximum willingness to pay for a good, without a study designed to investigate this situation such an extension may not be warranted.

The third, and potentially the most valuable extension of this research, is to investigate the degree to which hypothetical and actual measures of non-use values correspond. This dissertation uncovered a divergence between hypothetical and actual values for an environmental good which consisted solely of use values. Although this characteristic makes it possible to more accurately determine the actual values, it also limits the degree to which these conclusions can be extended to other valuation studies. Very little research has been performed which evaluates the degree to which hypothetical estimates of non-use values correspond to the actual non-use values. This line of research is very challenging due to the difficulty of defining a criteria by which to judge hypothetical estimates of non-use values.⁸² The final suggestion for further research is to develop entirely new methods for estimating the values that individuals place on environmental goods. Perhaps there are strategies which can more accurately estimate the value of an environmental good. One such method

⁸²The increased difficulty in identifying a criteria occurs because non-use values cannot be measured using methods which rely on observable behavior. Potential methods for obtaining the non-use value of an environmental good using observable behavior include utilizing voting behavior or contribution games.

which shows promise is conjoint analysis.⁸³ Valuation methods which are currently being employed might best be thought of as the best available technology, which should be discarded if superior methods are developed.

⁸³For a brief discussion of conjoint analysis refer to chapter two of this dissertation.

Bibliography

Bibliography

- Acton, Jan Paul. 1973. "Evaluating public progress to save lives: the case of heart attacks". Rand Research Report R-73-02. Santa Monica, CA: Rand Corporation, January.
- Adamowicz, W., J. J. Fletcher, and T. Graham-Tomasi. 1989. "Functional Form and the Statistical Properties of Welfare Measures". American Journal of Agricultural Economics. 71:414-422.
- Adamowicz, W., J. Louviere, and M. Williams. 1992. "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities". Unpublished manuscript.
- Adamowicz, Wiktor, L. and Theodore Graham-Tomasi. 1991 "Revealed Preference Tests of Nonmarket Goods Valuation Methods". Journal of Environmental Economics and Management. 20: 29-45.
- Balkan, E. and J. Kahn. 1988. "The Value of Changes in Deer Hunting Quality: A Travel Cost Approach". Applied Economics. 20:533-539.
- Baron, Jonathan 1994. "Rationality and Invariance: Response to Schuman". Manuscript, Department of Psychology, University of Pennsylvania.
- Becker, G. 1965. "A Theory of the Allocation of Time". Economic Journal 75:493-517.
- Bell, Frederick, W. and Vernon R. Leeworthy 1990. "Recreational Demand by Tourist for Saltwater Beach Days". Journal of Environmental Economics and Management. 18:189-205.
- Bergstrom, J. C., J. R. Stoll, and A. Randall (1990). "The Impact of Information on Environmental Commodity Valuation Decisions." American Journal of Agricultural Economics. 72:614-621.
- Bishop, Richard, C. and Thomas A. Heberlein. 1990. "The Contingent Valuation Method" in Johnson, Rebecca and Gary V. Johnson eds. Economic Valuation of Natural Resources: Issues Theory, and Applications. Boulder CO: Westview Press.
- Bockstael, Nancy E., W. Michael Hanemann, and Ivar E. Strand. 1989. "Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models", report to the Economic Analysis Division, U.S. Environmental Protection Agency, Washington, D.C.

- Bockstael, Nancy E. and Kenneth E. McConnell. 1983. "Welfare Measurement in the Household Production Framework". American Economics Review, 73(4):806-814.
- Bockstael, Nancy E., Kenneth E. McConnell and Ivar E. Strand. 1991. "Recreation" in Braden J., B. and C.D. Kolstad eds. Measuring the Demand for Environmental Quality. New York: Elsevier Science Publishing Company.
- Bockstael, Nancy E. and Ivar E. Strand. 1987. "The effects of common sources of regression error on benefit estimates". Land Economics. 63(1): 11-20.
- Bohm, Peter. 1972. "Estimating Demand for Public Goods: An Experiment". European Economic Review 3:111-130.
- Brookshire, David, S. and Don D. Coursey. 1987. "Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures". The American Economic Review. 77 (September): 554-566.
- Brookshire, David, S., Alan Randall, and John R. Stoll. 1980. "Valuing Increments and Decrements in Natural Resource Service Flows," American Journal of Agricultural Economics. 62(3): 478-488.
- Brookshire, David, S., Mark A. Thayer, William D. Schulze, and Ralph, C. D'Arge. 1982. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches". The American Economic Review. 72 (March): 165-177.
- Cameron, Trudy, A. 1992. "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods". Land Economics. 68 (August): 302-317.
- Carson, Flores, Martin, and Mitchell. 1994. "Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods." Unpublished manuscript.
- Cesario, F. J. 1976. "Value of time in recreation benefit studies". Land Economics. 68(2): 291-297.
- Cesario, F. J. and J. L. Knetsch. 1970. "Time Basis in Recreation Benefit Studies" Water Resources Research. 6 (June): 700-704.
- Cicchetti, C.J., A.C. Fisher, and V. K. Smith. 1973. "Economic Models and Planning Outdoor Recreation. Operations Research. 21 (September/October):1104-1113.

- 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.
- Ciriacy-Wantrup, S. V. 1947. "Capital Returns from Soil-Conservation Practices". Journal of Farm Economics. 29: 1181-1196.
- Clark, David E. and James R. Kahn. 1989. "The Two-Stage Hedonic Wage Approach: A Methodology for the Valuation of Environmental Amenities". Journal of Environmental Economics and Management. 16: 106-120.
- Clawson M. 1959. "Methods of Measuring the Demand for and Value of Outdoor Recreation". Resources for the Future (February).
- Clawson, M. and J. L. Knetsch. 1966. Economics of Outdoor Recreation. Baltimore: John Hopkins University.
- Conover, W. J. 1980. Practical Nonparametric Statistics, Second Edition. New York: Wiley.
- Cooper, Joseph, C. 1993. "Optimal Bid Selection for Dichotomous Choice Contingent Valuation Surveys". Journal of Environmental Economics and Management. 24:25-40.
- Cooper, Joseph, C. and John Loomis. 1992. "Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models". Land Economics. 68(May):211-224.
- Coursey, Don, L., John L. Hovis, and William D. Schulze. 1987. "The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value". The Quarterly Journal of Economics. 102:679-690.
- Cropper, M.L. and A.S. Arriaga-Salinas. 1980. "Inner-city Wage Differentials and the Value of Air Quality". Journal of Urban Economics 8:236-254.
- Cummings, Ronald, G., David S. Brookshire, and William D. Schulze, eds., 1986. Valuing Environmental Goods: An Assessment of the Contingent Valuation Method. Savage, Maryland: Rowman & Littlefield.
- Cummings, Ronald, G., Louis A. Cox Jr., and A.M. Freeman III. 1986. "General Methods for Benefit Assessment". In Benefits Assessment: The State of the Art. Bentkover, Judith, D., Covelio, Vincent, T., and Mumpower, Jeryl ed. Dordrecht: D. Reidel Publishing Company.

- Cummings, Ronald G., and Glenn W. Harrison. 1992. "Identifying and Measuring Non-Use Values for Natural and Environmental Resources: A Critical Review of the State of the Art." Washington, DC : American Petroleum Institute.
- Cummings, Ronald G., Glenn W. Harrison, and E.E. Rutsrom. 1993. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?" Economics Working Paper Series, B-92-12, Division of Research, College of Business Administration, The University of South Carolina (Columbia: February 1993).
- Darling, Arthur, H. 1973. "Measuring Benefits Generated by Urban Water Parks". Land Economics. 49: 22-34.
- Davis, Robert, K. 1963. "The Value of Outdoor Recreation: An Economic Study of the Maine Woods" Ph.D. dissertation, Harvard University.
- Davidson, Paul. 1972. "The Valuation of Public Goods". In Economics of the Environment: Selected Readings. Dorfman, Robert, and Dorfman, Nancy, S. ed. New York: W. W. Norton & Company, Inc.
- Deaton, A. and J. Muellbauer. 1988. Economics and Consumer Behavior. Cambridge: Cambridge University Press.
- Desvouges, William H., F.R. Johnson, R. Dunford, S. Hudson, K. Wilson, and K. Boyle. 1993. "Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy in Contingent Valuation: A Critical Assessment." J.A. Hausman (ed), New York: North Holland.
- Desvouges, William H., V. Kerry Smith, and Matthew P. McGivney. 1983. "A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements". Research Triangle Park NC: Research Triangle Institute.
- Dickie, M., A. Fisher, and S. Gerking. 1987. "Market Transactions and Hypothetical Demand Data: A Comparative Study". Journal of the American Statistical Association. 82:69-75.
- Dobbs, I.M. 1993. "Adjusting for Sample Selection Bias in the Individual Travel Cost Method". Journal of Agricultural Economics. 101: 335-342.
- _____. 1993. "Individual Travel Cost Method: Estimation and Benefit Assessment with a Discrete and Possibly Grouped Dependent Variable". American Journal of Agricultural Economics. 75 (February): 84-94.

- Doran, Howard 1993. "Testing Cost Estimates". American Journal of Agricultural Economics. 75:(February): 95-103.
- Fishbein, Marin, and Icek Ajzen. 1975. Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Reading, Mass.: Addison-Wesley.
- Fisher, Anthony, C. 1994. "The Conceptual Underpinnings of the Contingent Valuation Method". Department of Agricultural and Resource Economics, University of California at Berkeley.
- Freeman, A. Myrick 1993. The Measurement of Environmental and Resource Values. Washington DC: Resources for the Future.
- Gineo, William, M. 1990. "A Conjoint/logit Analysis of Nursery Stock Purchases". North East Journal of Agricultural Economics. April: 49-57.
- Green, Trellis, G. 1989. The Economic Value and Policy Implications of Recreational Red Drum Success Rate in the Gulf of Mexico. Department of Economics and International Business, The University of Southern Mississippi.
- Greene, William, H. 1990. Econometric Analysis. New York: MacMillan Publishing Co.
- Hammack, J. and Gardner M. Brown Jr. 1974. Waterfowl and Wetlands: Toward Bioeconomic Analysis. Baltimore: The John Hopkins University Press for Resources for the Future.
- Hanemann, W. Michael. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses". American Journal of Agricultural Economics. 66 (August): 332-341.
- Hanemann, W. Michael and Edward Morey. 1992. "Separability, Partial Demand Systems, and Consumer's Surplus Measures". Journal of Environmental Economics and Management. 22: 241-258.
- Hanley, N.D. and R. J. Ruffell. 1993 "The Contingent Valuation of Forest Characteristics: Two Experiments". Journal of Agricultural Economics. 44(2): 218-229.
- Harrison, Glenn, W. 1992. "Valuing Public Goods with the Contingent Valuation Method: A Critique of Kahneman and Knetsch". Journal of Environmental Economics and Management. 23:248-257.

- Hausman, J. A. 1981. "Exact Consumer Surplus and Deadweight Loss". American Economics Review. 71:662-676.
- Hayes, D., J. Shrogren, S. Shin, and J. Kliebenstein. 1993. "Valuing Food Safety in Experimental Auction Markets". Iowa State University, Ames Iowa.
- Hellerstein, Daniel. 1992. "Estimating Consumer Surplus in the Censored Linear Model". Land Economics. 68 (February):83-92.
- Hicks, John, R. 1939. "The Foundations of Welfare Economics". Economic Journal. 49:696-712.
- Hotelling, Harold. Dated 1947. Letter to National Park Service in An Economic Study of the Monetary Evaluation of Recreation in the National Parks. United States Department of Interior. National Park Service and Recreational Planning Division. 1949.
- Kanninen, Barbara, J. 1993. "Optimal Experimental Design for Double-Bounded Dichotomous Choice Contingent Valuation". Land Economics. 69 (May): 138-146.
- Kaoru, Yoshiaki. 1988. A Discrete Choice Benefit Analysis of Marine Recreational Fishing: Does the Site Definition Matter? Unpublished Work.
- Kealy, Mary Jo, Mark Montgomery, and John F. Dovidio. 1990. "Reliability and predictive validity of contingent values: does the nature of the good matter?". Journal of Environmental Economics and Management. 19(3): 244-263.
- Kealy, Mary Jo and Robert W. Turner. 1993. "A Test of the Equality of Closed-Ended and Open-Ended Contingent Valuations". American Journal of Agricultural Economics. 75(May):321-331.
- Kirk, Roger, E., 1990. Statistics: An Introduction. Third Edition. Fort Worth: Holt, Rinehart, and Winston, Inc.
- Kling, Catherine, L. 1988. "Comparing Welfare Estimates of Environmental Quality Changes from Recreation Demand Models". Journal of Environmental Economics and Management. 15: 331-340.
- _____. 1992. "Some Results in the Variance of Welfare Estimates from Recreation Demand Models". Land Economics. 68(August):318-328.

- Knetsch, Jack L. 1963. "Outdoor recreation demands and benefits", Land Economics. 39(4): 387-396.
- Kopp, Raymond, J., and V. Kerry Smith. 1993. Valuing Natural Assets: The Economics of Natural Resource Damage Assessment. Washington DC: Resources for the Future.
- LaPiere, Richard, T. 1934. "Attitudes vs. Actions." Social Forces. 13:230-237.
- Larson, Douglas, M. 1992. "Further Results on Willingness to Pay for Nonmarket Goods" Journal of Environmental Economics and Management. 23 (September): 101-122.
- _____. 1993. "On Measuring Existence Value". Land Economics. 69(November):377-388.
- Lee, Linda, K., and L. Joe Moffitt. 1993. "Defensive Technology and Welfare Analysis of Environmental Quality Change with Uncertain Consumer Health Impacts". American Journal of Agricultural Economics. 75 (May): 361-366.
- Loomis, John, B. 1989. "Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Response". American Journal of Agricultural Economics. 71(February):76-84.
- Loomis John B., Michael Creel, and Timothy Park 1991. "Comparing Estimates from Travel Cost and Contingent Valuation Using Confidence Intervals for Hicksian Welfare Measures". Applied Economics. 23(11): 1725-1731.
- Loomis, John and Joseph Cooper. 1990. "Comparison of Environmental Quality-Induced Demand Shifts Using Time-Series and Cross-Section Data". Western Journal of Agricultural Economics. 15: 83-90.
- MacKenzie, John. 1993. "Comparison of Contingent Preference Models" Journal of Agricultural Economics 75 (August): 593-603.
- Matthews, Stephen, B. and Gardner S. Brown. 1970. Economic Evaluation of the 1967 Sport Salmon Fisheries of Washington. Technical Report No. 2. Olympia: Washington Department of Fisheries. April 1970.
- McConnell, K.E. 1992. "On-Site Time in the Demand For Recreation" American Journal of Agricultural Economics. (November):918-925.

- McConnell, K. E., and I. E. Strand Jr. 1981. "Measuring the Cost of Time in Recreation Demand Analysis". American Journal of Agricultural Economics. 63:153-156.
- Mendelsohn, Robert, John Hof, George Peterson, and Reed Johnson. 1992. "Measuring Recreation Values with Multiple Destination Trips". American Journal of Agricultural Economics. 74(November): 926-933.
- Milon, Walter, J. 1988. "Travel Cost Methods for Estimating the Recreational Use Benefits of Artificial Marine Habitat." Southern Journal of Agricultural Economics. 20(1): 87-101.
- Milton, J. S. and Jesse C. Arnold. 1990. Introduction to Probability and Statistics, Principles and Applications for Engineering and the Computing Sciences. Second Edition. New York: McGraw-Hill Publishing Company.
- Mitchell, Robert, C. and Richard C. Carson. 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method. Washington, DC: Resources for the Future.
- Morey, Edward, Douglas W. Shaw, and Robert D. Rowe. 1991. "A Discrete-Choice Model of Recreational Participation, Site Choice, and Activity Valuation When Complete Trip Data are Not Available". Journal of Environmental Economics and Management. 20: 181-201.
- Myers, Raymond, H. 1990. Classical and Modern Regression with Applications. Belmont, CA: Duxbury Press.
- National Oceanic and Atmospheric Administration. January 15, 1993 Federal Register. Natural Resource Damage Assessments Under the Oil Pollution Act of 1990. Pp. 4601-4614.
- National Oceanic and Atmospheric Administration. January 7, 1994. Federal Register. Natural Resource Damage Assessments Pp. 1062-1124.
- National Oceanic and Atmospheric Administration. March 25, 1994. Federal Register. Natural Resource Damage Assessments Pp. 14262-14273.
- National Oceanic and Atmospheric Administration. May 4, 1994. Federal Register. Natural Resource Damage Assessments Pp. 23098-23110.
- Neill, Helen, R., Ronald G. Cummings, Phillip T. Ganderton, Glenn W. Harrison, and Thomas McGuckin. 1994. "Hypothetical Surveys and Economic Commitments". Land Economics. 70(2):145-154.

- Olson, James, M. and Mark P. Zanna. 1993. "Attitudes and Attitude Change". Annual Review of Psychology. 44: 117-154.
- Ozuna, Teofilo, Jr., Kee Yoon Jang, and John R. Stoll. 1993. "Testing for Misspecification in the Referendum Contingent Valuation Approach". American Journal of Agricultural Economics. 75(May):332-338.
- Palmquist, Raymond, B. 1984 "Estimating the Demand for the Characteristics of Housing". The Review of Economics and Statistics. 64(3):394-404.
- Parsons, George, R. and Mary Jo Kealy. 1992. "Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation". Land Economics. 68 (February): 93-106.
- Parsons, George, R. and Michael S. Needelman. 1992. "Site Aggregation in a Random Utility Model of Recreation". Land Economics. 68 (November): 418-433.
- Pindyck, Robert S. and Daniel L. Rubinfeld. 1991. Econometric Models and Economic Forecasts. New York: McGraw-Hill Publishing Company.
- Prince, Raymond, Michael McKee, Ben-David Shaul, and Mark Bagnoli. 1992. "Improving the Contingent Valuation Method: Implementing the Contribution Game". Journal of Environmental Economics and Management. 23:78-90.
- Randall, Alan. 1994. "The Travel Cost Method". Land Economics. 70 February: 88-96.
- Randall, Alan, Berry C. Ives, and C. Eastman. 1974. "Bidding Games for Valuation of Aesthetic Environmental Improvements". Journal of Environmental Economics and Management. 1: 132-149.
- Randall, Alan and John R. Stoll. 1983. "Existence value in a total valuation framework", in Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas. Robert Rowe and G. L. (eds.) Boulder CO: Westview Press.
- Ridker, Ronald, G. 1967. Economic Costs of Air Pollution. New York: Praeger.
- Rowe, Robert D., Ralph C.D'Arge, and David S. Brookshire. 1980. "An Experiment on the Economic Value of Visibility". Journal of Environmental Economics and Management. 7:1-19.

- SAS Institute Inc. 1989. SAS/Stat User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC: SAS Institute Inc.
- SAS Institute Inc. 1989. SAS/Stat User's Guide, Version 6, Fourth Edition, Volume 2, Cary, NC: SAS Institute Inc.
- Schuman, Howard. 1994. "The sensitivity of CV outcomes to CV survey methods". Draft paper for the DOE/EPA workshop on Using Contingent Valuation to Measure Non-market Values, May 19-20, 1994.
- Schuman, Howard and Michael P. Johnson. 1976. "Attitudes and Behaviors", in Alex Inkeles, ed., Annual Review of Psychology 2:161-207.
- Seller, Christine, John R. Stoll, and Jean-Paul Chavas. 1985. "Validation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques". Land Economics. 61(May):156-175.
- Shavell Steven. 1993. "Contingent Valuation of the Nonuse Value of Natural Resources: Implications for Public Policy and the Liability System?" In Contingnet Valuation: A Critical Assessment. J.A. Hausman (ed), New York: North Holland.
- Silberman, Jonathan, Daniel A. Gerlowski, and Nancy A. Williams. 1992. "Estimating Existence Value for Users and Nonusers of New Jersey Beach". Land Economics. 68 (May): 225-236.
- Shaw, W. D. 1992. Searching for the Opportunity Cost of an Individual's Time". Land Economics. 68 (February): 107-115.
- Smith, V., Kerry 1986. "A Comparison of Direct and Indirect Methods for Estimating Environmental Benefits". American Journal of Agricultural Economics. 68(May): 280-290.
- _____. 1989. "Taking Stock of Progress with Travel Cost Recreation Demand Methods: Theory and Implementation". Marine Resource Economics. 6:279-310.
- _____. 1990. "Can We Measure the Economic Value of Environmental Amenities?" Southern Economic Journal. 56: 865-878.
- _____. 1993. "Nonmarket Valuation of Environmental Resources: An Interpretive Appraisal" Land Economics. 69 (February): 1-26.

- Smith, V. Kerry, and William H. Desvouges. 1986. Measuring Water Quality Benefits. Boston: Kluwer-Nijhoff Publishing.
- Smith, V. Kerry, William H. Desvouges, and Ann Fisher 1986. "A Comparison of Direct and Indirect Methods for Estimating Environmental Benefits". American Journal of Agricultural Economics. 68(2): 280-290.
- Smith, V. Kerry and Yoshiaki Kaoru. 1990. "What Have We Learned Since Hotelling's Letter? A Meta-Analysis". Economics Letters. 32: 267-272.
- _____. 1990. "Signals or Noise? Explaining the Variation in Recreation Benefit Estimates". American Journal of Agricultural Economics. 72 (May): 419-433.
- Strong, E. J. 1983. "A note on the functional form of travel cost models with zones of unequal populations." Land Economics. 59: 342-349.
- Thaler, R. 1993. "Mental Accounting Matters" Unpublished.
- Thayer, Mark, A. 1981. "Contingent Valuation Techniques for Assessing Environmental Impacts; Further Evidence". Journal of Environmental Economics and Management. 8: 27-44.
- Trice, A. H., and S. E. Wood. 1958. "Measurement of Recreation Benefits". Land Economics. 34 (February):195-207.
- United States. Department of Interior. 1985. National Survey of Fishing, Hunting, and Wildlife Associated Recreation. Washington DC: Fish and Wildlife Service.
- Vaughan, W. J., and C. S. Russell. 1982. "Valuing a Fishing Day: An Application of Systematic Varying Parameter Model." Land Economics 58 (November): 450-463.
- 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: 400-407.
- Whittington, Dale, V. Kerry Smith, Apia Okorafor, Augustine Okare, Liu Okore, L. Jin, and Alexander McPhail. 1992. "Giving Respondents Time to Think in Contingent Valuation Studies; A Developing Country Application". Journal of Environmental Economics and Management. 22:205-225.

- Willig, R.D. 1976. "Consumer's Surplus without Apology". American Economic Review 66:589-597.
- Willis, Ken and Guy Garrod. 1991. "Valuing Open Access Recreation on Inland Waterways: On-site Recreation Surveys and Selection Effects." Regional Studies. 25:511-524.
- Ziemer, R.F. and W. N. Musser, and R.C. Hill. 1980. "Recreation Demand Equations: Functional Form and Consumer Surplus." American Journal of Agricultural Economics 62:136-141.

APPENDIX

Appendix

1. Now I would like to ask you a few questions about your deer hunting last year. In total, how many trips did you take to hunt deer in 1985?

2. Did you personally bag one or more deer in 1985?

3. Think about what it cost you for a TYPICAL deer hunting trip last year. Include your expenses for such things as gasoline and other transportation cost, food, lodging, ammunition, and equipment rentals. If you went deer hunting with family or friends, include ONLY YOUR SHARE of the costs.

Keeping all those expenses in mind, how much did a typical deer hunting trip cost you, on average, in 1985?

4. Now suppose the cost of your deer hunting trips last year had been significantly higher, but the cost per trip for other kinds of hunting, fishing, and recreational activities had not changed.

At \$(2 x the amount in #3) per trip, how many deer hunting trips would you have taken in 1985?

5. If deer hunting had been so expensive that you did not go deer hunting at all, what would you have done instead?

VITA

David Michael Fletcher was born in Orlando, Florida on December 9, 1963. He graduated from East Bay High School in June, 1982, and in August, 1987 received the degree of Bachelor of Science in Business Administration in accounting from the University of South Florida. He then pursued a career as a tax accountant at Florida Progress Corporation. In August, 1990 he entered The University of Tennessee, Knoxville and in August, 1995 received a Doctorate of Philosophy degree in Economics.

He is presently employed as an Assistant Professor of Economics and Environmental Studies at St. Andrew's Presbyterian College in Laurinburg, North Carolina.