DISSERTATION

CHALLENGES AND SOLUTIONS IN COMBINING RP AND SP DATA TO VALUE RECREATION

Submitted by

Juan Marcos González-Sepúlveda

Department of Agricultural and Resource Economics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2008

COLORADO STATE UNIVERSITY

May 27, 2008

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY JUAN MARCOS GONZALEZ ENTITLED CHALLENGES AND SOLUTIONS IN COMBINING RPAND SP DATA TO VALUE RECREATION BE ACCPETED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on graduate Work

David Mushinski

Gorm Kipperberg

Andrew F. Seidl

Adviser, John B. Loomis

Department Head, Stephen Davies

ABSTRACT OF DISSERTATION

CHALLENGES AND SOLUTIONS IN COMBINING RPAND SP DATA TO VALUE RECREATION

Valuing resources that lack a market could be a complicated endeavor due to the lack of appropriate prices. Non-market valuation methods have been the tools used to compensate for this shortcoming in the process of incorporating such resources into the economic analysis. Without these methods we would overlook the importance that such goods and services have to society and bias the related policy recommendations we present as economists.

This dissertation looks at joining two of the most commonly used non-market valuation methods, namely, the Travel Cost Model (TCM) and the Contingent Valuation Method (CVM), and their application to valuing recreation in El Yunque National Forest in Puerto Rico. The combination of TCM and CVM in a joint estimator allows us to test the consistency between the two methods and uncover potential issues that each may be suffering from. The study finds that the geographical limitations of the study can cause underestimation of willingness to pay when using TCM. Furthermore, it shows that CVM can suffer from the same sampling issues as TCM when the samples are collected on site. Besides pointing out these problems, this work presents alternative ways in which they can be addressed. Finally, we provide another example that imposing a common underlying utility can significantly improve the joint use of these models.

Juan Marcos González-Sepúlveda
Department of Agricultural and Resource Economics
Colorado State University
Fort Collins, Colorado
Summer 2008

ACKNOWLEDGEMENTS

This research was possible thanks to the support of the National Science Foundation NSF grant # 038414.

Special thanks to my advisor and friend John Loomis. Your advice has inspired me to be not only a better economist, but also a better person. To my committee, Dr. Seidl, Dr. Kipperberg and Dr. Mushinski, for making this work so much better with your recommendations and support. To my colleagues and friends in the department, for sitting with me and help crack down some of the complicated challenges I was faced with during this process. To all of you, thank you.

DEDICATION

Mami, gracias por tomar en serio todas mis locuras.

Papi, gracias por enseñarme que la única pregunta tonta es aquella que no se hace.

Tío, gracias por abrirme los ojos a este camino y ayudarme en cada paso que he dado en él.

TABLE OF CONTENT

Introduction 1

A Joint Estimation Method to Combine Dichotomous Choice CVM Models with Count Data TCM Models Corrected for Truncation and Endogenous Stratification 9

Introduction 9
Alternative Ways to Combine TCM and CVM Data 13
Data 15
Likelihood Estimation 17
Estimating CVM Parameters 17
Estimating the TCM parameters 19
Simultaneous Estimation 21
Convolutions Method for Testing Differences in WTP 23
Testing Efficiency Gains of Joint Estimation 25
Results 26
Conclusions and future research 30
References 32
Appendix A. Deriving the Joint Density Function 37
Appendix B. Empirical Convolutions Method 40

Do CVM Welfare Estimates Suffer from On-Site Sampling Bias: A Comparison of On-Site and Household Visitor Surveys 41

Introduction 41
Testing for On-Site Sampling Bias in CVM 44
Endogenous Stratification in the CVM 45
Incidental Truncation in CVM 46
Adjusting the On-Site Conditional WTP to a Population Unconditional WTP 48
Method to Test for and Correct On-Site Sampling Bias in CVM 50
Data Sources 53
Results 54
Conclusions 57
References 58

Spatial Limits of the TCM revisited: Island Effects 61

Introduction 61
A Truncated Spatial Market 63
Methodology for Comparing TCM and CVM 64
Comparing TCM and CVM WTP Results 65
Simulating a Truncated Spatial Market 67
Empirical Applications 68
El Yunque National Forest, Puerto Rico 68

Snake River, Wyoming 70
Results 70
Conclusions 74
References 77
Appendix A. Results from individual regressions by site 84

A Utility Consistent Joint Estimation of Count Data and Dichotomous Choice Models 86
Introduction 86
Deriving the Utility Framework 89
Joint Estimation 95
Welfare Measures 98
Hypothesis Testing 99
Data 100
Results 102
Conclusions 104
References 106

Concluding Remarks 110

General Appendix A 112

General Appendix B 115
References 124

INTRODUCTION

Non-market valuation methods have been used by environmental economists for years. They represent an alternative approach to obtaining information on people's preferences about goods that are not suitable for the development of markets. Goods like clean air, standing forests, biodiversity, water clarity, open space, recreation and so on, provide important amenities and other non-market benefits for which we do not typically observe direct payments. Without markets, little information can be directly obtained about the benefits consumers enjoy from having such goods available and their willingness to trade such benefits for other goods.

Economic theory suggests that markets convey a great deal of information about trade offs through prices and their relative levels observed in transactions. Unfortunately, such transactions, and appropriate price ratios, are not observed in the case of public goods or those private goods that impose an externality (either positive or negative). In the former it is not possible to have a naturally occurring price due to the non-excludable and non-divisible nature of a public good. For the latter, prices are not a consistent reflection of the actual trade-offs that take place with the consumption of a good because there are costs or benefits that are not accounted for. In these contexts, prices are not present or lack the necessary information to truly reveal people's preferences among goods. Non-market valuation methods provide powerful tools to indirectly obtain this information.

Economists divide the different valuation methods into two general categories; revealed preference (RP) and stated preference (SP) method. The first one gathers and analyzes data from observed behavior that either relates the quantity consumed of the good of interest with an implicit price that would better capture the expenses incurred by consumers, or alternatively, looks at actual quantities consumed and prices paid for a related good either substitute or complement. The second method simulates a hypothetical market for the good and obtains information about people's behavior and preferences through stated responses.

Under the category of RP methods, economists typically use Travel Cost and Hedonic Models. Travel Costs Models use the cost incurred by visitors as the price they are willing to pay for the amenities they enjoy at a recreation site. Hedonic Models decompose prices of market goods (such as homes) into the marginal value of its components, including environmental amenities such as views, noise pollution levels, and so on.

In the case of SP methods, Contingent Valuation (CV) or Behavior (CB) Methods are the most commonly used[1]. These models get at people's willingness to pay or react to hypothesized changes in current market conditions, entirely new markets, or changes in policies. Their stated responses to randomized changes in price or environmental quality uncover people's preference structure and allow researchers to quantify the appropriate trade offs they would make if faced with such choices in real life.

None of these methods is free of criticism. SP methods have been criticized for being a way of obtaining a hypothetical answer to a hypothetical question (Bohm, 1972, Bishop and Heberlein 1979, 1986). Although several validation studies have shown that should not be the case (Carson et al. 1996; Loomis 1989), it is still believed that these

[1] Another type of SP method, Stated Choice Experiments, of which CV is a special case, is also a very used by economists (more so than CB).

models can be susceptible to various framing issues arising from the description of the hypothetical scenario. RP methods on the other hand, can be sensitive to the specification chosen and the definition of the travel cost variable. More importantly RP methods can never capture passive use values, which are important in many environmental contexts.

The Empirical Context

This dissertation is part of a broader project that tries to establish the relationship between human actions and hydrological and biological aspects of El Yunque National Forest in Puerto Rico. Surprisingly, the results from an individual TCM and CVM generated statistically different values for the sites studied[2]. This raised a question about the policy implications of having dramatically different results and how these differences can be explained and accounted for. Immediately, the traditional criticism presented above became the obvious target to explain this difference. Particularly, the hypothetical bias in SP methods and the TCM inability to capture passive use (non-use) values.

Despite this initial suspicion, the payment mode used in the SP question provided a framework that strongly suggests the information gathered is strictly limited to use values. This made both methods consistent in terms of the values they should provide. On the other hand, evidence in the literature seem to suggest that the hypothetical bias that can be found in the CVM should not as significant as the one found in this study and is, if anything, expected to yield lower willingness to pay (WTP) than RP methods (Carson 1996). Furthermore, because of the lack of a theory about the sources of hypothetical bias, among other factors, the ability to determine what may be responsible for this bias is greatly limited (Murphy et al. 2005). Instead, we look into other potential sources of

[2] The specific results are presented in the first chapter but it is worth mentioning now that individual use of these models yielded a value that was roughly 6 times higher for the SP method.

distortion and ways in which we can combine the two valuation methods to explain and deal with the challenges uncovered in this work.

In 1992, Trudy Cameron published a seminal paper that proposed a method to combine stated and revealed preference models. The idea was to complement each of these methods and enrich the information considered when getting at the bottom line of all these models, people's willingness to pay for a good or service that has no market. Her work used a joint estimation method that basically estimated a utility differential (characteristic of CVM) that was dependent on the consumption of a recreational site that was in turn determined by a conventional TCM.

Since then others have seen the opportunity to use this idea to enhance each of these models in different ways. From testing for consistency between data sets (McConnell et al. 1999) to incorporating non-use values to the estimations of revealed preference models (Eom and Larson 2006), economists have found use to the idea of combining valuation methods (Englin and Cameron 1996, Azevedo et al. 2003).

This dissertation follows the line of work that Cameron started in 1992. It uses the idea of joint estimation to address the marked difference obtained between separate TCM and CVM results for the valuation of several recreation sites in El Yunque National Forest in Puerto Rico. However, this work enhances that analysis that Cameron presented by updating the estimator she developed to use a more appropriate count data model in the TCM portion. Like McConnell et al. (1999), we look at the correlation of the errors in the two models and test whether the unobservable factors captured by them are statistically different. According to McConnell et al., a statistically insignificant correlation parameter would suggest that the preference structure that motivates the data generated by each model is different or somehow distorted. This gives us some grounds to look at each of these models independently and evaluate the problems that could be causing the difference.

In the TCM explorations, the manuscript looks at the spatial limitations that may be affecting the behavior observed from the local visitors on the island. We show that certain spatial market characteristic of the TCM, can be truncated, biasing the values obtained through such a model. In the case of Puerto Rico this truncation comes from the geographical limits of the island. As we move away from the site of interest we observe an abrupt end to the concentric circles that define the spatial market for the sites. In other words, we cannot observe trips from areas where the cost to visit is higher than the maximum cost associated to visiting from the farthest point of the island. This is particularly a problem since we cannot include non-local visitors due the difficulty to tease out the portion of their expenses that is attributable to consuming the site of interest. This problem is known as multiple destination bias (Smith and Knopp, 1980). Then, using the information on local's visiting decisions alone causes underestimation of the value obtained for the site and bias the valuation of other amenities at the site. In the case of the CVM no spatial market is required and the limitation described here is not present. However, we find that other problems could be affecting the CVM and its use in a joint model. This required looking deeper into what exactly we are obtaining from our CVM measure.

Despite having corrected the TCM estimation for on-site sampling problems, a common practice in the TCM literature, no effort was done initially to explore the need for a similar correction in the CVM. To our knowledge, very little has been done to explore the possibility that CVM estimates may suffer from a similar on-site sampling problem. The results show that it is possible that the values obtained from an uncorrected estimation of the CVM may significantly differ from a corrected TCM result and that jointly estimating the models can help in the identification of such a problem. Using the information in the TCM reveals the conditional nature of our CVM estimates upon observing a positive number of trips from users. This problem is known as incidental trunctation (Greene, 2003).

In the end, the research presented here makes use of economic theory to overcome the issues that we uncover. We decide to impose consistency through a common utility framework and try to get at a single WTP for sites and site characteristics from combining the two valuation methods. Contrary to Cameron however, we do not define a somewhat arbitrary utility functional form, but derive a consistent one from the uncorrected and corrected count data model for trip demand. We find that imposing a utility consistent framework not only results in a single measure of willingness to pay, but also a more statistically efficient one as well. By making use of the information conveyed in both the TCM and CVM data in this manner, we manage to reduce the confidence intervals around the parameters estimated by 84%. This, in turn, reduces the variability in the associated willingness to pay in a similar fashion.

This type of improvement in the precision of our estimates could have significant impacts on management decisions and benefit transfer applications. Accurate estimates of benefits can better inform maintenance decisions in the face of budget reductions as well as benefits foregone from closing the sites of interest. They would also provide more informative single point estimates for benefit transfers.

References

Azevedo, C. D., J. A. Herriges and C. L Kling "Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations." *American Journal of Agricultural Economics* 85 (Aug. 2003): 525-537.

Bishop, R. C. and T. A. Heberlein "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61 (1979) 926–930.

Bishop, R. C. and T. A. Heberlein "Does Contingent Valuation Work?" *Valuing Environmental Goods: A State of the Art Assessment of the Contingent Valuation Method.* R. Cummings, D. Brookshire and W. Schulze, eds., Totowa, New Jersey: Rowman and Allenheld, 1986.

Bohm, P. "Estimating the Demand for Public Goods: An Experiment." *European Economic Review* 3 (Nov. 1972): 111–130.

Cameron, T. "Combining Contingent Valuation and Travel Cost Data for Valuation of Nonmarket Goods." *Land Economics* 68 (Aug. 1992): 302-317.

Carson, R., N. Flores, K. Martin and J. Wright "Contingent Valuation and Revealed Preference Methodologies: Comparing Estimates for Quasi-public Goods." *Land Economics* 72 (Feb. 1996): 80–99.

Englin, J. and T.A. Cameron "Augmenting Travel Cost Models with Contingent Behavior Data: Poisson Regression Analyses with Individual Panel Data." *Environmental and Resource Economics* 7 (Sept. 1996): 133-147.

Eom, Youg-Sook and D. Larson "Improving environmental valuation estimates through consistent use of revealed and stated preference information." *Journal of Environmental and Economic Management* 52 (July 2006): 501- 516

Greene, William H. "Econometric Analysis" Prentice Hall 5th Edition, New Jersey, 2003.

Loomis, J.B. "Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses." *American Journal of Agricultural Economics* 71(Feb. 1989):76-84.

McConnell, K., Q. Weninger and I.E. Strand "Joint Estimation of Contingent Valuation and Truncated Recreational Demands." *Valuing recreation and the environment: Revealed preference methods in theory and practice.* Herriges, J. A. and C. L. Kling, eds. Cheltenham, Glos, England: Edlward Elgar Publishing Limited, 1999.

Murphy, J.J., P. Geoffrey Allen, T.H. Stevens and D. Weatherhead "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation" *Environmental and Resource Economics*. 30 (March 2005): 313-325.

Smith, V.K. and R.J. Kopp, 1980. The Spatial Limits of the Travel Cost Recreational Demand Model. Land Economics 56 (1): 64-72.

CHAPTER ONE

A Joint Estimation Method to Combine Dichotomous Choice CVM Models with Count Data TCM Models Corrected for Truncation and Endogenous Stratification

Introduction

In 1992 Cameron proposed a procedure that combined Revealed Preference (RP) and Stated Preference (SP) methods in a simultaneous estimation framework. The purpose of this was to allow communication between models and to arrive at a robust estimation of both sets of parameters. In Cameron's study, Contingent Valuation method (CVM) estimation is combined with a TCM in a structural way, allowing CVM parameters to be conditional to expected demand levels for each individual. This first attempt used a probit and a normal distribution joint process. The simultaneous estimation done in Cameron's paper relates the errors in both methods assuming a bivariate normal distribution, conditioning the probit part of the estimation to the error structure in the TCM portion. Since the publication of this paper, determining the consistency of SP and RP has become an important part in the recreation economics literature (Adamowicz, Louviere and Williams 1994; Azevedo, Herriges and Kling 2003; McConnell, Weninger and Strand 1999).

SP uses hypothetical scenarios to create or extend existing market conditions for a public good and assess marginal consumer behavior to changes in fees or quality. RP considers observed behavior from consumers to uncover a demand schedule, usually to arrive at the benefit consumers receive with the current price and quantity. These models are set up to look at different sides of the same problem. They differ in their approach, but aim to obtain the same information from survey data. TCM looks to estimate an ordinary demand function through which economists can calculate respondent's willingness to pay (WTP). While CVM obtains surplus measures directly looking at utility differentials between residual income and the visitor's stated behavior. Consistency between the two models requires that the site demand function and utility difference function come from the same underlying utility function (McConnell, Wenington and Strand). Traditionally, this theoretical expectation has been imposed through parameter restrictions (Cameron 1992) or conversion of one type of data to the other (McConnell, Wenington and Strand 1999; Loomis 1997; Englin and Cameron 1996).

When looked at individually, neither of the available methods under both types of models is free of criticism. SP models, typically developed in the form of Contingent Valuation methods, are of concern because of the hypothetical nature of the "transactions" used. Although several validation studies have been done (Bowker and Stoll 1988; Loomis 1989; Carson, et al. 1996) showing that CVM results provide welfare estimates that are comparable to RP results, criticism of CVM techniques have become more focused and direct overtime (Boyle 2003). RP models, typically in the form of Travel Cost Models (TCM) and Random Utility Travel Cost Models (RUM-TCM) are criticized because of the sensitivity of their welfare estimates to treatment of travel time and econometric issues.

However, both SP and RP have useful properties that aid researchers in their assessment of nonmarket values. SP models allow the researcher to explicitly evaluate policy relevant scenarios that may involve changes in resource quality beyond the levels observed in the RP data. This "data augmentations" approach avoids extrapolating beyond the range of the RP data when evaluating substantial improvements in environmental quality. RP data resembles what economists are used to dealing with when they estimate demand for a good that has a market. The fact that TCM behavior reflects actual decisions that involve real payments provides very useful information for the estimation process.

For these reasons we adopt the spirit of Randall's suggestion that we learn everything that can be learned from combining these data without imposing preconceived notions regarding the superiority of one type of data over another. As Azevedo, Herriges and Kling (2003) mention, discrepancies between the results obtained with these two methods need not be a failure of either one. On the contrary, these differences should be taken as an indication that the two sources are correcting the limitations that the other has.

For this research we also follow the spirit of Cameron's work, by combining CVM and TCM data to estimate joint parameters. Unlike Cameron's approach however, our attempt is primarily computational and does not use a combined utility function to channel the TCM model information into the CVM choice parameters. This leaves us with a joint error structure but eliminates the need for parameter restrictions as no utility function needs to be determined (thus, parameters are not to be constrained across equations). In a way, our approach looks at these equations as a pair of seemingly unrelated regressions where the connection between equations lies in the error structure rather than the parameters themselves. When using both models with the same group of respondents the unobservable factors that affect respondents' number of trips demanded are also likely to affect respondents' answers to the CVM question. These unobservable factors are contained in the error term of each model, suggesting that the errors of these models could be related (McConnell, Weninger and Strand 1999). Bringing the two models together and allowing a correlation parameter between their respective error terms provides a way to look into this possibility and test it. Again, the purpose here is NOT to impose but to test consistency.

This approach has a drawback that should be mentioned. That is, with no parameter restrictions we can *potentially* obtain two different welfare measures[3] for the same policy change. Despite this potential drawback, we see value in relaxing the assumption that both models respond to a particular underlying distribution arbitrarily chosen by the researcher. Estimating separate parameters and accounting only for the potential relation in the error structure of the models represents, in our view, a looser enforcement of the known theoretical relation between the models. Furthermore, in terms of applicability of the model, this would only be a problem really if any of the WTP measures changes the balance between the costs and benefits of a project.

Another contribution of this paper is to update the joint estimation process presented by Cameron by taking advantage of the evolution in parametric estimation models for TCM data. Fully parameterized trip frequency count data models have gained ground with the use of Poisson, Negative Binomial and Multinomial Count Distributions in recreation literature (Creel and Loomis 1990; Hellerstein and Mendelsohn 1993). They are seen as a logical extension to accommodate the particular properties of trip data (Shonkwiler 1999). In fact, it has been argued that the evolution of fully parametric trip

---

[3] Carson et al. used over 600 different CVM and TCM estimates and concluded that differences between CVM and TCM WTP were not statistically significant. If anything, CVM WTP measures are generally below TCM WTP estimates (roughly 0.9 of TCM estimates).

frequency models have made RP models trustworthy (Hellerstein 1999). With this in mind we use a Poisson and Negative Binomial distribution to exploit the count nature of the TCM data. Furthermore, these distributions are modified to account for on–site sampling, a problem also known as endogenous stratification.

To assess whether welfare calculations differ between individual and joint estimations we use an empirical numeric procedure known as *complete combinatorial convolutions*. Poe, Giraud and Loomis (2005) proposed this method as an alternative to empirically determine the probability that a random variable is statistically different to another. We recognize that individual's willingness to pay (WTP) in both CVM and TCM models is a random variable and test whether calculated consumer surplus changes significantly from one case to another (joint and individual estimation).

The following sections will expand on the econometric estimation process and the use of the convolutions method. Results and conclusions are also presented.

Alternative Ways to Combine TCM and CVM Data

There is a continuum of TCM and CVM questions, ranging from seasonal WTP for both (Cameron 1992) to marginal trips for both (Loomis 1997). Loomis (1997) combined TCM and CVM in a series of dichotomous choices. In this view, the revealed trip making behavior reflects an implicit yes to the first of the bid questions at existing travel cost, whereas the CVM question represents the second response to a higher bid in a panel. McConnell, Weninger and Strand (1999) also look at combining TCM and CVM by treating both as utility differentials. Like Loomis, McConnell argues that the original trip decision is an implicit yes to a first dichotomous choice question with a bid equal to the actual travel cost. His RUM argument is very appealing because it also allows for a change in the visitor's preference structure after more information about the site becomes available through a visit. Although useful, the problem with using such an approach is that you need to discard the trip frequency information from the TCM to be able to use it in a dichotomous choice panel context.

Others, like Englin and Cameron (1996), do the opposite, setting up the CVM question in a way that mimics the TCM framework. Their study looks at a change in trips in response to higher travel costs. The problem here is that asking visitors to reassess a full season of trips given a marginal change in price on site might be too much of a strain, thus becoming a source of possible bias or item non responses. The argument is basically the same used with open ended questions where respondents have a hard time pinpointing the actual WTP from a wide number of possible values. When visitors have to reassess the number of trips made in a season they are in essence asked to choose a new value for trip number in an open ended format. Certainly this problem becomes less relevant with visitors that have fewer visits as this limits the remaining number of visiting options available.

To our knowledge, the closest prior effort to test for consistency was done by Azevedo, Herriges and Kling (2003). However, three differences need to be pointed out. First, they use the same approach that Enlgin and Cameron (1996) had used before. They asked respondents to reassess their behavior for an entire season subject to marginal changes in costs. This provides a nice dataset where you can readily pool RP and SP responses into a single framework. In addition, Azevedo, Herriges and Kling (2003) still rely on the use of a censored normal distribution for their estimations. They do not take advantage of the newer, more appropriate count data models. They do this because this allows them to use a bivariate normal distribution which provides a familiar framework to correlate the errors between the two scenarios. The third difference lies in the way they test for consistency. Because they focus on the statistical discrepancies between TCM and CVM parameters, differences in the actual variable of interest, WTP, are not addressed directly.

The objective of this paper is to simultaneously estimate both models to take advantage of the commonalities between the two methods without: 1) discarding TCM trip frequency information, 2) forcing users to reassess their visits for the full season and 3) imposing consistency between the two models (e.g. instead, allowing testing for consistency). Our paper fills an important empirical gap in the analysis of combined RP and SP data: The case of TCM, with CVM on the most recent trip. This combination is often used in the literature. Examples of separate use of these particular data setup can be found in studies that range from deer hunting (Loomis, Pierce and Manfredo 2000), mountain biking (Fix and Loomis 1998) to recreation demand in developing countries (Chase et al. 1998).

Data

Data for this study come from a research project that is currently being conducted in El Yunque National Forest in the northeastern part of Puerto Rico. Surveys were administered during the summers of 2004-05 as part of a comprehensive study on the impact of site characteristics on social and physical conditions in and around the forest streams.

In person interviews were conducted at nine recreation sites along the Mameyes and Espíritu Santo rivers. Data include visitor's demographics, site characteristics (fixed and variable), trip information and a contingent valuation question in the form of; "Taking into consideration that there are other rivers as well as beaches nearby where you could go visit, if the cost of this visit to this river was $____ more than what you have already spent, would you still have come today?" Bid amounts ranged from $1 to $200 per trip.

Over 700 observations were obtained and coded, of which 450 observations were used in this analysis. The reason for the reduction in observations is because only trips where visiting the site was the main reason for traveling are considered valid for the TCM. This is done to deal with multiple destination problems (274 trips were not single destination trips) that are typically pointed out as a source of distortion in travel cost models. Also, because of the complicated form of the corrected negative binomial distribution, we eliminated four visitors who took more than 12 trips because they appear to be from visitors that are somehow quite different than the vast majority who take a small fraction of these trips. This is not uncommon, as pointed out by Englin and Shonkwiler (1995) where they too limited their corrected Negative Binomial to visitors with fewer than 12 trips.

The same variables were chosen in both models to be able to compare "apples and apples" between the two models. The variable travel cost was created from a set of variables available. Our definition of travel cost follows the conventional formula:

$$TC = (.33\times \text{ per minute income})\text{travel time} + \left(\frac{\text{gas cost}}{\text{\# of adults}}\right)$$

where travel time (in minutes) and gas costs are round trip measures.

The first term of this definition looks at the opportunity cost of the time spent in the trip (assuming this time was taken away from income generating activities). The second term looks at the actual cost of traveling to the site incurred by each adult in the household.

The following is a summary statistics of the sites studied and the variables considered. Table 1.1 presents the mean, maximum, minimum and the number of observations per site included.

The price variable in the TCM is of course travel cost as defined above. The bid amount visitors were asked to pay is the price variable in the CVM. Common explanatory variables for both models were mean annual discharge (as means of flow), distance the river pools were to the bridge access and road width (as a measure of accessibility).

Likelihood Estimation

Estimating CVM Parameters

Because CVM directly deals with consumer reactions to marginal changes they represent a straightforward way to obtain compensated welfare measures. In our study a dichotomous choice WTP question format is used. The welfare measure from a WTP question in CVM can be summarized in the following equation:

$$v(p^0,Q^0,y) = v(p^1,Q^1,y-c) \qquad (1)$$

where $v()$ is an indirect utility function, p^0 is the current price level of the good considered, Q^0 is the current quantity of the good consumed and y is income. On the other side of the equation, p^1 and Q^1 represent the new price and consumption level and c is the Hicksian compensating surplus, or WTP. In words, this equation states that maximum

WTP is the amount that makes utility levels equal when considering different price levels, quantities and disposable income. Note that under the current condition (*0*), disposable income is y, whereas in the alternative scenario (*1*), it is the difference between y and c.

What CVM allows us to do is to determine what the visitors' WTP is for the good in question. In other words, we uncover the population parameter c. In the case of recreation or site valuation the two levels available for consumption are typically all or nothing. Put differently, we uncover the WTP that makes the visitors indifferent between visiting a site or not on their most recent trip.

Because our WTP question format of "take it or leave it" involves a dichotomous choice of continuing to visit at the hypothetically higher travel cost or staying home, economists have used logit and probit likelihood functions to obtain WTP measures. This study uses a probit for the CVM portion of the parameter estimation. The general form of a probit likelihood function is derived from the Bernoulli distribution. A probit link is associated to ensure a nonnegative and bounded probability value (between 0 and 1) while conditioning the individual probability function to the set of parameters to be estimated.

$$lnL = y_{cvm} * ln(\pi) + (1-y_{cvm}) * ln(1-\pi) \qquad (2)$$

where $\pi = \Phi(X\beta)$ and y_{cvm} is the individuals response to the CVM question. It is important to point out that $\Phi(\,)$ stands for the standard normal cumulative density function; X refers to the set of variables we are conditioning our probability to and β is the set of parameters to be estimated. Among the set of variables X we have the bid amount or price increase per trip.

Estimating the TCM parameters

For the TCM portion of our estimation we use a Poisson and a Negative Binomial. Both these distributions are used in the estimation of recreation demand because they are count data models. This means that they take advantage of two important characteristics that count data share: non-negative and discrete outcomes. The Poisson and Negative Binomial distribution have been used successfully in the past to estimate seasonal demand for sites.

One important consideration that was raised by Shaw (1988) and later showed empirically by Creel and Loomis (1990) is that truncated versions of these distributions should be used when on-site sampling takes place. Truncation of the dependent variable arises because all visitors must take at least one trip to be sampled. In addition, we also correct for what is known as endogenous stratification or the fact that on-site sampling results in an over-representation of more frequent visitors in the sample data.

In general, correcting for truncation is done by dividing our probability distribution function by the probability of the ruled out (i.e., unobserved) outcomes. Analytically this could be represented as:

$$Pr(Y=y \mid y>a) = Pr(Y=y) \,/\, Pr(Y>a) \quad (3)$$

In our particular case $a = 0$, so:

$$Pr(Y=y \mid y>0) = Pr(Y=y) \,/\, (1-Pr(Y=0)) \qquad (4)$$

Note that because we are using count data models, we only need to find the probability that Y equals 0 and use its complement by subtracting it from 1.

When using the Poisson distribution, the resulting truncated version looks like:

$$Pr(Y=y \mid y>0) = \frac{\left(e^{-\lambda}\lambda^{y}\right)}{y!\left(1-e^{-\lambda}\right)} \qquad (5)$$

where $\lambda = e^{(X\beta)}$; and a resulting log likelihood function that can be represented in the following way:

$$ln\left(L_{poisson}\right) = -\lambda - ln\left(y_{TCM}!\right) - ln(1-e^{-\lambda}) \qquad (6)$$

Alternatively, the Poisson distribution has a very particular and useful property for correcting for endogenous stratification. That is that the truncated Poisson distribution provides the same results as using a regular (without truncation) Poisson when subtracting 1 from the dependent variable *Y*.

However, the Poisson imposes the restriction that the mean of the distribution equals its variance, something often rejected by trip data. A more general form of the Poisson count data that tests for and relaxes this mean-variance equality is the Negative Binomial model. The standard likelihood function of the on-site corrected Negative Binomial is:

$$L_{rp} = \left(\frac{\Gamma\left(\alpha + y_{TCM}\right)}{\Gamma\left(\alpha\right)\left(y_{TCM}-1\right)!}\right)\left(\frac{(\alpha)}{(\lambda+\alpha)}\right)^{\alpha} \frac{\lambda^{(y_{TCM}-1)}}{(\lambda+\alpha)^{y_{TCM}}} \qquad (7)$$

This was derived by Englin and Shonkwiler where again $\lambda = e^{(x\beta)}$ and α is the overdispersion parameter. In the case of the Negative Binomial distribution this convenient property of the Poisson for correcting for on-site sampling does not hold and a more complicated correction to the likelihood function is needed. See Englin and Shonkwiler (1995) for the derivation and expression.

Simultaneous Estimation

Using Cameron's structure we define our joint estimation process taking advantage of the known fact that a joint probability is equal to a conditional probability multiplied by a marginal probability:

$$f(x,y) = f(x \mid y) f(y) \qquad (8)$$

Just as in her case, we define the conditional probability in a direct manner by making the CVM estimation conditional to the TCM expected outcome. This expectation is used as an avidity measure to "inform" the CVM part of the estimation. It is however a statistical convenience to be able to incorporate a correlation parameter while treating the individual likelihood functions separately. This is an ad hoc approach that simplifies the estimation process by allowing the researcher to simply multiply the two probabilities obtained from our individual models. This in turn allows the possibility of simply adding the two log likelihood functions together when using a bivariate distribution.

When choosing the bivariate distribution for this application we are faced with a particular challenge. Because we use a count data distribution for our TCM estimation we cannot use a regular bivariate normal distribution as has been used in the past. To accomplish the simultaneous estimation of these equations we use the joint Poisson and Probit distribution derived in Cameron and Englin (1997). Although developed for a different purpose the joint density is ideal for the job. Unfortunately, Cameron and Englin (1997) did not derive a Negative Binomial and Probit joint density so another contribution of this paper is to derive such a joint estimator to incorporate overdispersion in our regression. Furthermore, both our densities are also modified to incorporate endogenous stratification. Appendix A shows in detail how these distributions were derived. Analytically, our new Negative Binomial joint likelihood function for the i[th] observation looks like:

$$L_i = \left[(\theta_i)^{y_{cvm,i}} (1-\theta_i)^{1-y_{cvm,i}} \right] \times \left(\frac{\Gamma(\alpha + y_{TCM,i})}{\Gamma(\alpha)(y_{TCM,i}-1)!} \right) \left(\frac{(\alpha)}{(\lambda_i + \alpha)} \right)^{\alpha} \frac{\lambda_i^{(y_{TCM,i}-1)}}{(\lambda_i + \alpha)^{y_{TCM,i}}} \quad (9)$$

where $\theta_i = \Phi\left((x_i\beta + \sigma\rho Z_i)/(1-\rho^2)^{0.5}\right)$ and $Z_i = \left(y_{tcm,i} - E(y_{tcm,i})\right) / \left(Var(y_{tcm,i})\right)^{0.5}$.

The log likelihood version of the joint estimation is simply the sum of the new CVM Probit likelihood and the chosen TCM likelihood function. The Probit portion is modified using the normalized TCM variable and accounting for individual variances and their joint covariance. Because we expect that the error term in the CVM equation should change with changes in the expected trip demand we also allow for this heteroskedastic process by setting $\sigma = \mathrm{e}^{(\gamma E(y_{tcm}))}$.

One point of clarification is necessary before finalizing this section. Special care must be taken when using the NB modified distribution. Because we are correcting it for endogenous stratification, the first and second moments used in the definition of Z are not the ones usually considered, but are also modified to account for the correction. Englin and Shonkwiler (1995) define these corrected moments for the Negative Binomial as:

$$E(y \mid y>0) = \lambda + 1 + \alpha_0 \qquad (10)$$

and

$$V(y \mid y>0) = \lambda + \alpha_0 + \alpha_0\lambda + \alpha_0^2 \qquad (11)$$

where $\alpha_0 = \alpha/\lambda$.

To summarize, we will estimate recreation benefits with three empirical models: (1) the dichotomous choice CVM estimated with a probit model; (b) the TCM using Poisson and NB; (c) a joint RP-SP model. From each of these models an estimator of net

WTP for a trip is calculated. Now we turn to evaluation of whether these benefit estimates are different from each other and their respective confidence intervals (CI) as a measure of the precision of the benefit estimates with each of the three methods. To do this we use a method proposed by Poe, Giraud and Loomis (2005) called empirical convolutions. The next section presents these methods and relates it to the task at hand, comparing the resulting WTP from all the models used in this paper.

Convolutions Method for Testing Differences in WTP

We use the method of convolutions to compare WTP estimates. Convolution is a mathematical operator that takes two functions and produces a third function that represents the amount of overlap between them. Poe, Giraud and Loomis (2005) proposed an alternative that can use a complete combinatorial approach to measure the difference between independent distributions. As mentioned before, convolutions create a third random variable that is formed by some relationship between the original functions considered. In Poe's example, this relationship is a difference between the two random variables of interest. This new random variable can be expressed as:

$$Z = X - Y \qquad (12)$$

Although several approaches have been used to assess differences between benefit estimates, some important issues are addressed with the use of the complete combinatorial approach. With this method we do not have sampling errors from using random sampling or overstate significance using Nonoverlapping Confidence Intervals. More importantly, this method does not require the assumption of normality for the difference parameter obtained.

The complete combinatorial method assumes that the researcher generates two independent distributions that approximate random variables *X* and *Y*. The way in which these empirical distributions are obtained does not affect the operation by which we determine the difference between them. Poe, Giraud and Loomis (2005) follow the argument that resampling methods approximate the underlying distribution of two independent random variables or calculated parameters. Each event in both distributions is given the same probability, although repeated outcomes are easily incorporated without losing generality. Poe, Giraud and Loomis (2005) showed that this empirical application can be related to the summation of polynomial products which, itself, goes back to the formal definition of the convolutions method. For more details on the approach used by Poe et al. you can refer to Appendix B. at the end of this paper.

In our study, *X* and *Y* refer to WTP vectors for the individual and joint estimations, respectively. A vector with random draws from the feasible values for each WTP is generated using the Krinsky Robb approach. A total of 4,000 draws were made and sorted. Each element of these vectors is subtracted from the other as suggested in Appendix B. This results in 4,000! possible combinations of the elements in both vectors. To obtain the one and two sided p-value the proportion of non-positive values is calculated. This represents the empirical probability that $\{x - y\} \leq 0$ or the area in one distribution that overlaps the other. We use the convolutions method to test consistency between CVM and TCM joint and individual estimation. This method aids us in looking beyond mean values in the WTP distributions and allows us to determine statistically whether the difference between the two estimation approaches is significant in the part that matters the most, surplus values.

Testing Efficiency Gains of Joint Estimation

As explained above the method known as convolutions allow us to assess the probability that two empirical distributions are different (whether $WTP_{joint}=WTP_{individual}$). In our particular case we want to test whether the distribution of the WTP obtained from a joint estimation is statistically different from the one obtained in the individual estimation process. This allows us to test whether simultaneous estimation yields significantly different benefit estimates. There are other important ways in which we can evaluate how different these results are from the ones obtained in separate regressions. For this matter we rely on more traditional hypothesis testing methods. That is, we use two different hypothesis tests to determine whether 1) the data generating processes of both equations are related in some way and, 2) if the resulting parameters for joint and individual estimations are equal. Formally this would be:

$$H_0: \rho = 1 \quad \text{and} \quad H_1: \rho \neq 1 \qquad (13)$$

$$H_0: \beta^{joint} = \beta^{individual} \quad \text{and} \quad H_1: \beta^{joint} \neq \beta^{individual} \qquad (14)$$

To determine whether to accept the null hypotheses in (13) and (14) we use the traditional t-test and likelihood ratio approach, respectively. We assess whether Rho is statistically different than one by using a t-test. To test equality of joint and individual coefficients we use the sum of log likelihoods of individual estimations against the joint estimation likelihood value. Together with the convolutions method, these set of tests should aid us in having a clearer idea of whether simultaneous estimation in this empirical case provides more efficient parameters.

Results

Results for the models estimated are summarized in Table 1.2. The values shown are the parameters estimated value and their corresponding (t-values). This table shows results for the individual and joint estimations using the Negative Binomial (NB) distributions, as preliminary statistical results indicated that the overdispersion parameter was statistically significant. This suggests that the Negative Binomial is closer to the actual data generating process and thus should be used rather than the Poisson when determining WTP.

As can be seen, theoretically consistent results were obtained for both TCM and CVM regressions. Results seem to suggest that our empirical case supports the theoretical expectation of negative slope parameters for travel cost and bid amount variables. The table not only reports the individual log likelihoods for the separate estimations, but also includes the sum of both TCM and CVM likelihood values. Results also suggest that the CVM and TCM results were very robust because all parameters from individual and joint estimations remain very close under the two estimation approaches.

One notable thing is that in both our separate and joint estimations, calculated WTP for CVM and TCM were considerably different. The two-tail p-value for the empirical convolution between the TCM and CVM WTP (for the joint and the individual Negative Binomial estimation) was around 0.02. This suggests that the disparity found between the two WTP measures is not an artifact of our joint estimation but instead could be the reason behind the little improvement found between these approaches.

Results for the likelihood ratio test performed between simultaneous and individual regressions are included in Table 1.2 also. The individual likelihood values for the separate regressions are reported along with the pooled log likelihood value. The difference between the sum of the individual log likelihoods and the simultaneous estimation likelihood is multiplied by 2 to obtain the likelihood ratio statistic χ^2 reported. The likelihood ratio value computed is not significant for the χ^2 test with two degrees of freedom (critical value for 90% confidence level equals 3.84). With regard to the hypothesis tests in (13) and (14), we see that in the joint estimation Rho appears an insignificant variable. Through both an insignificant Rho value and likelihood ratio for the joint model, the joint estimation process, as used here, does not seem advantageous in our case study over the separate regressions approach. Finally, our estimate for σ also appears to be insignificant suggesting that our error term in the CVM portion does not vary with changes in the expected number of trips.

As for the convolutions results, testing for significant differences in mean WTP, the most commonly used confidence levels (90 and 95%) are reported in Table 3. The values presented as maximum and minimum WTP in each case come from our convolutions method, thus these would vary in case of replication due to the random nature of the process. The mean values presented are the ones obtained directly from the parameters estimated using the appropriate WTP formulas.

On the other hand, we fail to reject the null hypothesis of equality or no difference in separately estimated versus joint estimation of TCM and CVM benefits. Note that p-value under this test represents the probability that the difference between the two empirical distributions is less or equal to zero. These results seem to reflect the small gain in efficiency obtained with the joint estimation process in the case for our data. In our table, the comparisons between the joint and individual empirical WTP variables appear, for all practical purposes, identical for both the TCM and the CVM. The similarity of consumer surplus estimates from the individual and joint models can be seen in the near equivalence of the Travel Cost coefficients in Table 1.2 In the individual Negative Binomial and Joint Negative Binomial model, the coefficients are again almost identical (-.1250 and -.1263) yielding consumer surplus per day of $8 and a price elasticities of -1.04 and -1.06 respectively. This also suggests that, in our particular dataset, we do not observe any significant connection between the two models through the unobservables. Perhaps, imposing a theoretical relation between the parameters in the models could increase the relation between the error structures.

Since all comparisons between joint and individual estimations show us a two-tail p-value close to 1 (0.97 for the TCM and 0.98 for the CVM) we can understand that the entirety of one of the distribution tails is covered by the tail of the other distribution, thus one empirical distribution lies on top of the other. It is important however to recall the huge difference in WTP between the two methods (CVM and TCM). Perhaps this is one reason why our results are not benefiting from the joint estimation. This may be what McConnell, Weninger and Strand called Transitory Preference Structure, where the unobservables between the two equations are not related. It is called this because it is assumed that under this situation visitors completely update their information set, leading to a new preference structure[4]. Other possible explanations include problems with the way the CVM question was presented and/or with the way travel cost was determined, but both of these were done following very standard assumptions.

---

[4] Although unlikely, it is worth mentioning that the nature of the rainforest under study causes significant and sudden changes in precipitation and water flow levels. These sudden changes can considerably alter the nature of the scenario faced by visitors when compared to the information individuals had at hand when they made their visiting decision.

An alternative explanation for the difference between TCM and CVM WTP could lie in the geographical characteristics of the studied area. Particularly because we are dealing with an on-site sample (no zeros or choke prices are observed), we may be facing a truncated spatial TCM market for the studied sites. We call this issue "Island effect" and it basically says that the implicit spatial market requirement in the TCM could be broken by the geographical limitations that the island dimensions impose, biasing TCM welfare measures downward. Because the maximum amount that local visitors are able to pay might be limited by the size of the island, the observed variation in the implicit price could be truncated in our application. In other words, we do not observe the full range of prices that locals are willing to pay for the services received at these sites hence the inverse demand function estimated by the TCM does not reflect the full benefit accrued to visitors on each visit. In fact, any point that should lie above the "spatial choke price" would not be observed. Instead, individuals with a WTP above the choke price would be found somewhere below their true demand points. If this is the case, not only our TCM WTP will be underestimated because of the portion missing above the spatial choke price but, because we are using fully parametric estimations, our results would suffer from further bias due to the fact that our methods will take the biased observations under the choke price as good and try to accommodate our regression results to them. Again, this could be a problem particularly because we do not observe zeros in our data hence we do not have observations on the price axis that would tilt our regressed demand curve up towards the real demand function.

Conclusions and future research

This paper provides an empirical modeling procedure that allows for testing whether joint estimation of stated and revealed preference models increase efficiency when compared to individual estimations and consistency between TCM and CVM responses. In our data, the CVM WTP question involved willingness to pay to visit the site under current conditions, a scenario quite conceptually similar to what is estimated with TCM. In this situation the improvement from joint estimation was quite small. However, joint estimation may result in larger and significant efficiency gains in the situation where the CVM WTP scenario deviates substantially from the existing situation in terms of quality of the site. Empirically testing this conjecture awaits suitably designed CVM and TCM datasets.

Another avenue of future research would be to integrate both models more, perhaps updating the joint utility theoretical approach that Cameron (1992) used to reflect the utility structure of count data models presented by Hellerstein and Mendelsohn (1993). Another alternative is to derive the expected constraints for different utility specifications and again use the simultaneous equation or estimation only to test which utility specification is supported by the data.

For this case our simultaneous estimation process can be seen as a general unconstrained version of Cameron's earlier work and opens the door to determine which type of joint preferences should be used prior to the actual constrained estimation. Due to the complexity of estimating a constrained utility theoretic specification, more information on the constraints that are supported by our empirical analysis should save researchers a great amount of effort while providing a better understanding of the behavior that guides both stated and revealed preferences.

At the methodological level, a contribution of this paper is updating the TCM portion of the joint estimation statistical technique used by Cameron to reflect the count data models now commonly used for recreational demand modeling. Using count data models represents an improvement over the original simultaneous estimation suggested by Cameron (1992).

References

Adamowicz W., J. Louviere and M. Williams "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities." *Journal of Environmental Economics and Management* 26 (May 1994): 271-292.

Azevedo, C. D., J. A. Herriges and C. L Kling "Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations." *American Journal of Agricultural Economics* 85 (Aug. 2003): 525-537.

Bowker, J.M. and J.R. Stoll "Use of Dichotomous Choice Nonmarket Methods to Value the Whooping Crane Resource." *American Journal of Agricultural Economics* 70 (May 1988): 372-381.

Boyle, K.J. "Contingent Valuation in Practice" *A primer on Nonmarket Valuation*, Champ, P., K.J. Boyle and T.C. Brown, eds. Netherlands: Kluwer Academic Publishers, 2003.

Carson, R., N. Flores, K. Martin and J. Wright "Contingent Valuation and Revealed Preference Methodologies: Comparing Estimates for Quasi-public Goods." *Land Economics* 72 (Feb. 1996): 80-99.

Cameron, T. "Combining Contingent Valuation and Travel Cost Data for Valuation of Nonmarket Goods." *Land Economics* 68 (August 1992): 302-317.

Cameron, T., Englin, J., 1997. "Respondent Experience and Contingent Valuation of Environmental Goods." *Journal of Environmental Economics and Management Volume* 33 (1997): 296-313.

Chase, L. C., D.R. Lee, W.D. Shulze and D.J. Anderson "Ecotourism Demand and Differential Pricing of National Park Access in Costa Rica." *Land Economics* 74 (Nov. 1998): 466-482.

Creel, M. and J. Loomis "Theoretical and Empirical Advantages of Truncated Count Data Estimating for Analysis of Deer Hunting in California." *American Journal of Agricultural Economics* 72 (May 1990): 434–41.

Englin, J. and T.A. Cameron "Augmenting Travel Cost Models with Contingent Behavior Data: Poisson Regression Analyses with Individual Panel Data." *Environmental and Resource Economics* 7 (Sept. 1996): 133-147.

Englin, J. and J.S. Shonkwiler "Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand Under Conditions of Endogenous Stratification and Truncation." *Review of Economics and Statistics* 77(February 1995):104-112.

Fix, P. and J. Loomis "Comparing the Economic Value of Mountain Biking Estimated Using Revealed and Stated Preference." *Journal of Environmental Planning and Management* 41 (1998): 227-236.

Hellerstein, D., 1999. "Can We Count on Count Models?" *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*, Edward Elgar, Cheltenham, UK.

Hellerstein, D. and R. Mendelsohn "A Theoretical Foundation for Count Data Models." *American Journal of Agricultural Economics* 75 (August 1993): 604-611.

Loomis, J.B. "Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses." *American Journal of Agricultural Economics* 71 (February 1989): 76-84.

Loomis, J. "Panel Estimators to Combine Revealed and Stated Preference Dichotomous Choice Data." *Journal of Agricultural and Resource Economics* 22 (1997): 233-245.

Loomis, J., C. Pierce and M. Manfredo "Using the Demand for Hunting Licenses to Evaluate Contingent Valuation Estimates of Willingness to Pay." *Applied Economics Letters* 7 (2000): 435-38.

McConnell, K., Q. Weninger and I.E. Strand "Joint Estimation of Contingent Valuation and Truncated Recreational Demands." *Valuing recreation and the environment: Revealed preference methods in theory and practice*. Herriges, J. A. and C. L. Kling, eds. Cheltenham, Glos, England: Edlward Elgar Publishing Limited, 1999.

Poe, G.L., K.L. Giraud and J.B. Loomis "Computational Methods for Measuring the Difference of Empirical Distributions." *American Journal of Agricultural Economics* 87 (May 2005): 353-65.

Randall, A. "Beyond the Crucial Experiment: Mapping the Performance Characteristics of Contingent Valuation." *Energy and Resource Economics* 20 (1998): 197-206.

Shaw, D. "On-Site Samples Regression: Problems of Non-negative Integers, Truncation and Endogenous Stratification." *Journal of Econometrics* 37 (February 1988): 211-223.

Shonkwiler, J.S. "Recreation Demand Systems for Multiple Site Count Data Travel Cost Models." *Valuing recreation and the environment: Revealed preference methods in theory and practice*. Herriges, J. A. and C. L. Kling, eds. Cheltenham, Glos, England: Edlward Elgar Publishing Limited, 1999.

Table 1.1 Summary Statistics by Site.

		Angelito Trail	El Verde Bridge	Jimenez Bridge	Jimenez Waterfall	Juan Diego	La Mina	La Vega	Puente Roto	Sonadora	Waterfall
Obs.		28	23	18	19	42	235	49	254	37	73
Mean Annual Discharge	Mean	1.272	1.282	1.632	0.219	0.106	0.238	1.667	1.365	0.273	0.665
Dist Bridge to Pool	Mean	0	0	145	52	93	35	0	0	5	30
Travel Cost	Mean	$17.54	$14.88	$14.57	$18.49	$13.72	$12.88	$18.09	$17.16	$24.77	$15.82
	Min	$1.65	$1.28	$1.41	$2.90	$1.24	$0.14	$1.41	$0.41	$3.05	$0.14
	Max	$55.09	$42.81	$61.50	$75.21	$63.88	$65.50	$71.88	$140.13	$92.19	$77.75
Bid	Mean	$83.93	$57.83	$69.44	$50.26	$86.19	$58.89	$63.35	$54.37	$62.43	$86.90
	Min	$5.00	$5.00	$5.00	$5.00	$5.00	$1.00	$1.00	$1.00	$5.00	$5.00
	Max	$180.00	$180.00	$200.00	$200.00	$200.00	$200.00	$200.00	$200.00	$200.00	$200.00

Table 1.2 Results from Individual and Joint Estimations

Variable	Separate Estimation		Joint Estimation
Negative Binomial			
Intercept	0.8022		1.0412
	(0.3342)		(0.4374)
TC	-0.1250***		-0.1263***
	-(2.9963)		-(3.1975)
Road	0.4683		0.3512
	(0.5610)		(0.4433)
Mean Annual Disch.	-3.8705		-3.8259
	-(1.5616)		-(0.9745)
Dist. Bridge to Pool	-0.0420		-0.0324
	-(0.8490)		-(0.7999)
Alpha	3.9238***		3.9131***
	(8.5631)		(8.1664354)
Probit			
Intercept	1.4846		1.3562***
	(1.3320)		(3.3690)
Bid	-0.0107***		-0.0095***
	-(8.8900)		-(4.5358)
Road	-0.0770		-0.0556
	-(0.5637)		-(0.8563)
Mean Annual Disch.	-0.0142		-0.0485
	-(0.0213)		-(0.2652)
Dist. Bridge to Pool	-0.0010		-0.0011
	-(0.1993)		-(0.4629)
Rho			0.0642
			(0.3901)
Sigma			0.0253
			(0.6286)
LL Neg. Bin.	-764.5786		
LL Probit	-236.4633		
LL Joint	-1001.0418		-1000.7478
Likelihood Ratio		0.588	
Implied WTP			
TCM	$8.00		$7.92
CVM	$110.12		$114.32

Results present coefficients and t-values.

* Significant at the 90% confidence level, ** significant at the 95% confidence level, *** significant at the 99% confidence level.

Table 1.3 Summary for Convolutions WTP Confidence Intervals for Individual and Joint Models.

	Joint			Individual		
TCM (Negative Binomial)						
CI	**MIN.**	**MEAN[a]**	**MAX.**	**MIN.**	**MEAN[a]**	**MAX.**
95	\$4.89	\$8.00	\$9.28	\$4.84	\$7.92	\$11.96
90	\$5.19	\$8.00	\$9.45	\$5.16	\$7.92	\$11.52
CVM (Probit)						
95	\$95.37	\$110.12	\$160.62	\$96.30	\$114.32	\$126.78
90	\$97.33	\$110.12	\$156.13	\$98.23	\$114.32	\$123.57

[a]Means are calculated using $1/\beta_{tc}$ for the TCM and $\beta_0 / abs(\beta_{bid})$ where β_0 is a grand constant term (it includes all non bid coefficients multiplied by the respective mean value of the variables). Minimum and maximum values come from the convolutions method. These represent the minimum and maximum values of the random WTP vectors generated and compared under each estimation type.

Appendix A. Deriving the Joint Density Function

The derivation of the Joint Negative Binomial and Probit distribution comes almost directly from Cameron and Englin (1997). In that article the authors look at two different random variables Z_1 and Z_2 and relate them to the Probit and Probit distributions. For a corrected Negative Binomial and a Probit we just need to follow Cameron and Englin's steps but change the Poisson portion for a corrected Negative Binomial. So we start by defining two random variables Z_1 and Z_2 and relating them to the following moments:

$$E[Z_1] = \lambda + 1 + \alpha; \qquad V[Z_1] = \lambda + \alpha + \alpha\lambda + \alpha^2;$$

$$E[Z_2] = 0; \qquad V[Z_2] = 1;$$

We can relate Z_1 and Z_2 to the variables of interest in the following way:

$$Trip = Z_1;$$

$$WTP = \sigma\left[\rho\left(\left(Z_1 - E[Z_1]\right)/\sqrt{V[Z_1]}\right) + \sqrt{\left(1-\rho^2\right)}Z_2\right] + \mu$$

It can be shown that the moments of these new random variables are:

$$E[Trip] = \lambda + 1 + \alpha; \quad V[Trip] = \lambda + \alpha + \alpha\lambda + \alpha^2;$$

$$E[WTP] = \mu; \qquad V[WTP] = 1;$$

The covariance between the two random variables is determined by:

$$E\left[\left(Trip - E[Trip]\right)\left(WTP - E[WTP]\right)\right] = E\begin{bmatrix}\left(Trip - E[Trip]\right)\times \\ \left(\sigma\left[\rho\left(\left(Trip - E[Trip]\right)/\sqrt{V[Trip]}\right) + \sqrt{\left(1-\rho^2\right)}WTP\right]\right)\end{bmatrix}$$

$$E\left[\left(Trip - E[Trip]\right)\left(WTP - E[WTP]\right)\right] = \rho\sigma\left(\lambda + \alpha + \alpha\lambda + \alpha^2\right)^2$$

Therefore the covariance term can be defined as ρ.

Now, the joint density of these two variables assuming independence between them is defined by:

$$f(Z_1,Z_2)=\left[\left(\frac{\Gamma(\alpha+Z_1)}{\Gamma(\alpha)(Z_1-1)!}\right)\left(\frac{(\alpha)}{(\lambda+\alpha)}\right)^{\alpha}\frac{\lambda^{(Z_1-1)}}{(\lambda+\alpha)^{Z_1}}\right]\left[\frac{1}{\sqrt{2\pi}}\exp\left(\frac{(-Z_2)^2}{2}\right)\right]$$

By solving *Trips* and *WTP* for Z_1 and Z_2, substituting in the joint density and scaling by the following Jacobian:

$$J=\left(1-\rho^2\right)^{-0.5}$$

we obtain the joint density function:

$$f(Trips,WTP)=\left[\frac{1}{\sqrt{2\pi}}\sigma^2\left(1-\rho^2\right)\right]^{-0.5}\left[\left(\frac{\Gamma(\alpha+Trips)}{\Gamma(\alpha)(Trips-1)!}\right)\left(\frac{(\alpha)}{(\lambda+\alpha)}\right)^{\alpha}\frac{\lambda^{(Trips-1)}}{(\lambda+\alpha)^{Trips}}\right]$$
$$\times\exp\left(\left(\frac{WTP-\mu-\sigma\rho\left(\left(Trips-E[Trips]\right)/\sqrt{V[Trips]}\right)}{\sqrt{\sigma^2\left(1-\rho^2\right)}}\right)^2\right)$$

Because it is easier to deal with a joint density that is defined in terms of a conditional and a marginal density function we try to do this for our joint distribution. We know that the marginal density of trips is defined as:

$$g(Trips)=\left(\frac{\Gamma(\alpha+y_{TCM})}{\Gamma(\alpha)(y_{TCM}-1)!}\right)\left(\frac{(\alpha)}{(\lambda+\alpha)}\right)^{\alpha}\frac{\lambda^{(y_{TCM}-1)}}{(\lambda+\alpha)^{y_{TCM}}}$$

By dividing our joint density *f(Trips, WTP)* by the marginal of trips we obtain the conditional density:

$$h(WTP\mid Trips)=\left[\frac{1}{\sqrt{2\pi}}\sigma^2\left(1-\rho^2\right)\right]^{-0.5}\times\exp\left(\left(\frac{WTP-\mu-\sigma\rho\left(\left(Trips-E[Trips]\right)/\sqrt{V[Trips]}\right)}{\sqrt{\sigma^2\left(1-\rho^2\right)}}\right)^2\right)$$

Note that this is the normal distribution with the following moments:

$$E[WTP \mid Trips] = \mu + \sigma\rho\left(\left(Trips - E[Trips]\right)/\sqrt{V[Trips]}\right)$$

$$V[WTP \mid Trips] = \sigma\left(1-\rho^2\right)$$

Now we have everything we need to define our joint density as a product of a marginal and a conditional probability function. Because our CVM response is really a latent variable we know that:

$$y_{cvm} = 0 \text{ if } (\text{Bid}<\text{WTP})$$
$$y_{cvm} = 1 \text{ if } (\text{Bid} \geq \text{WTP})$$

For this setup (and when assuming the variable WTP follows a normal distribution) we use a Probit link with a Bernoulli Distribution. With this final assumption we can present our joint likelihood function.

$$L_i = \left[(\theta_i)^{y_{cvm,i}}(1-\theta_i)^{1-y_{cvm,i}}\right] \times \left(\frac{\Gamma(\alpha + y_{TCM,i})}{\Gamma(\alpha)(y_{TCM,i}-1)!}\right)\left(\frac{(\alpha)}{(\lambda_i+\alpha)}\right)^{\alpha} \frac{\lambda_i^{(y_{TCM,i}-1)}}{(\lambda_i+\alpha)^{y_{TCM,i}}}$$

where $\theta_i = \Phi\left(\left(x_i\beta + \sigma\rho Z_i\right)/\left(1-\rho^2\right)^{0.5}\right)$ and $Z_i = \left(y_{tcm,i} - E(y_{tcm,i})\right)/\left(Var(y_{tcm,i})\right)^{0.5}$

Appendix B. Empirical Convolutions Method

The empirical convolutions method was first proposed by Poe, Giraud and Loomis. It uses all possible differences between randomly selected values of two random variables to determine the probability that these variables are statistically the same. Note that (12) can also be presented by adding the X distribution to the distribution of Y flipped around zero (thus obtaining the negative value).

$$Z = X + (-Y)$$

Assuming that the corresponding probability functions of X and Y are $f_X(x)$ and $g_Y(y)$ respectively, the distribution of their sum is represented by the following integral:

$$\begin{aligned} f \otimes (-g) &= h_z(z) \\ &= \int_{-\infty}^{\infty} f_x(z-(-y))\, g_y(-y)\, dy \end{aligned}$$

This expression provides the probability that each combination of the original function produces. This can be shown to be related to the sum of the product of each combination from a polynomial multiplication.

The complete combinatorial approach offers a simpler way to use the empirical convolutions method. The empirical distribution of the difference can be expressed as:

$$\hat{X}_i - \hat{Y}_j = \hat{X}_i + (-\hat{Y}_j) \qquad \forall i = 1,2,3...,m \qquad j = 1,2,3...,n$$

where each difference is given the same weight.

CHAPTER TWO

Do CVM Welfare Estimates Suffer from On-Site Sampling Bias: A Comparison of On-Site and Household Visitor Surveys

Introduction

On-site sampling is a useful and cost efficient sampling technique that has been used for years by recreation economists. Because visitors to a particular site represent a small portion of the total population, obtaining a large enough visitor sample from a general population survey can be an expensive and overwhelming task. On-site surveying represents an inexpensive alternative that ensures the sampling of groups that can, potentially, be most affected by policy or management decisions in recreational sites. The problem is that, as observed with almost every shortcut to a goal, on-site sampling benefits come at the expense of other sampling issues. These issues may force economists to refrain from using the statistical results of their surveys at an existing site to quantify the general population's value of a new proposed recreation site.

In 1988 Shaw recognized these problems and proposed a statistical correction to address them in the Travel Cost Model (TCM). The correction he proposed was meant to account for two important potential sources of bias when sampling visitors to recreational sites:

a) *Endogenous stratification*: Refers to the problem that arises when people that visit the studied site more frequently have a greater probability of being sampled and greater avidity for the site of interest.

b) *Truncation*: If the researcher is only relying on interviewing people that visited the site, the sample will be truncated at zero (no zero visits will be observed).

Throughout the years economists have embraced the idea that the correction proposed by Shaw provides useful information and allows them to extend the model conclusions to the general population. Relevant empirical applications of Shaw's correction in the TCM literature include the use of "corrected" count data models (Poisson and Negative Binomial particularly) to study deer hunting (Creel and Loomis; 1990), hiking (Englin and Shonkwiler; 1995), river recreation (Loomis; 2003), marine recreation (Bhat; 2003) and ecotourism (Chase et al.; 1998) among many others.

Although the issues covered with these corrections have been very common for TCM estimations, they have not been incorporated to other valuation methods commonly used by economists. Contingent Valuation Methods (CVM), at times, also rely on samples that are taken on site. This paper looks into the statistical concerns pointed out by Shaw and tests whether CVM suffers from them.

We argue that, even though CVM models are different from TCM in terms of the parametric distributional assumptions and the approach they follow to assess people's willingness to pay for a good, the sampling issues that affect TCM may still be present in CVM. Although the domain of the distributions assumed in most CVM studies are not directly affected by the nature of on site collected data, on-site sampling may over-represent more avid users, and certainly more avid than the general population and the sample frame is still truncated although incidentally. In on-site CVM samples, our responses are conditional to observing the visitor at the site. This implies that our observed answers are conditioned to having an individual with a strictly positive number of visits, i.e. a willingness to pay in excess of their current travel costs.

This paper takes advantage of a rare opportunity in which we have access to both a CVM visitor on-site survey and household survey. We use the data available to assess the willingness to pay from different sampling frames which include on-site sample, visitors from household survey, and total population from household survey. We then compare the resulting WTP from each sample and test whether the on-site CVM WTP shares the sampling problems explained above. To test for the effect of on-site sampling on willingness to pay we use an empirical convolutions method (Poe, Giraud and Loomis; 2005) that helps us to determine whether on-site sampling provides a statistically different measure of willingness to pay for visitors when compared to visitors in the general population survey. To test for incidental truncation we use an incidental truncation model (ITM) that incorporates the truncation in the TCM into the CVM estimation. Then we test for the statistical significance of the difference between the corrected CVM WTP using the ITM and the corresponding population analog measure. Finally, we propose a method to correct these problems in the CVM. The method proxies the first, best scenario where researchers would use the information provided by the TCM to statistically correct WTP measures in CVM estimations obtained using on-site samples. Of course, in many cases it may make little sense to use this best case because having TCM information may render this whole CVM effort unnecessary. However, when TCM data is not available we use the proportion of the population that visited the site to correct the CVM WTP and account for the conditional nature of the on-site CVM. We show that this correction improves our unconditional on-site visitor WTP measure and brings the corrected WTP into conformity with the population WTP estimates.

In the next sections we present a brief explanation of the issues (endogenous stratification and incidental truncation) and how they apply to the dichotomous choice CVM context. We then present the method we used to test whether these issues are present in the CVM case and discuss the data used to perform these empirical tests. After that we present a detailed explanation of the proposed correction. Finally, we show the results of our tests and correction as well as some concluding remarks and conclusions.

Testing for On-Site Sampling Bias in CVM

The two main problems discussed above (endogenous stratification and truncation) are really an issue when policy makers need valuation information on the general population and not just site visitors. When this is the case, the results obtained from an uncorrected estimation would provide higher WTP values that could mislead benefit transfers of proposed sites. Being able to correct our estimation process to account for on-site bias and obtain an unconditional measure of the population WTP (not conditional on being a visitor) expands the usefulness and applicability of primary valuation research. In fact, accounting for the bias caused by endogenous stratification and truncation can allow researchers to transfer WTP estimates with greater confidence that the results used are indeed a representative measure of their population preferences. With this in mind, verifying whether CVM estimates are susceptible to on-site bias and defining a correction for such a problem is an relevant addition to the nonmarket valuation literature.

Although the problems of endogenous stratification and truncation are typically found together, Martinez-Espineira, Amoako-Tuffour and Hilbe (2006) showed that when their impact on WTP estimates is separated it is truncation that has the larger part of the bias in on-site samples. If this is the case, focusing on (incidental) truncation should take care of the greater among both problems. Just as in their paper, we estimate a truncated model (although incidentally) and suggest that any remaining bias could be related to endogenous stratification. We expect that results from our empirical application will also show that (incidental) truncation is the source of most of the potential bias in on-site CVM samples.

Endogenous Stratification in the CVM

As mentioned above, endogenous stratification refers to the problem that arises when sampling on-site: that the researcher is more likely to sample frequent users that have greater avidity, hence higher WTP, for the site of interest. Statistically speaking, this means that the visitors intercepted have a different visit's probability distribution (Moeltner and Shonkwiler, 2005) violating the random sampling requirement to make results externally valid. Shaw, Englin and Shonkwiler (1995) found that not paying attention to such "sample-size" concerns could lead to wrong TCM welfare measures for single sites. Further efforts by Moeltner and Shonkwiler (2005) consider the same issues for multiple sites in a multivariate random utility model framework.

Discussion of endogenous stratification in the CVM literature is, to our knowledge, rare at best. In 1988 Nowell, Evans and McDonald recognized a length-bias sampling problem in CVM. They showed that not including the length of time people plan to stay at a site biases the WTP estimates. In essence, the claim is that, in an on-site sampling exercise, the probability of interviewing someone at the site is directly proportional to the length of the visit. In the Nowell, Evans and McDonald study, they show that a simple correction weighting the observations by the length of the stay can correct this problem and result in unbiased measures of WTP for visitors.

For the CVM trip frequency bias, a similar idea is presented. When sampling on-site, the probability of sampling an individual is clearly *conditional* to their decision to visit the site and the likelihood of sampling then is proportional to their number of trips. In other words, the individual must at least value its visit as much as the associated costs to get there. Like in the TCM, the systematic omission of non-visitors in the survey becomes a problem when the researcher is interested in using the results to say something about the general population or in benefit transfer to proposed sites. This might be encountered if one wanted to do a benefit transfer from the study site to a policy site, and only knew the size of the population around the policy site. If indeed non-visitors are not present in the sample due to their lower WTP we would expect to have a lower unconditional WTP. Nowell, Evans and McDonald already showed that, as long as certain conditions are met, this is true for visitors that spend more time on-site than others.

Incidental Truncation in CVM

Truncation occurs when the values of the independent variables in the model are known only when the dependent variable is observed (Ozuna, Jones and Capps; 1993). Indeed, when surveying on-site every person included in the sample will have at least one visit to the site and the researcher will be missing information on people that were not surveyed because they were not at the site when the sampling was done. Under these circumstances, using ordinary least squares parameters could lead to biased welfare measures if the group of interest in the study is different from site users. Incorporating this information to the parametric estimation can also result in more efficient estimators which, in turn, would provide more accurate welfare measures for changes in the goods considered.

Because dichotomous choice CVM models typically use probit or logit models the researcher might assume that the problem of truncation does not apply to this valuation method. Despite this, it should be remembered that CVM and TCM are different fronts of the same economic problem. With this in mind, using a sampling technique that causes problems in one would likely have some effect on the other. The concept of incidental truncation, widely accepted in the labor literature, sheds light on this issue by presenting the possibility that the distribution of a random variable might be conditioned on observing a related variable. In our case, the CVM answer is conditioned on observing the visitor at the site. Put differently, with on-site surveys we only observe a response for the bid question if the person has at least one visit to the site, requiring that WTP be greater than their travel costs.

Using the TCM and a dichotomous choice CVM conventional representation we can show analytically how our on-site CVM sample is incidentally conditioned to the TCM truncation. The general form of each individual model equation is as follows:

$$\text{TCM} \quad q_i = f\left(tc_{ij}, z_i, v_j\right) = x_i\beta + \mu_i \quad \text{where } x = \left(tc_{ij}, z_i, v_j\right) \text{ and,} \quad (1)$$

$$\text{CVM} \quad bid\_answer_i = g\left(b_i, z_i, v_j, \lambda_i\right) = z_i\delta + \varepsilon_i \quad \text{where } z = \left(b_i, z_i, v_j, \lambda_i\right) \quad (2)$$

Also, q_i represents the number of trips taken by individual *i*, tc_{ij} is the cost of traveling that the *i*th individual faced to visit site *j*, z_i is a set of individual characteristics, v_i is a set of site characteristics, bid_answer_i is the answer that individual i gave to the stated preference question, b_i is the bid amount offered to each individual, and λ_i is the on-site condition variable.

Relating these equations to the incidental truncation concept, we note that the expected value of the dependent variable in the CVM model can be expressed as:

$$\begin{aligned} E\left[bid\_answer_i \mid bid\_answer_i \text{ is observed}\right] &= E\left[bid\_answer_i \mid q_i > 0\right] \qquad (3) \\ &= E\left[bid\_answer_i \mid E\left[q_i\right] > 0\right] \\ &= z_i\delta + E\left[\varepsilon_i \mid \mu_i > -x_i\beta\right] \\ &= z_i\delta + \varsigma_\lambda \lambda_i\left(\alpha_\mu\right) \end{aligned}$$

where $\alpha_\mu = -x_i\beta/\sigma_\mu$ and $\lambda_i\left(\alpha_\mu\right) = \phi\left(-x_i\beta/\sigma_\mu\right)/\Phi\left(-x_i\beta/\sigma_\mu\right)$. As a consequence our incidentally truncated CVM to the population estimation recognizes that we have only visitors in our sample and allow us to account the effect this would have over the WTP.

Adjusting the On-Site Conditional WTP to a Population Unconditional WTP

In order to be able to generalize the WTP results from the on-site dichotomous choice CVM we have shown that something must be done to account for the fact that the on-site sample is conditioned on having a positive number of visits to the site. Generalizing the results obtained on-site can have a great deal of relevance in benefit transfer cases and, as shown here, could represent a considerable correction to the WTP estimates. An alternative and simple way to do this correction is the use of an adjustment factor that is equal to the percentage of a general population that would visit the site of interest. This fraction is multiplied by the on-site WTP, thereby adjusting it to represent the population value.

To justify the use of this adjustment factor we need to think of what we are obtaining when we calculate the WTP for the on-site sample. As previously mentioned, the on-site WTP calculation is a conditional random variable that depends on whether the respondent has visited the site of interest. The visitor WTP is in essence just part of the true population WTP for the site. Analytically we can say that:

$$E\left[WTP^{pop}\right]=\left(WTP^{v}\cdot P_{v}\right)+\left[WTP^{nv}\cdot\left(1-P_{v}\right)\right] \qquad (4)$$

where $E\left[WTP^{pop}\right]$ is the expected value of the population WTP, WTP^{v} is the net willingness to pay or consumer surplus that visitors have, P_v is the probability of being a visitor and WTP^{nv} is the net willingness to pay of non-visitors. The equation above says that the expected population WTP equals the WTP of each of the two possible groups (visitors and non-visitors) multiplied by the respective probability of being in that group. At first glance we can see that WTP^{v} is what is obtained with the on-site survey. Furthermore, we also know that, given the way our survey is setup, WTP^{nv} has to be zero. This is the case because non-visitors face a travel cost that is already higher than what they are willing to pay for their first trip. Due to the non-divisible nature of trips, non-visitors optimal choice is to not visit the site at all and hence have no consumer surplus or zero net WTP. With this in mind our equation above becomes:

$$E\left[WTP^{pop}\right]=\left(WTP^{v}\cdot P_{v}\right)+\left[0\cdot\left(1-P_{v}\right)\right] \qquad (5a)$$

or

$$E\left[WTP^{pop}\right]=\left(WTP^{v}\cdot P_{v}\right) \qquad (5b)$$

In conclusion, this shows that we can use the probability of being a visitor to adjust our conditional WTP measure and use our on-site results to infer something about the general population. Researchers can use this probability in new policy sites to transfer on-site WTP measures from other sites. This probability can be either an informed estimate based on existing studies like the USFWS National Survey, or simple survey information on what percentage of the population of interest might visit the new site.

Method to Test for and Correct On-Site Sampling Bias in CVM

To test whether there is endogenous stratification and truncation in dichotomous choice CVM we compare WTP results from an on-site sample to those obtained by using a household survey for the visiting group and corresponding population, respectively. By using a simple logit regression on both datasets we obtain the WTP for each. Once that is done, an empirical convolutions method is used to determine the statistical significance of any difference between the WTP measures. To do this we use the parameters obtained from using a representative population sample (using a sample that was not obtained on-site) and the corresponding standard errors to calculate a random vector of WTP (WTP_{pop}) with its own confidence interval. A similar random WTP vector is calculated $(WTP_{\text{on-site}})$ from the parameters and standard errors of the on-site CVM estimation. The convolutions method takes all possible differences between these two random vectors and determines the probability that they are different. Empirically, the percentage of values that are less than zero as a result of this convolution procedure are believed to be overlapping values in the corresponding distributions, and used to determine the empirical probability of finding the WTP amount in both distributions (empirical p-values). Formally, the study will first look to test the following:

$$H_o : WTP_{pop} = WTP_{on-site} \qquad (6a)$$

$$H_a : WTP_{pop} \neq WTP_{on-site} \qquad (6b)$$

Further tests can be performed to determine the effects that each of these sampling issues may have on the WTP values. For the endogenous stratification problem we utilize the population survey and look exclusively at the households that reported visits to the site of interest. We then use this portion of the household survey to run a logit regression and calculate their WTP. Because we focus on visitors from the household sample both the household visitor sample and on-site sample are still incidentally truncated. Nevertheless, the probability of sampling each observation is not the same under each group. In the on-site case, we face the aforementioned problem where the observed visitors are not a random sample of the general population and on-site visitors sampled are more avid. On the other hand, the visitor portion of the household was randomly selected from the population, eliminating the endogenous stratification issue. By comparing these two groups' WTP using the convolutions method we can test if indeed on-site CVM suffers from endogenous stratification.

Once we determine what effect endogenous stratification may have on on-site samples' WTP, we could attribute any remaining difference to incidental truncation. However, to make a more precise assessment we run an incidentally truncated CVM logit model and test whether the resulting WTP changes from that of the on-site sample. To do this we consider two equations. One describes the number of seasonal visits to a site (TCM); the other estimates the maximum amount visitors are willing to pay to visit a site (CVM). For the TCM, a Poisson distribution (corrected for truncation and endogenous stratification) is used. It is important to mention that the study uses White's Robust Variance Covariance Matrix to determine each coefficient's standard error. This corrects for the possibility of overdispersion. The specification for the TCM simply included an intercept and travel cost. We use the TCM part of the estimation to generate a variable named *Lambda* (following Greene's [2003] notation). As in (3) *Lambda* is defined as the *inverse Mills' ratio*. This is the ratio between the probability mass function and cumulative mass function of a distribution of each observation. The *inverse Mills' ratio* is meant to account for the incidental truncation nature of our CVM sample (a form of selection bias).

Our estimation follows Heckman's two-step estimation process. He proposed to start with an estimation of the extensive margin using a probit or logit model; then suggested using the estimated parameters to come up with *Lambda* or *inverse Mills' ratio*. *Lambda* is then included in the intensive margin model to account for the censored nature of the data. In our case, we start by estimating the intensive margin parameters (which are used to construct *Lambda*) then we include *Lambda* in the dichotomous choice CVM responses as one of the independent variables. With the empirical results from this regression we calculate another CVM WTP and compare it to the population WTP. Again, we use the convolutions method to evaluate whether any resulting difference in WTP is statistically significant.

The use of these three tests should provide us with enough empirical evidence to show whether this particular application of on-site CVM suffers from endogenous stratification and incidental truncation. In the next section we discuss about the data used to run these tests.

Data Sources

The Snake River in Jackson Hole, Wyoming was selected as the recreation site of interest for this analysis. This stretch of the Snake River south of Grand Teton National Park provides a wide spectrum of recreational activities. These activities included fishing from shore, fishing from boats, scenic raft trips, as well as hiking/fishing/jogging along the levees.

Visitor On-Site Sampling

Visitors to one of four areas along the Snake River were given a mail-back survey packet during weekdays and weekends during the month of August through Labor Day weekend in September of 2000. The four sampling locations included a boat put-in and take-out point used by private and commercial rafters, as well as two levee areas used for fishing, hiking, and jogging. A random sample of visitors was intercepted as they returned to their vehicles at each location. Visitor names and addresses were recorded so that a reminder postcard and second mailing of the survey to non-respondents could be performed. Only individuals over 18 years of age were requested to fill out a survey. We only had 19 refusals, for a refusal rate of just 3%. There were 657 surveys handed out and the overall response rate was 65%.

The same 12-page survey booklet was mailed to 800 randomly selected Teton County residents and 800 randomly selected Wyoming residents along with a $1 incentive on the first mailing. After two mailings, the response rate, net of undelieverables and deceased, for the sample of Teton County residents was 59% or 372 returned surveys. For Wyoming residents, the net response rate was 52.2%, 386 returned surveys.

The standard questions necessary to implement a TCM demand model as specified above were asked. This included gasoline costs, travel time, annual number of trips, time available for recreation, income, etc.

The dichotomous choice CVM recreation WTP question was asked immediately following the questions asking the respondent to record their trip expenses. The exact wording of the question was: "As you know, some of the costs of travel, such as gasoline, have been increasing. If the cost of this most recent visit to this section of the Snake River had been $X higher, would you have still made this visit? Yes No". The $X varied from a low end of $1 and $2 to a high end of $90 to $150.

Results

Four specifications are presented in Table 1. The first one shows the results for the regular on-site sample. The second one shows the results for only visitors within our household sample. The third and fourth show unconditional models that are meant to represent the parameters for visitors and non-visitors. In column three is our incidental truncation model which is our on-site visitor sample with the additional variable lambda that addresses the incidental conditioning of our CVM model to the TCM truncation. In this column we also present the TCM specification chosen as the first step of our estimation process. The specification used in this case uses only travel cost and an intercept, but includes no demand shifters. The fourth column has the parameters estimated for the household sample including non-visitors.

Results from our models shown at the bottom of Table 2.1 indicates that each sampling approach and model provide a different mean WTP estimates. These estimates range from $61.63 in the on-site visitor sample to $36.58 for the general population. As expected, mean WTP diminishes when we consider general population samples that include non-visitors. Additional relationships between resulting mean WTP calculations are: (1) there seems to be little difference between the regular on-site sample WTP and visitors in the household's survey suggesting endogenous stratification is not a problem in this sample despite the average number of trips being 43.24 for visitors and 10.30 for households; (2) in the incidentally truncated model the variable lambda is statistically significant and the resulting WTP from this model ($45.39) is much closer to the total population household sample ($36.58).

As expected, WTP from the incidentally truncated model carries most of the on-site bias in this case. From a total mean WTP difference of $25 between the on-site observations and the general population ones, we can attribute $16 to incidental truncation alone. This makes for 64% of the total difference and matches quite closely the results obtained by Martinez-Espineira et al. (2006) for the TCM.

Testing Differences in WTP

Because WTP is our variable of interest we compare at the calculated means for WTP with each model, and use the convolutions method to statistically test whether the WTP estimates are significantly different from each other. Table 2.2 shows the results of this process. Four pair wise comparisons were done between the alternative sampling frames. The first three comparisons look at the on-site sampling result versus the household visitors, corrected on-site and total household sample. The first result that is clearly observed is that the conditional on-site sample WTP is statistically different from the total household population WTP. The convolutions method tells us that only 3% of the times do these two WTP measures will coincide. This of course raises a flag for researchers that would want to use on-site WTP calculations to say something about the general population (non-visitors included) or transfer the on-site benefit estimates to a new proposed site which might have a different proportion of the population as visitors.

On the other hand, the second comparison shows that it may be acceptable to use this on-site WTP estimates for visitors. The similarity of the on-site WTP measure with the WTP of household visitors (90% similar when considering the two-tails p-value) provides compelling evidence that on-site surveys provide useful information about visitors to a site even for those that were not intercepted at the site and the sampled visitors take multiple trips.

With incidental truncation, the evidence is not as clear. On the one hand, our convolutions result shows that the resulting distribution for the WTP in our incidental truncation model lays almost exactly in between the on-site and the total household WTP. Differences between the ITM and these two other sample frame models are nearly identical (17% for the on-site versus ITM and 19% for the ITM versus the total household sample) and in both cases we have to reject that the resulting WTP is statistically different to the other. On the other hand, our lambda variable is significant to the 95% level which says that there is a statistically significant incidental truncation, but in this case it does not translate into statistically different WTP measures.

As for the simple adjustment factor approach, we obtain the percentage of visitors to the site by looking at the household sample and calculating the portion of the sample that visited the site of interest. We find reports of visits in almost 60% of the household surveys. Multiplying this percentage by the on-site WTP reduces the net WTP from $61.63 to $37.91. This surplus is almost identical to the one in the household general sample ($36.58) suggesting this simple correction to an on-site sample is a tenable approximation to a household sample.

Conclusions

Results from this study show that on-site dichotomous choice CVM WTP estimates have a conditional nature that has to be recognized when using the estimated parameters to infer WTP about the general population. Although it appears that endogenous stratification is not a problem in our data, the problem of incidental truncation appears to be an issue that can help to explain the difference between on-site and general population WTP. The proposed correction presented here is a useful tool to extend on-site sampling results to the general population. The mean WTP in our household sample was roughly $36 while the simple corrected on-site WTP was $38. We believe that, despite requiring some more information on the percentage of potential visitors to a policy site, the simple nature of the correction makes it a valuable addition to the dichotomous choice CVM model, specifically when considering benefit transfer to a new proposed site.

References

Bhat, M.G. "Application of non-market valuation to the Florida Keys Marine Reserve Managament." *Journal of Environmental Management* 67(2003): 315-325

Chase, L. C., Lee, D. R., Shulze, W. D., Anderson D. J. "Ecotourism Demand and Differential Pricing of National Park Access in Costa Rica" *Land Economics* 74 (1998): 466-482.

Creel, M., Loomis, J.B. "Theoretical and Empirical Advantages of Truncated Count Data Estimating for Analysis of Deer Hunting in California." *American Journal of Agricultural Economics* 72 (1990): 434-41.

Englin, J., Shonkwiler, J. S. "Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand Under Conditions of Endogenous Stratification and Truncation." *Review of Economics and Statistics* 77 (1995): 104-112.

Greene, William H. "Econometric Analysis" Prentice Hall 5th Edition, New Jersey, 2003.

Loomis, J. "Travel Cost Demand Model Based River Recreation Benefit Estimates with On-Site and Household surveys: Comparative Results and a Correction Procedure." *Water Resources Research* 39 (2003): 1105

Martinez-Espineira, R., Amoako-Tuffour, J., Hilbe, J. "Travel Cost Demand Model Based Recreation Benefit Estimates with On-site and Household Surveys: Comparative Results and a Correction Procedure: Reevaluation." *Water Resources Research* 42(2006): W10418.

Moeltner, K. and J.S. Shonkwiler "Correcting for On-Site Sampling in Random Utility Models." *American Journal of Agricultural Economics* 87 (2005): 327-39

Nowell, C, M.A. Evans, and L. McDonald "Length-Biased Sampling in Contingent Valuation Studies." *Land Economics* 64 (1988): 367-71

Ozuna, T., Jones, L.L. and O. Capps "Functional Forms and Welfare Measures in Truncated Recreation Demand Models." *American Journal of Agricultural Economics* 75 (1993): 1030-1035

Poe, G.L., Giraud, K.L. and J.B. Loomis "Computational Methods for Measuring the Difference of Empirical Distributions." American *Journal of Agricultural Economics* 87(2005): 353-65

Shaw, D. "On-Site Samples Regression: Problems of Non negative Integers, Truncation and Endogenous Stratification." *Journal of Econometrics* 37 (1988): 211-223.

Table 2.1 Results from Dichotomous Choice models for the Snake River area.

	On-site visitors	Visitors Household Survey	Corrected On-Site Visitors	Household Population
TCM				
Intercept			3.8869***	
			(0.0069)	
Travel Cost			-0.0092	
			(0.0069)	
CVM				
Intercept	0.4446	0.8339***	0.2839	-0.7057***
	(0.2835)	(0.231)	(0.2868)	(0.1689)
Bid	-0.0197***	-0.0208***	-0.0238***	-0.0159***
	(0.0054)	(0.0039)	(0.0061)	(0.0033)
Income	6.64E-06**	1.05E-06	6.34E-06**	7.47E-06***
	(2.75E-06)	(2.02E-06)	(3.01E-06)	(1.69E-06)
Lambda			1.8936**	
			(0.6762)	
Log likelihood	-113.297	-226.07	-105.771	-345.639
Pseudo R2	0.1082	0.0983	0.1544	0.0749
Mean WTP	$61.63	$59.77	$45.39	$36.58

* Significant at the 90% confidence level, ** significant at the 95% confidence level, *** significant at the 99% confidence level.

Table 2.2 Willingness to pay convolutions and p-values for pair wise differences between samples.

	On-Site Visitors WTP			On-Site Visitors WTP			On-Site Visitors WTP			ITM WTP[a]		
Confidence Int.	MIN.	MAX.	MEAN	MIN.	MAX.	MEAN	MIN.	MAX.	MEAN	MIN.	MAX.	MEAN
95	$ 44.15	$113.50	$61.63	$ 44.23	$116.90	$61.63	$44.43	$113.60	$61.63	$33.16	$82.84	$45.71
90	$ 46.48	$ 99.73	$61.63	$ 46.40	$100.50	$61.63	$46.32	$ 98.83	$61.63	$34.78	$74.56	$45.71
	Total Household WTP			Household Visitors WTP			ITM WTP[a]			Total Household WTP		
Confidence Int.	MIN.	MAX.	MEAN	MIN.	MAX.	MEAN	MIN.	MAX.	MEAN	MIN.	MAX.	MEAN
95	$ 28.04	$ 54.99	$36.58	$ 46.50	$ 85.26	$59.77	$32.28	$ 81.00	$45.71	$28.14	$56.43	$36.58
90	$ 29.16	$ 51.05	$36.58	$ 48.04	$ 79.13	$59.77	$33.96	$ 72.49	$45.71	$29.13	$51.55	$36.58
Convolutions Result												
P-value (one side)		**0.03**			**0.45**			**0.17**			**0.19**	
P-value (two side)		**0.06**			**0.90**			**0.34**			**0.38**	

Minima and maxima were obtained using the convolutions method described by Poe et al. (2005). Mean WTP values were calculated using the mean WTP formula for logit models $WTP = \ln\left(1 + \exp^{\text{grand intercept}}\right) / \left(\text{-bid coefficient}\right)$.

[a] ITM WTP is the Incidental Truncation method adjusted WTP

CHAPTER THREE

Spatial Limits of the TCM revisited: Island Effects

Introduction

The ideas behind the Travel Cost Model (TCM) were first suggested by Harold Hotelling in 1949 and first implemented by Marion Clawson and Jack Knetsch (1966). The model recognizes that recreation sites, even when people do not pay entrance fees, have an implicit price that stems from the travel costs involved with visiting the site. The idea of using *travel cost* served to develop a demand-based model (analog to those commonly used in regular goods' demand) that could be estimated to value recreational uses of the environment (Parsons, 2003). Implicitly then, the TCM also relies upon the notion of a spatial market where visitors' willingness to trade travel costs for site visits reveals their willingness to pay (WTP) for the site and its characteristics. By looking across people who live at different distances from the recreation site and hence face different travel costs, the model allows researchers to estimate a "revealed" demand curve for a site and its components.

More than 20 years ago Smith and Kopp (1980) raised the issue of spatial limits of the TCM. They showed that differing spatial limits could considerably affect the WTP derived from the model. Since then, the TCM literature has shifted its attention from these spatial concerns and has mainly focused on determining the travel cost incurred by each visitor. Recent efforts include studies that look at the opportunity cost of time (Larson and Shaikh, 2001), latent separability of costs (Blundell and Robin, 2000) and how to separate on-site time from travel time (Shaw, 1992; McConnell, 1992). In addition, past research has focused on the assumptions of the TCM that distant visitors actually incur the travel cost exclusively to visit the site of interest (the so-called multiple destination trip bias problem) (Haspel and Johnson, 1982; Mendelsohn et al., 1992), but very little research has focused on physical or natural spatial limits to the travel cost model. The closest concern in using TCM is in urban recreation settings where there may be insufficient variation in travel costs to fully reflect a visitor's WTP (Loomis and Walsh, 1997).

A similar, but somewhat different problem arises in the case of recreation that take place on small islands such as Hawaii, Puerto Rico, Jamaica, etc., i.e., islands with significant resident populations that make primary purpose visits to high quality local sites. The difficulty is that the maximum travel cost that a single destination visitor can incur is limited or truncated by the physical size of the island. If the site is of high value to the locals, such that their maximum WTP exceeds the maximum distance necessary to drive to the site, this value will not be reflected in a typically estimated trip frequency model (e.g., count data model of recreation). That is, the choke price may be constrained below the maximum WTP by the physical distance of the island. In this case, TCM will underestimate visitors' maximum WTP because it appears to the model that visitation stops at this physically imposed choke price, and there is no WTP beyond this level.

In our data from Puerto Rican residents visiting streams on El Yunque National Forest, the maximum observed travel cost was approximately $30 (strongly influenced by the 100 mile width of the island). To allow respondents WTP to not be constrained by this physical limit on the choke price, we asked them a dichotomous choice contingent valuation question if they would still take their most recent trip at a random **increase** in the bid amount that was upwards of $200. This additional question allowed us to look at the same valuation problem from a stated preference perspective and proves useful as it shows how much the TCM underestimates people's WTP with spatial truncation. The foundation of our prior that TCM and CVM WTP would normally be equal is based on Carson, et al. (1996) comprehensive comparison of nearly 100 TCM and CVM studies that showed no statistical difference in WTP between the two methods.

In the next sections we elaborate on the idea of truncated spatial markets and how this can affect the WTP measures that researchers obtain when using TCM. Then, we discuss the empirical application in which this truncation is seemingly observed, explain the methodology followed to determine individual's WTP under the models we use and present the results obtained from them. We also simulate a truncated market to determine whether the observed impacts are consistent with our intuition behind this potential problem. Finally, we present some concluding remarks and recommendations.

A Truncated Spatial Market

The TCM assumes that people from different points can travel to a given site. Because a main component of the implicit price in the model has to do with transportation costs, travel cost is understood to increase in a continuous fashion as one gets further away from the site of interest. Figure 3.1.A shows a representation of this spatial property of the travel cost. In the representation one can see that the cost of visiting a site increases as we move to the outer rings of the diagram. On the other hand, figure 3.1.B shows what would happen if the spatial market was truncated and the geographical area around the site was limited. In this case, the maximum amount observed is lower than the one we see in diagram A. Even if the site was worth more to the average person in the inner rings, visitors are prevented from revealing it because no visitors can come from greater distances due to the size of the spatial market. In essence, the demand curve is truncated at the maximum amount of money needed to visit the site from any particular point of the island.

As presented in figure 3.2, the reduction in WTP (hence net consumer surplus) caused by spatial truncation comes from two different sources. First, when calculating consumer surplus from visitors' revealed preferences, the researcher does not observe any portion of the demand curve that is above the geographical choke price P_c. The area above this price is not revealed to the researcher, thus it cannot be accounted for despite being a real benefit to consumers. Furthermore, because more often than not, TCM valuation studies make use of fully parametric regressions (count data models), the demand curve estimated by them adjusts itself to the information it has, tilting the schedule down towards the choke price.

As a result of this the estimated demand curve ($D^{est.}$) appears flatter than the actual demand schedule (D^{actual}). Not only would the researcher miss the portion of the demand that is above the truncated price level, but it would also force the estimated demand curve to adjust to this lack of information beyond P_c and cause a further underestimation of consumer surplus.

Methodology for Comparing TCM and CVM

To measure the degree of underestimation in visitors WTP from the TCM in a constrained island environment, we compare our TCM estimates to those estimated from a dichotomous choice Contingent Valuation Method (CVM) from a dataset obtained on an island. Furthermore, to confirm that this underestimation could be the result of a truncated spatial market, we artificially truncate the market for a second dataset that does not suffer from limited geography. Then we look at the effect that eliminating the "outer rings" of the market has on our WTP values.

Likely, any difference between TCM and CVM estimations is not due to hypothetical bias or other biases associated with CVM. In 1996, Carson et al. used over 600 different CVM and TCM estimates and concluded that differences between CVM and TCM WTP were not statistically significant. If any, CVM WTP measures are generally below TCM WTP estimates (roughly 0.9 of TCM estimates). The argument is that CVM does not suffer from the physical limits as it increases the travel cost by a random amount, so a difference between the two WTP measures could be attributed to the *Island Effect* explained above.

Comparing TCM and CVM WTP Results

Determining the level of underestimation requires the appropriate estimation of both the TCM and dichotomous choice CVM. In the TCM case, we use a traditional count data model. To account for possible overdispersion a negative binomial distribution was chosen and robust standard errors were obtained for each coefficient in the specified model. Two sets of parameters were estimated under the TCM. The first one uses the on-site correction described by Englin and Shonkwiler (1995) and the second one is an uncorrected version. Corrected on-site WTP values are expected to be smaller than the uncorrected WTP values because they are meant to obtain the surplus of the general population rather than just the visiting portion. With this in mind, the study also looks at the corrected and uncorrected CVM equivalent so both conditional and unconditional estimates can be compared. For the dichotomous choice CVM a logistic distribution was chosen. We use weighted estimations of the parameters to account for the increased probability of sampling more avid users with on-site data. The method derived from Loomis (2007) weights each observation by the inverse of the seasonal trips made by each visitor and seeks to reduce the weight users that show greater avidity will have in the process of estimating the model parameters.

In both models (CVM and TCM) the observations considered were limited to those individuals who indicated that visiting the site was the main purpose of their visit, a common practice to control for the possible multiple destination problem mentioned before and found sometimes in on-site samples. We also incorporated the travel time as a separate variable to control for the possible underestimation of the TCM WTP.

Once the coefficients for the models are obtained mean WTP measures are calculated following TCM and CVM theories and considering the distributional assumptions made. An empirical convolution process follows in order to statistically determine whether differences in WTP measures are significant. The method proposed by Poe et al. in 2005 is intended to find all possible differences between two sets of values. By exploiting the distributional assumptions about the model parameters we generate a random vector of WTP values within the coefficients' confidence intervals following a Krinsky-Robb method. The convolutions method then looks at these vectors and determines the probability that one WTP distribution lies on top of the other. The resulting p-values are then used as statistical ground to test that CVM and TCM WTP measures are indeed different.

Simulating a Truncated Spatial Market

To determine whether a truncated market can be responsible for any difference in the WTP obtained with the TCM we simulate spatial truncation in an otherwise conventional market. By removing observations from the "outer rings" we are able to gradually observe the impact that truncation has on three key elements in the analysis of travel cost data: vertical intercept of our implied demand curves, expected number of trips and net WTP. We would expect these variables to react differently when the spatial market they rely upon is truncated and understanding their trends will help us isolate the impact that spatial truncation or *Island Effect* has on the benefit estimates.

The value for each of the three variables of interest is calculated at different levels of spatial truncation. These levels are determined by the statistical properties of the distance traveled variable. The expected number of trips and WTP can be readily found but the vertical intercept value had to be obtained by setting the number of trips to zero in our estimated demand function and solving for the intercept of the equation. The first estimation is done with a full dataset. The following estimations eliminate the observations that are above two standard deviations from the distance traveled mean, one standard deviation, at our original mean, one standard deviation and two standard deviations below. In each round we toss out the observations that are further than the pre-determined levels. Then we re-estimate the model and calculate the implied variables of interest. These values are tabulated and compared for each level of truncation to determine whether there is a progression that implies a difference between a full spatial market and a truncated one.

Empirical Applications

El Yunque National Forest, Puerto Rico

The first part of this study uses a data set from a survey administered in the El Yunque National Forest (YNF) in Puerto Rico. The YNF is the only rainforest in the National Forest System. It is also one of the only protected "old growth" rainforests on the island of Puerto Rico, as well as a cultural and historical landmark for locals. The on-site surveys contain information on trip demand for the 2005 season and a CVM question that was meant to complement the trip assessment. In person interviews were conducted at ten recreation sites along the Mameyes and Espíritu Santo rivers, which include several scenic waterfalls. Data contained demographic information of the users, distance and time traveled, characteristics of the visited sites and a contingent valuation question in the form of; "Taking into consideration that there are other rivers as well as beaches nearby where you could go visit,if the cost of this visit to this river was $____ more than what you have already spent, would you still have come today?" Bid amounts ranged from $1 to $200 per trip.

Over 700 observations were obtained and coded, of which 430 observations were used in this analysis. The reason for the reduction in observations is because only trips where visiting the site were the main reason for traveling were considered valid for the TCM. This is done to deal with multiple destination problems (274 trips were not single destination trips). As mentioned before, these observations are typically pointed out as a source of distortion in travel cost models. Also, because of the complicated form of the corrected negative binomial likelihood function, we eliminated observations from visitors who took more than 100 trips because they appear to be from visitors that are somehow quite different than the vast majority who take a small fraction of these trips.

Variables in the models include an intercept, travel cost per person (in the TCM case) and a bid amount each visitor was asked to pay (in the CVM case). To avoid the subjective use of the wage rate as a proxy for opportunity cost of time, we use the actual gas expense divided by the number of adults in each group as our definition of the travel cost. The travel time variable was included separately and was multiplied by the wage rate to reduce the correlation with our travel cost. Since value of travel time is a separate variable, the proportion of the wage rate used will have no effect on the TC coefficient, as the coefficient on the travel time will implicitly reflect whatever fraction of the wage rate visitors are responding to. The model also includes mean annual stream discharge (as a measure of average seasonal flow), distance of bridge to pool (as a measure of accessibility) and median grain size (measure of substrate sand size). A dummy variable was also included to indicate whether there were picnic areas at the site and restaurants in the area of interest. A dummy variable (gender) was also used to define whether the visitor was male or female. Four other dummy variables allow us to separate the effects of the two sites for which we have a large number of observations (La Mina and Puente Roto) and the sites for which preliminary analysis suggested some unobservable site characteristics that influence number of trips and/or WTP (Juan Diego and El Verde Bridge). These dummies work as separate intercepts for the TCM demand equation estimated and allow the unbiased calculation of our WTP parameter. The following is a table that presents the summary statistics for the observations considered for the variables used.

Snake River, Wyoming

For the second part of this study we use data from the Snake River in Jackson Hole, Wyoming. This stretch of the Snake River south of Grand Teton National Park provides a wide spectrum of recreational activities that include fishing from shore, fishing from boats, scenic raft trips, as well as hiking/fishing/jogging along the levees.

Visitors to one of four areas along the Snake River were given a mail-back survey packet during weekdays and weekends during the month of August through Labor Day weekend in September of 2000. The four sampling locations included a boat put-in and take-out point used by private and commercial rafters, as well as two levee areas used for fishing, hiking, and jogging. A random sample of visitors was intercepted as they returned to their vehicles at each location. Visitor names and addresses were recorded so that a reminder postcard and second mailing of the survey to non-respondents could be performed. Only individuals over 18 years of age were requested to fill out a survey. We only had 19 refusals, for a refusal rate of just 3%. There were 657 surveys handed out and the overall response rate was 65%.

Results

El Yunque National Forest

The first set of results that we present here belongs to the initial comparison between the two non-market valuation techniques in the context of an island market. As explained in the methodology section, due to the on-site nature of our survey, we estimate models uncorrected and corrected for on-site sampling. The uncorrected version of the models provides us with a conditional WTP measure as it can only be used to say something about the preferences of those who visit the sites of interest. In the case of the corrected estimation a more general assessment of the preferences in the population can be obtained.

Four models were used for the comparisons intended in the first part of this study. The results of these models are summarized in Table 3.2. It should be mentioned that the highly significant value for our overdispersion parameter in the TCM results suggests our data suffers from this common problem and so we correctly chose a negative binomial distribution for the estimation. As expected, the WTP measures for the corrected TCM and CVM are lower than the uncorrected ones and both TCM WTP values are well below their CVM analogs.

In all cases, the values obtained in the regression follow what theory suggests with a negative and significant bid and travel cost coefficient. These coefficients yield a $21 WTP per day trip for the corrected TCM, $26 for the uncorrected TCM and $110 and $116 for the equivalent CVM estimations. This means that our CVM results are roughly 423% greater than the estimated WTP in the TCM.

Results from the empirical convolutions show that in both cases (corrected and uncorrected) the CVM WTP is statistically different from the TCM WTP measures. In both cross model comparisons (TCM versus CVM WTP) our two tail p-value was zero. This shows that neither TCM WTP distributions overlap the CVM WTP counterparts so the estimated WTP are in fact statistically different. This is not surprising considering the

WTP obtained for the uncorrected dichotomous choice CVM is more than four times greater than the uncorrected TCM WTP and more than five times greater than the WTP obtained for the corrected estimations. As for a comparison within TCM models or within CVM models (corrected versus uncorrected versions), we find that neither can be considered statistically different, although the TCM results are more different from each other than the CVM results.

Table 3.3 addresses one of the assumptions made in the models presented above. In the four specifications used with the TCM and CVM models data from all sites were pooled together. This imposes the restriction of equality across the parameters that define the demand for each site other than what is accounted for in the intercept shifters. It should be noted that only one of our site intercept shifters was significantly different from zero (Juan Diego) in the TCM and La Mina, El Verde Bridge and Puente Roto in the corrected CVM. To determine whether pooling of data across sites influenced our results, we estimate individual site demands and dichotomous choice models. The site specific regression equations are presented in Appendix A. We look for two indicators in these regressions: (1) whether our differences in WTP between TCM and CVM in the pooled dataset WTP is present across the individual sites, and (2) whether the WTP measures obtained from individual site regressions are consistent with WTP by site from the pooled model[5].

Table 3.3 shows the uncorrected TCM and CVM WTP values for all the sites studied. Neither the TCM nor CVM regressions are corrected for endogenous stratification due to the relatively low number of observations available for each site and the complexity of

[5] This would suggest that our decision to pool the data was appropriate because imposing equality across site parameters is supported by our data.

the TCM likelihood function of a corrected Negative Binomial distribution. This however should not affect our ability to recognize that the difference in conditional WTP between TCM and CVM. Differences between TCM and CVM WTP for the sites range from $54 to $98. This means that CVM estimates were between two and seven times greater than the TCM WTP. The results from our separate regressions for La Mina and Puente Roto show that there is no significant deviation between the WTP obtained for these sites and the pooled version or the results from the rest of the sites. This makes it unlikely that the voluminous data for the sites are somehow driving our results in the pooled model.

The last piece of evidence in the first part of the empirical study is presented in figure 3.3. Figure 3.3 shows that the effect of the island's physical size limit determining the choke price in the "continuous" count data model also biases the slope coefficient. So the reduced WTP with the TCM is a combination of the censored choke price and its effect on the price coefficient. Figure 3.3 also illustrates what the implied demand curve from the CVM looks like[6].

Snake River Simulation

As for the second part of our study, table 3.4 presents a summary of the negative binomial count data TCM model as we eliminate more distant observations from our spatial market in the Snake River data. Each level is obtained by limiting the distance to the site that is considered in the estimation. As we limit the maximum distance observations that are included we are effectively eliminating the "outer rings" of the spatial market.

---

[6] The form of the implied demand curve was obtained by taking the price coefficient as the demand slope and taking the current (expected) number of trips at the current sample price as given. The intercept then follows as a consequence of the linearization assumed for the depiction of the schedules.

Table 3.4 summarizes the regression results at each level of truncation. As expected, all travel cost coefficients are negative and statistically significant. Also, the number of observations is reduced as we reduce the spatial market. It is worth mentioning that our goodness of fit also decreases by roughly 45%.

In table 3.5 we see how the simulated truncation of the spatial market considered in each estimation has several relevant effects on the variables of interest. First, we see a dramatic increase in the number of expected seasonal trips. This is partly due to a decrease in the average price faced by users. This implies that the intersection between the site demand curve and the price happens at a point further out in the quantity axis. As for the change in the vertical intercept, as expected, we see a decrease. This is not as pronounced as we expected, but nevertheless follows the intuition presented before. Lastly, we look at the change in WTP. The progress in this variable has the expected direction but the level at which it decreases is not as pronounced as we observed in the island case study. The smaller-than-expected changes in the vertical intercept and the individual WTP could be due to the fact that we are just eliminating the furthest observations but no real geographical choke price is present, hence, we do not observe clustering of the observations around it. This in turn makes our estimation less prone to use misleading information from the clustered data points.

Conclusions

We tested for the effect of truncation of the spatial market on the WTP derived from the TCM. In the case of Puerto Rico, the count data TCM corrected for on-site sampling bias had a negative and statistically significant travel cost coefficient but yielded an average net WTP of only $21 per trip. The dichotomous choice CVM had a negative statistically significant bid coefficient and yielded an average net WTP of $117 per trip. As can be seen, this is a sizeable difference given that both are modeling the exact same people at the exact same sites. Our interpretation is that the higher WTP estimate from the dichotomous choice CVM is more reflective of the high quality visitor experience associated with the El Yunque National Forest, the only tropical rainforest in the National Forest System, a set of sites with plunging waterfalls and excellent pools for swimming.

Normally, we do not expect a large difference in terms of WTP between TCM and CVM recreation values (Carson et al. 1996). Because of the way the CVM question was phrased and presented we find it very unlikely that our results are picking up anything but use values. This makes the results for our TCM and the dichotomous choice CVM comparable in terms of the values that they look to obtain. With this compatibility and, based on the empirical evidence presented by Carson et al. (1996) that shows that hypothetical bias should not lead to statistically different measures of WTP, it would be reasonable to consider the marked differences in surpluses as the result of a truncation in the spatial market for the sites due to the small size of the island of Puerto Rico.

We also rule out the possibility that using on-site TCM results corrected for endogenous stratification and truncation could be responsible for the observed difference. Comparing both corrected and uncorrected CVM to TCM WTP we do not observe a reduction in the gap between the two measures.

Instead, the results obtained in this study suggest that the reason behind the marked difference between the WTP measures can be attributed to an Island Effect. To show this is possible we performed a TCM analysis at a single recreation site with simulated “islands” of different sizes. The results show that as a geographical market grows in size relative to the quality of the recreation site WTP estimates increase, the number of expected trips decreases and the demand curve becomes steeper and more binding. Alternatively, on islands smaller than Puerto Rico (e.g. Virgin Islands) the underestimation of WTP by TCM could be even larger. Researchers need to be aware of this concern when performing local recreation site valuation on islands where most of the visitor use is by island residents.

References

Blundell, R. and J.M. Robin "Latent Separability: Grouping Goods Without Weak Separability." *Econometrica* 168 (Jan. 2000): 53-84.

Carson, Richard T., N. Flores, K.M. Martin and J.L. Wright "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics* 72 (Feb. 1996): 80-99.

Clawson, M. and J. Knetsch "Economics of Outdoor Recreation" John Hopkins University Press, Baltimore, MD (1996).

Englin, J., Shonkwiler, J.S. "Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand Under Conditions of Endogenous Stratification and Truncation" *Review of Economics and Statistics* 77 (1995): 104-112.

Haspel, A.E. and F.R. Johnson "Multiple Destination Trip Bias in Recreation Benefit Estimation" *Land Economics* 58 (Aug 1982): 364-372.

Larson, D. and S. Shikh "Empirical Specification Considerations for Two-Constraint Models of Recreation Demand." *American Journal of Agricultural Economics* 83 (May 2001): 428-440.

Loomis, J.B. "Correcting for On-Site Visitor Sampling Bias When Estimating the Regional Economic Effects of Tourism." *Tourism Economics* 13 (March 2007): 41-47

Loomis, John B., and Richard Walsh "Recreation Economic Decisions: Comparing Benefits and Costs." State College, Pennsylvania: Venture Publishing, Inc. (1997)

McConnell, K. "On-Site Time in the Demand for Recreation." *American Journal of Agricultural Economics* 74 (1992): 918-924.

Mendelsohn, R.J. Hof, G. Peterson and R. Johnson "Measuring Recreation Values with Multiple-Destination Trips." *American Journal of Agricultural Economics* 74 (1992): 926-933

Parsons, G.R. "The Travel Cost Model" *A primer on Nonmarket Valuation*. Champ, P., Boyle, K.J., Brown, T.C., eds., Vol. 3. Kluwer Academic Publishers, Netherlands, 2003.

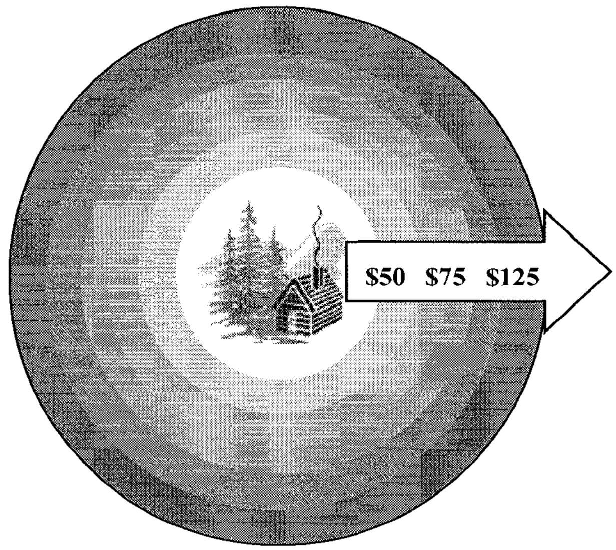
Poe, G.L., Giraud, K.L. and J.B. Loomis "Computational Methods for Measuring the Difference of Empirical Distributions." *American Journal of Agricultural Economics* 87 (2005): 353-65.

Shaw, W.D. "On-Site Samples Regression: Problems of Non-negative Integers, Truncation and Endogenous Stratification." *Journal of Econometrics* 37 (1988): 211-223.

Shaw, W.D. "Opportunity Cost of Time" *Land Economics* 68 (1992): 107-115

Smith, V.K. and R.J. Kopp "The Spatial Limits of the Travel Cost Recreational Demand Model." *Land Economics* 56 (1980): 64-72

Figure 3.1 A) Continuous Spatial Market Assumed by TCM and



B) Example of Truncated Spatial Market

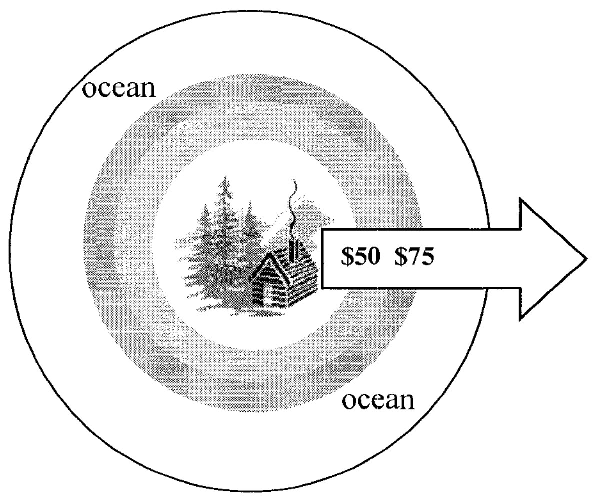


Table 3.1 Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Bid	63.8700	59.4178	1	200
Travel Cost	2.89	2.90	0	30
Travel TimexWage Rate*	3.86511	5.213116	.05167	51.15327
Gender	0.5249	0.4999	0	1
Education	13.4841	3.1879	0	28
Juan Diego	0.0595	0.2368	0	1
La Mina	0.3016	0.4594	0	1
El Verde Bridge	0.0218	0.1463	0	1
Puente Roto	0.3353	0.4726	0	1
Mean Annual Discharge	0.8267	0.5753	0.106	1.667
Dist. Pool to Bridge	23.8889	31.7407	0	145
Median Grain Size	462.9603	566.4335	102	2337
Picnic	0.4782	0.5000	0	1
Restaurants	0.1356	0.3427	0	1

* For consistency, the wage rate was adjusted to a per minute measure because travel time is included in minutes.

Figure 3.2 Truncated Demand Schedule

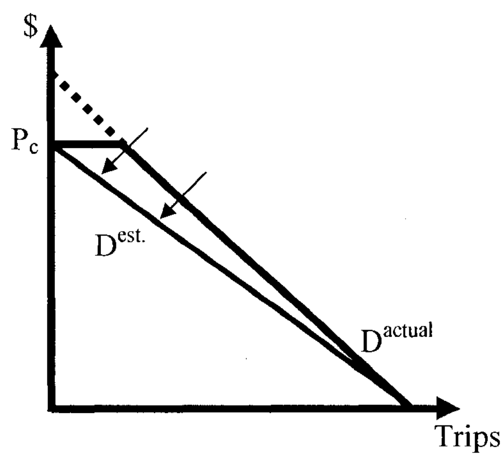


Table 3.2 Results from parametric regressions using CVM and TCM models corrected and uncorrected on-site sampling.

	TCM		CVM	
	Corrected	**Uncorrected**	**Corrected**	**Uncorrected**
TC gas (TCM) or Bid (CVM)	-0.0462***	-0.0300***	-0.0185***	-0.0165***
	(0.0146)	(0.0081)	(0.0027)	(0.0021)
Value of Travel Time	-0.0669**	-0.0376***		
	(0.0280)	(0.0126)		
Gender	0.4443***	0.3590***	0.6795***	0.3023
	(0.1785)	(0.1357)	(0.2775)	(0.2294)
Education	0.0036	-0.0027	-0.0371	-0.0371
	(0.0272)	(0.0204)	(0.0436)	(0.0360)
Site Specific Intercepts				
Juan Diego	-1.6846***	-1.2669***	-1.2135	-1.1070
	(0.6605)	(0.5044)	(0.9035)	(0.8566)
La Mina	-0.5954	-0.3507	-0.9422*	-1.0600**
	(0.4203)	(0.3289)	(0.5281)	(0.4829)
El Verde Bridge	0.6862	0.4452	-1.0028	-1.1759*
	(0.6224)	(0.4448)	(0.7183)	(0.6492)
Puente Roto	0.1496	0.1420	0.6666*	0.5962*
	(0.4434)	(0.2870)	(0.3779)	(0.3548)
Mean Annual Discharge	-0.8537**	-0.5829*	-0.8241*	-0.7347*
	(0.4031)	(0.2986)	(0.4351)	(0.4330)
Dist. Bridge to Pool	0.0088**	0.0074**	0.0055	0.0040
	(0.0044)	(0.0035)	(0.0054)	(0.0052)
Median Grain Size	-0.0002	-0.0001	7.510E-06	-0.0004
	(0.0002)	(0.0001)	(0.0003)	(0.0003)
Picnic Areas	-0.5590***	-0.3012***	0.1152	0.0930
	(0.0954)	(0.0495)	(0.1285)	(0.1086)
Restaurants	0.1641	0.1882	0.2481	0.3079
	(0.2517)	(0.2077)	(0.4259)	(0.3654)
General Intercept	-14.9508	2.3193***	2.8054***	2.9896***
	(0.7472)	(0.5343)	(0.9469)	(0.8541)
Overdispersion	17.3963***	0.9297***		
	(0.4144)	(0.0660)		
N	439	439	452	452
LL	-1013.2405	-1147.6780	-228.9062	-239.3316
WTP	$21.63	$33.29	$117.34	$119.69

* Significant beyond 90% confidence level, ** Significant beyond 95% confidence level, *** Significant beyond 99% confidence level

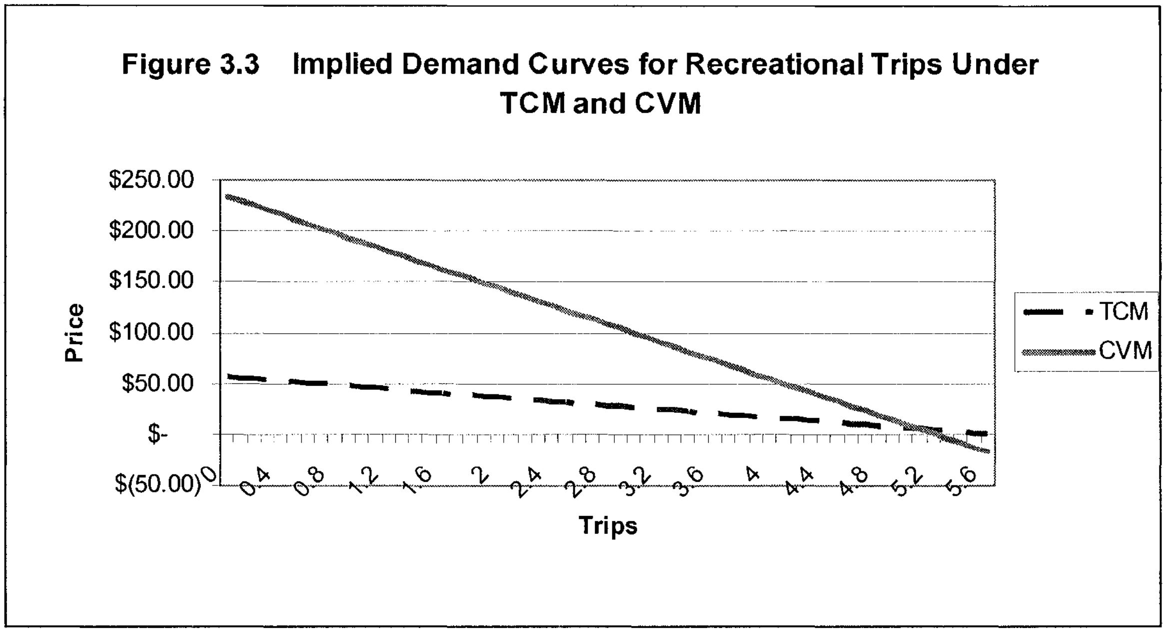


Table 3.3 Individual regressions for sites

	Willingness to Pay per day Trip	
	TCM	CVM
Juan Diego	$ 62.76[a]	$ 160.00
Jimenez Waterfall	-	$ 84.72
La Mina*	$ 18.17	$ 93.84
Sonadora	-	$ 151.89
Waterfall	$ 15.40	$ 111.27
El Verde Bridge	-	$ 39.15
Puente Roto*	$ 68.65	$ 142.44
Jimenez Bridge	$ 33.28[a]	$ 127.38
La Vega	$ 34.98[a]	$ 88.00

* These represent the two sites with most observations (La Mina and Puente Roto).

a. These values were calculated with TC parameters that were not statistically significant.

Table 3.4 Regression results for each level of truncation (std. errors in parentheses)

Max Dist.	Travel Cost	Constant	N	LL	R-squared
4000	-0.0078***	3.7114***	335	-7413.1453	0.2639
	(0.0019)	(0.0995)			
3231	-0.0080***	3.7598***	313	-7015.7000	0.2618
	(0.0020)	(0.0990)			
2462	-0.0085***	3.7721***	305	-6913.2112	0.2558
	(0.0023)	(0.0992)			
1693	-0.0091***	3.7982***	277	-6643.1674	0.2155
	(0.0028)	(0.0996)			
924	-0.0092**	3.8245***	243	-6276.0274	0.1450
	(0.0040)	(0.1010)			

* Significant beyond 90% confidence level, ** Significant beyond 95% confidence level, *** Significant beyond 99% confidence level

Table 3.5 Results from artificially truncating the spatial market

Max. Dist.	λ^a	WTP per Trip	Vertical Intercept
4000	6.98	$127	$ 471.35
3231	7.24	$125	$ 469.98
2462	10.40	$116	$ 443.78
1693	21.81	$110	$ 417.38
924	37.42	$109	$ 415.71

a. λ represents the expected number of trips as determined by the TCM.

Appendix A. Results from individual regressions by site (std. errors in parentheses)

TCM for El Yunque National Forest

Site	Variables: Travel Cost	Total Value of Travel Time	Gender	Constant	Overdispersion	N	LL	WTP per day trip
Juan Diego	-0.0159	-0.0388	0.5160	1.1956***	0.5438**	27	-60.4016	$ 62.76
	(0.0232)	(0.0362)	(0.4030)	(0.3624)	(0.2634)			
Jimenez Waterfall	0.0801	-0.2776**	1.6313	1.5689	1.5860***	15	-45.8682	-
	(0.1549)	(0.1215)	(0.9824)	(1.2942)	(0.3114)			
La Mina	-0.0550***	-0.1332***	0.5482**	2.1881***	1.0200***	146	-422.5468	$ 18.17
	(0.0171)	(0.0376)	(0.2418)	(0.2502)	(0.0944)			
Sonadora	0.0053	-0.0457*	-1.1079	1.8820***	0.4970**	22	-47.5569	-
	(0.0187)	(0.0279)	(0.4243)	(0.3877)	(0.2171)			
Waterfall	-0.0649**	-0.0064	-0.0361	2.1224***	0.6716***	37	-96.6763	$ 15.40
	(0.0281)	(0.0174)	(0.4015)	(0.4471)	(0.1414)			
Angelito Trail	-0.0613	-0.0736*	-0.8834	2.4356***	0.8163***	18	-39.8742	$ 16.32
	(0.0802)	(0.0387)	(0.5553)	(0.8986)	(0.3150)			
El Verde Bridge	0.01808	-0.1265**	-1.26308**	2.61391***	0.44705***	10	-23.4672	-
	(0.0972)	(0.0610)	(0.6174)	(0.5968)	(0.1544)			
Puente Roto	-0.0146*	-0.0483***	0.7560***	1.1725***	0.8056***	164	-387.1899	$ 68.65
	(0.0083)	(0.0154)	(0.2099)	(0.1510)	(0.1200)			
Jimenez Bridge	-0.0300	-0.1504**	0.8723	2.6072*	1.3188***	12	-38.4547	$ 33.28
	(0.3444)	(0.0663)	(0.9093)	(1.4458)	(0.1515)			
La Vega	-0.0286	-0.0967***	-0.3591	1.7650***	0.6635***	27	-58.6193	$ 34.98
	(0.0474)	(0.0392)	(0.5033)	(0.7171)	(0.2557)			

* Significant beyond 90% confidence level, ** Significant beyond 95% confidence level, *** Significant beyond 99% confidence level

CVM for El Yunque National Forest

Site	Variables					
	Travel Cost	Gender	Constant	N	LL	WTP per day trip
Juan Diego	-0.0092	-1.4088	1.9303*	29	-17.7041	$ 160.00
	(0.0062)	(0.9109)	(1.0733)			
Jimenez Waterfall	-0.0512***	-3.1722***	5.5903***	15	-3.7495	$ 84.72
	(0.0185)	(1.2460)	(1.6717)			
La Mina	-0.0192***	0.3405	1.4380***	148	-81.2141	$ 93.84
	(0.0043)	(0.3887)	(0.3127)			
Sonadora	-0.0268***	-2.2002	5.2845**	25	-8.2785	$ 151.89
	(0.0100)	(2.0619)	(2.6054)			
Waterfall	-0.0145**	-0.1650	1.4865*	36	-22.1325	$ 111.27
	(0.0068)	(0.7823)	(0.8480)			
Angelito Trail						
El Verde Bridge	-0.0499	-1.4371	2.5842	11	-4.1894	$ 39.15
	(0.0445)	(1.5824)	(2.2733)			
Puente Roto	-0.0159***	1.1160***	1.5980***	166	-80.6185	$ 142.44
	(0.0034)	(0.4390)	(0.2984)			
Jimenez Bridge	-0.0198	-0.8438	2.7979	12	-5.9685	$ 127.38
	(0.0136)	(2.4742)	(2.4212)			
La Vega	-0.0604***	4.2464**	3.0369***	27	-9.4400	$ 88.00
	(0.0241)	(2.1421)	(1.0480)			

* Significant beyond 90% confidence level, ** Significant beyond 95% confidence level, *** Significant beyond 99% confidence level

CHAPTER FOUR

A Utility Consistent Joint Estimation of Count Data and Dichotomous Choice Models

Introduction

Joint estimation of travel cost models (TCM) and contingent valuation methods (CVM) have become increasingly common in recent years. Many attempts have been made to complement the information provided by these two valuation methods and deal with the shortcomings that each one has as well as expanding the values that they were meant to capture (Cameron 1992; Azevedo, Herriges and Kling 2003; Adamowicz, Louviere and Williams 1994; Englin and Cameron 1996; McConnell, Weninger and Strand 1999; Eom and Larson 2007). This paper looks at expanding the literature in joint estimations of TCM and CVM models by developing a utility framework that relates the count data models used in TCM and the probit models used in dichotomous choice CVM estimations.

Travel cost models are revealed preference methods and have been traditionally favored by economists for the realistic circumstances under which information about preferences is obtained. Contingent valuation methods on the other hand, are stated preference methods and they are particularly useful to obtain values for levels of non-market goods that are currently not observed or are new to the population of interest. Furthermore, CVM is the only method available to economists to obtain values that are not associated with the direct use of market or non-market goods. These non-use values include existence, bequest and option values and may represent the majority of the total value for certain unique natural environments. Despite being useful tools to uncover people's preferences, neither of these methods is free of criticism and their limitations have been widely explored in the literature (Larson and Shaikh 2001, Blundell and Robin 2000, Haspel and Johnson 1982, Boyle 2003). For the TCM, determining the implicit travel cost, dealing with multiple destination trips and length of destination stay bias have been long studied challenges (Smith and Kopp 1980, Loomis 2007). For the CVM, the hypothetical nature of the questions used to elicit people's WTP has been the target of criticism as results can be susceptible to the researchers' representation of the market for the good in question (Boyle 2003).

We combine the two valuation methods using Eom and Larson (2006) and updating Cameron (1992). The estimation follows Cameron who estimates a joint set of parameters by assuming a quadratic utility function that is dependent on two goods, a numeraire good and trips to a site. As Cameron, we use data gathered from a dichotomous choice contingent behavior question that is answered affirmatively if the difference between the utility obtained from the proposed scenario is larger than the alternative one. The change between the scenarios in this application is strictly linked to an increase in the cost per visit at the current number of trips or avoiding the increase entirely by reducing the trips taken. However, we assume a semi-log specification typical of count data models as the Marshallian demand for visitors and derive a consistent utility differential using Eom and Larson (2006).

Since the Marshallian trip demand function obtained through the TCM is the result of an optimization problem that relates back to the individual's utility function, we can convert the utility differential typically estimated through the CVM into an indirect utility differential by substituting the trip demand into the direct utility functions. The challenge however, is that the researcher has to make sure that the utility function and the trip demand function are theoretically connected. In Cameron's case, she chooses a somewhat arbitrary utility function and derives the appropriate specification for trip demand. However, this process results in an uncommon, although consistent, trip demand function rarely used by economists. This paper approaches the same issue in a slightly different, yet meaningful, manner. Since most TCM studies nowadays make use of count data models for the estimation of the trip demand function, we start from a traditional semi-log trip demand function and instead, work back to a consistent utility function that we can use to determine the utility differentials for the CVM estimation process. This ensures that we can estimate a demand function that is consistent with the underlying preferences stated through a dichotomous choice question **and** with the count data nature of trip demand.

Generating a joint model framework has important consequences. As suggested by Azevedo, Herriges and Kling (1999), a joint model allows us to take advantage of each data type's strength without imposing preconceived notions regarding the superiority of one of data over the other. In addition, combining a travel cost and a dichotomous choice model in a single framework disciplined by a common utility framework also ensures that we have a single welfare measure for the site of interest. This is of particular use when we find significant differences in WTP when estimating the models separately and a single value is needed for policy purposes.

Another important gain from combining the two valuation methods has to do with efficiency. The additional information, otherwise incorporated through a panel setup of the data, is expected to considerably reduce the variability of parameter estimates. This reduction stems from the idea that now we are incorporating information on the intensive margin of the demand for trips and the extensive or margin of participation with an increase in the price. In other words, we use the horizontal distance from the origin to the observed number of trips given a fixed price (TCM) and the extent to which users will remain users as we increase the price they face (CVM).

To test gains in efficiency we estimate a separate regression for each model and a joint estimation with the consistent parameter restrictions. Then we compare the confidence intervals between the price coefficients and the associated WTP of each.

In the next sections we present the derivation of this consistent estimation process. We also present the implied parameter restrictions. After that, we present an empirical application of the model and the results obtained from it. Finally, we present some conclusions about the effectiveness of the joint estimation.

Deriving the Utility Framework

We start by explaining the theoretical underpinnings of the optimization process that relates the two data sources we use here. Economic theory establishes that individuals have a set of preferences that can be represented with a function. This utility function depends on the consumption of different goods and services given their relative full prices and available full income[7]. In the simple scenario we explore, this utility function U is dependent upon the consumption of two goods, a numeraire labeled z and trips to the site of interest q. Individuals maximize this function by choosing the level of z

[7] We use the terms full prices and full income to imply that individuals consider not only their monetary prices and incomes, but also associated time prices and time budget.

and q that produces the highest level of utility without exceeding their available time and money budget[8] (Larson and Shaikh 2001).

$$\max_{z,q} U = f(z,q) \qquad \text{s.t.} \quad M_Z^F \times z + M_q^F \times q = Y^F \quad (1)$$

For convenience we normalize the full price of z. This results in a simpler full budget constraint that only has the full price of the trips and full income as exogenous (from now on referred to as M and Y respectively).

From this optimization process one can derive the optimal level of (Marshallian) trips demanded given the full prices and income faced. In the case of the count data models this solution has a semi-logarithmic form:

$$q(M,G,Y) = \exp(\alpha + \beta M + \gamma G + \delta Y) \quad (2)$$

where M is the full price or marginal cost per visit, G represents a set of site characteristics, Y is full income and q is the number of trips to the sites of interest.

An alternative way to solve the same optimization problem is to fix the utility level and determine the consumption that would reach that fixed level while minimizing the associated expenses. In this setup we obtain an expenditure function instead of a Marshallian demand function. Economic theory tells us that the solutions to these problems are related through Shephard's Lemma. Such a relationship also exists between the expenditure function and the indirect utility function.

All these connections between the Marshallian demand, indirect utility function and expenditure function provide the necessary links to derive a form of each that is

[8] To ensure that money and time constraints are binding two numeraires are required, each with only one type of price. A single numeraire problem as the one presented here would then have to be thought of as a vector of several goods, one of which only requires money payments and a second one that requires only time (Larson and Shaikh 2001). This does not alter the steps taken to derive a Marshallian trip demand or the analysis that follows.

consistent with the others. Particularly, they allow us to get from the count data Marshallian demand function expressed above to the appropriate expenditure function and, from there, to the consistent indirect and regular utility functions.

Now, looking at our particular application, it is important to mention that the information assumed to be obtained from the CVM question refers to whether a visitor would have taken the last trip to a site if the marginal price of that visit was increased. This means that: 1) the surveyed individuals must have at least one trip to the site of interest, 2) we do not alter any of the site characteristics for the dichotomous choice question and 3) that only the marginal price of the last visit is affected by the hypothetical scenario. This defines the way the utility differential is believed to motivate CVM answers is set. Although the next paragraphs make use of this particular information set in the derivation of the consistent utility difference and trip demand function, we show that this is only a special case that can be extended to all hypothetical changes in price with an associated loss of access or trips taken.

In the context of our particular application, a respondent would only say yes if:

$$\Delta U = \max_{q} U\big(Y - M(q-1) - (M+B), q\big) - \max_{q} U\big(Y - M(q-1), (q-1)\big) > 0 \quad (3)$$

where B is the bid increase to marginal cost (tc per visit), and ΔU is the change in the level of utility experienced by respondents. The first term is the utility obtained from the consumption of z $(Y\text{-}M(q\text{-}1)\text{-}(M\text{+}B))$ and its purchased with the income left after consuming q number of trips. That is, the visitor can consume her income less the current price per trip (M) times (q-1) trips and the current price plus the bid increase (M+B) times the last trip taken. The second utility term is a function of the income left after consuming only (q-1) trips at the current travel cost level.

Simplifying this expression results in:

$$\Delta U = \max_q U(Y - Mq + M - M - B, q) - \max_q U(Y - M(q-1), (q-1)) > 0$$
$$= \max_q U(Y - Mq - B, q) - \max_q U(Y - M(q-1), (q-1)) > 0 \qquad (4)$$

Or

$$\Delta V = V(Y - B, M, G) - V(Y, M, G) > 0 \qquad (5)$$

If we restrict $q \geq 1$, this is consistent with the CVM question suggested above, so the second term becomes $U(Y - M(q-1), (q-1)) = U(Y - 0, 0)$ in the case where q assumes the lowest possible value. Recall that the dichotomous choice question presented to visitors referred to an *ex post* increase of the price for the current visit.

Since we do not observe people's utility function we need to use the information obtained from their visiting decisions and their choice between the two scenarios presented above. By relating the changes in the two scenarios of the SP portion to different income levels and number of visits to the site of interest, one can incorporate the optimization results that are estimated with the TCM. Based on this, we start from the assumed form of the count data Marshallian demand (semi-logarithmic) and determine the utility function that would be consistent with the relationship specified. Then, we use the consistent preference structure derived as the appropriate expression of people's utility. With this utility function one can calculate the correct expression for the difference between the two scenarios implied by the CVM question. When the optimal level of q is substituted into this expression this difference really becomes a change in the indirect utility levels.

Economists have used count data models to estimate seasonal trip demand functions for recreational purposes because they are discrete distributions that are supported over a strictly positive domain. This of course matches the form of trip. These models, typically Poisson or Negative Binomial, assume a semi-logarithmic form $\left(q(M,G,Y)=\exp(\alpha+\beta M+\gamma G+\delta Y)\right)$ and make use of travel costs to obtain the needed variation in price. From Eom and Larson (2006) the demand function defined by the count data model typically used in TCM ($q(M,G,Y)=\exp(\alpha+\beta M+\gamma G+\delta Y)$) has a quasi-expenditure function of the form:

$$\widetilde{E}(M,G,c(U))=\frac{-1}{\delta}\ln\left(\frac{-\delta}{\beta}e^{(\alpha+\beta M+\gamma G)}-\delta c(U)\right) \qquad (6)$$

where $c(U)$ is a constant of integration. Contrary to Eom and Larson, we assume weak complementarity of the site attributes (no non-use value) and set this constant of integration as a utility index U[9]. Hence, the associated quasi-indirect utility function can be shown to be:

$$\widetilde{V}(M,G,Y)=-\frac{1}{\delta}e^{-\delta Y}-\frac{1}{\beta}e^{(\alpha+\beta M+\gamma G)} \qquad (7)$$

This quasi-indirect utility function can be used to determine the corresponding direct utility function[10] with the following form:

$$U(M,G,Y)=e^{-\delta Y}\left(\frac{-(\beta+\delta q)}{\delta\beta}\right) \qquad (8)$$

[9] Assuming non-use values would complicate the form of the expenditure function requiring that c(U) was also a function of site characteristics. This would imply that non-users could also receive some benefit from the levels of the characteristics considered even when they do not visit the sites. Since the CVM question does not consider any quality changes or non-use values it does not make sense to relax the weak complementarity assumption.

[10] The reason why we present the direct utility function is to point out where is the number of trips entering this function. Later on we substitute q for the utility maximizing number of trips q^*, recovering a version of the quasi-indirect utility function that is now dependent upon the number of trips predicted by the TCM.

Where q is the number of trips consumed and is a choice variable that will be set in a way that will maximize individual utility. The optimal level of trips chosen will be defined as q^* and is estimated through the TCM as $q(M,G,Y) = \exp(\alpha + \beta M + \gamma G + \delta Y)$. Now we have a consistent definition of the utility function and the demand functions used in count data models.

The next step is to go back to our original representation of change in utility and use this to redefine this differential in terms of the consistent utility with the optimal level of trips (hence indirect utility function) derived here.

$$\widetilde{V}(Y,M,G) = e^{-\delta Y}\left(\frac{-(\beta + \delta(q^* - 1))}{\delta\beta}\right) \qquad (9)$$

$$\widetilde{V}(Y - B,M,G) = e^{-\delta(Y-B)}\left(\frac{-(\beta + \delta q^*)}{\delta\beta}\right) \qquad (10)$$

$$\Delta\widetilde{V} = \left[e^{-\delta(Y-B)}\left(\frac{-(\beta + \delta q^*)}{\delta\beta}\right)\right] - \left[e^{-\delta Y}\left(\frac{-(\beta + \delta(q^* - 1))}{\delta\beta}\right)\right] \qquad (11)$$

A more general representation of this problem would be:

$$\widetilde{V}_0(Y_0,M,G) = e^{-\delta Y_0}\left(\frac{-(\beta + \delta(q_0^*))}{\delta\beta}\right) \qquad (12)$$

$$\widetilde{V}_1(Y_1,M,G) = e^{-\delta Y_1}\left(\frac{-(\beta + \delta q_1^*)}{\delta\beta}\right) \qquad (13)$$

Where $Y_0 = Y$ or the entirety of the remaining income, $Y_1 = Y - B$ is the remaining income less the increase in price B, q_0^* equals the optimal number of trips when no increment in cost of fee is charged and $q_1^* = q_0^* - d$ or the original optimal solution for seasonal visits minus the number of trips that visitors will not be able to take (d) if they do not pay the required increase in cost of fee.

Going to back to the application at hand, we can represent the utility differential as:

$$\Delta\widetilde{V} = e^{-\delta Y}\left(e^{\delta B}\xi - \eta\right) \tag{14}$$

where,

$$\xi = \left(\frac{-\left(\beta + \delta q^{*}\right)}{\delta\beta}\right)$$
$$\eta = \left(\frac{-\left(\beta + \delta\left(q^{*} - 1\right)\right)}{\delta\beta}\right) \tag{15}$$

It is easy to see that our application is simply a special case of a broader problem where we constrain visitors to lose only their last trip if they do not pay a general increase in cost B. This makes our results easily applicable to other dichotomous choice scenarios including the one presented by Cameron (1992) in her paper, but now updated to the count data TCM.

The expressions derived will serve to determine the parameter constraints required to obtain consistent estimates when using a count data TCM and a dichotomous choice model that follow the scenarios presented above. The next section explores the statistical considerations that need to be taken into account in order to jointly estimate the parameters in the utility differential and trip demand function.

Joint Estimation

To jointly estimate TCM and CVM parameters we use the joint probability function derived by González-Sepúlveda, Loomis and González-Cabán (2008). This estimator combines a negative binomial distribution corrected for on-site samples and a probit distribution. The joint distribution was derived by taking advantage of the fact that joint densities can be split into the product of a conditional and a marginal density function. In their application, González-Sepúlveda et al. made no effort to establish a common utility framework between the two set of parameters obtained from the TCM and CVM models. Instead, the estimation is meant to test whether there was a correlated underlying error structure without assuming any particular unobserved relation between the two model expressions.

However, the literature on joint revealed-stated preference estimations has stressed on the need for a utility consistent combination of the two valuation methods (Eom and Larson 2006). In this sense, this paper extends the initial effort to update joint estimation methods and use more current count data models, but now also with the underlying utility forms that are implied by the typical semi-logarithmic form assumed in the TCM.

The use of a conditional and a marginal probability function in the place of a joint probability appropriately requires one of the two equations in the model to be conditioned on the other. The requirement is satisfied by the setup we present here where the assumed utility difference is conditional on the number of trips determined by our trip demand function. However, since González-Sepúlveda et al. use a negative binomial distribution corrected for on-site sampling, some changes need to be done to be able to use this joint distribution with our uncorrected estimation. From the original form of the joint density function presented by González-Sepúlveda et al. we can derive the uncorrected equivalent for the joint estimation.

$$L_i = \left[(\pi_i)^{y_{cvm,i}}(1-\pi)^{1-y_{cvm}}\right] \times \left(\frac{\Gamma(\theta+q_i)}{\Gamma(\theta)\Gamma(q_i+1)}\right)\left(\frac{\theta}{(\lambda_i+\theta)}\right)^{\theta}\left(\frac{\lambda_i}{(\lambda_i+\theta)}\right)^{q_i} \quad (19)$$

where $$\pi_i = \Phi\left(\left(\frac{\Delta U}{\sigma} + \rho Z_i\right)\Big/\left(1-\rho^2\right)^{0.5}\right),$$

$$\lambda_i = \exp(\alpha + \beta M + \gamma G + \delta Y),$$

$$\theta = \frac{\theta_0}{\lambda_i},$$

$$Z_i = \left(y_{tcm,i} - E\left(y_{tcm,i}\right)\right)\Big/\left(Var\left(y_{tcm,i}\right)\right)^{0.5}$$

and $$E\left(y_{tcm,i}\right) = \lambda_i + \theta_0$$

The parameter ρ is the correlation parameter between the two models, σ is the standard error of our probit estimation, and $\Phi(\cdot)$ refers to the cumulative standard normal density function of the values within the parentheses.

Another issue is that the original estimator for on-site samples was not concerned with was the problem of scale between models. The scale problem occurs when we want to directly compare the structural parameters for two different samples or two different models. When that is the case the preference parameters are confounded by different scales. With parameter restrictions across models this becomes a serious problem. Particularly in the case of the probit estimator where population parameters are only obtained up to scale, we need to pay special attention to σ which will scale the parameters in the utility difference function (Adamcowikz et al. 1993).

This problem however, is overcome by the non-linear form of the derived utility differential. In the traditional specification it is not possible to identify all the parameters for the independent variables and the scale parameter implicit in the conversion from a normal to a standard normal distribution. With the specification used here identification is no longer a problem and so we can confidently impose cross equation constraints without worrying about the different scales we may have in the two types of data. We set this parameter equal to a constant value due to the complicated for of our likelihood function and the failure to converge with alternative specifications.

The same is true with the level of identification of the TCM specification. Using the Negative Binomial distribution, we also estimate an overdispersion parameter that lies in the definition of the variance of the dependent variable. This extra parameter serves as a scalar that modifies the expected value of the error term in the TCM as needed.

Using this distribution we are able to estimate a joint set of parameters for the CVM and TCM data. In conjunction with the parameter constraints presented in the section before, we estimate our joint models and present the results in the next section.

Welfare Measures

The way the model has been setup allows us to easily obtain two different commonly used measures of people's willingness to pay, consumer and compensating surplus. Both tell us the benefit that consumers get from visiting the sites of interest at different price levels and given the current site characteristics. The difference between the two has to do with the way they consider what a change in relative prices (of the sites and the consumption bundle z) does to quantity consumed. With compensating surplus we find the benefit obtained from changes in relative prices while leaving out the income effects that such changes may bring. Consumer surplus on the other hand, looks at the same benefits but it does not discriminate between effects that come from the new ratio between the prices and the associated perceived wealth level.

To obtain the consumer surplus (or willingness to pay) for a person that take a single trip we simply use the inverse of the price coefficient β (Englin and Shonkwiler 1995). Alternatively, we can get at the median compensating surplus by looking back at the utility differential we presented before and solving for the level B that would leave the visitor indifferent between the two scenarios. By doing so, we obtain the following expression:

$$\frac{\ln(-\beta-\delta(q-1))-\ln(-\beta-\delta q)}{\delta}=B \qquad (20)$$

One can use this expression and compare its value to the more traditional consumer surplus to see if there is a significant difference between the two. If so, that would suggest there is a significant income effect (not considered in the compensated measure) and that consumer surplus might not be an ideal way of getting at people's willingness to pay.

Hypothesis Testing

Bringing together information on the intensive margin for trips and the extensive margin of participation given an increase in the marginal prices faced by the respondent, the model presented here incorporates more information and greatly reduces the uncertainty around our price coefficients. This suggests that the price parameter estimated for the joint estimation is expected be more efficient than the one obtained from an individual regression.

From the prior section we know that both surplus measures are dependent on the parameters estimated with the models. It is easy to see then that reducing the variability of these parameters could in turn reduce the variability of our surplus measures.

To test whether this is true we look at the confidence intervals of the parameters under each model and use them to determine the corresponding confidence intervals of the consumer and compensated surplus. In the case of consumer surplus the relation between these two variables is easy to obtain since one is simply the negative inverse of the other. For the compensated surplus however we have a highly non-linear relationship between two parameters in the derivation made. To obtain the confidence interval for the compensated surplus then, we randomly select 4,000 values for the two variables in the definition of compensated surplus (β and δ). Then, we simply look at the 4,000 definitions that result from the random draws in a Krinsky Robb fashion. Finally, we compare the width of the resulting confidence intervals to look for gains in efficiency between the joint and individual estimations.

Data

In person interviews were conducted at El Yunque National Forest in Puerto Rico during the summer of 2005. The surveys administered during the interviews were collected as part of a comprehensive study on the impact of site characteristics on social and physical conditions in and around the forest streams.

Ten recreation sites along the Mameyes and Espíritu Santo rivers were chosen for this study. Data include visitor's demographics, site characteristics (fixed and variable), trip information and a contingent valuation question in the form of; "Taking into consideration that there are other rivers as well as beaches nearby where you could go visit, if the cost of this visit to this river was $____ more than what you have already spent, would you still have come today?"

Over 700 observations were obtained and coded, but only 250 observations were eliminated because respondents reported that their visit to the site was not the main purpose of their current trip. Also, because of the complicated form of the corrected negative binomial distribution, we eliminated visitors who took more than 20 trips. This is not uncommon, as pointed out by Englin and Shonkwiler (1995), where they limited their corrected Negative Binomial to visitors with fewer than 12 trips.

We decompose travel cost as the sum of two expenses; money and time. This specification separates the two types of expenses and uses actual money (gas) cost and travel time separately. The specification helps us avoid the arbitrary use of a particular fraction of the wage rate as the opportunity cost of time. The gas cost was divided by the number of adults in the group for this specification.

One important consequence of separating the two expenses is the need to include two separate budgets as well. Since the optimization process now responds to a two constraint (money and time) optimization process, our resulting trip demand should be a function of both budget levels. This however requires more information on the visitor's available time than we have in our dataset. Without differentiated levels of time budgets we can only assume a single one for our entire sample. With this assumption, the information contained in this variable gets accounted for by the demand intercept, leaving the price parameter unchanged.

While the price variable in the TCM is the travel cost in the form defined above, the bid amount visitors were asked to pay in the CVM question were simply a random sample of price increases that ranged from $1 to $200 per trip. The site characteristics included were mean annual discharge (as means of flow), distance from the river pools to the bridge access (as a measure of accessibility), median grain size and pool volume (as a proxy for the size of the pool).

Results

Tables 4.1 and 4.2 show the results obtained in the different models used in our study and the impact they have on the variable of interest, WTP for the site. In table 4.1 we present the results obtained for all the models used. The table shows the coefficient and standard errors for the specifications chosen in the individual and joint estimations. As expected, the price parameters for all specifications and estimation methods are negative and statistically significant with an implied price elasticity of -0.19 for the individual estimation and -0.08 for the joint model.

In general, the joint estimation increases the WTP through a reduction in the price parameters obtained. This is encouraging because we now have a single WTP measure in the site users as our uncorrected joint model provides a surplus that is almost identical to the one obtained with the CVM. But the real gain from estimating our parameters jointly seems to come from a dramatic reduction in the standard errors of the parameters. This should not be surprising because with our joint estimation we consistently add two sources of information.

The reduction in the parameters' standard error results in tighter confidence intervals of the surpluses measured through the model. This is shown in table 4.2 where we present the minimum and maximum values of the implied WTP at a 90% confidence interval for each model and both surplus types. We also include the mean values and the span (or difference) between the two limits implied by our standard errors.

It is easy to see that the joint estimates are by far narrower than the individual counterparts. At the 90% level, the confidence interval from the traditional single estimation model was a staggering $184 for the consumer surplus and $133 for the compensated one. For the jointly estimated parameters these confidence intervals where far narrower ranging from $26 to $32. This notable difference is of course due to the reduced variation in the joint estimation and may have particular impacts when using the results for benefit transfer purposes. For policy analysis a tighter confidence interval aid us in valuing alternative management actions that may change recreation use or valuing recreation sites for maintenance or keeping open in the face of reduced agency budget.

Table 4.3 presents the comparison between the consumer and compensated surplus for each model. For the compensated measure we use both the expected number of trips and the sample the average. In the case of the consumer surplus we simply take the inverse of the price parameter for each model. The results show that there is virtually no difference between the two types of surpluses. In no case the difference between the two exceeds 46 cents. This is not surprising because our income variable is also statistically insignificant which suggest there is no income effect. Without any income effect the gap between the two can be expected to be zero.

Conclusions

This paper updates the estimation methods used when looking at combining travel cost and dichotomous choice contingent valuation data. In the sprit of Cameron (1992), we look at the appropriate parameter restrictions that would make both models consistent with a common underlying utility function. However contrary to choosing an arbitrary utility function and solve for a consistent demand equation we start with the most common estimation model for the TCM and work our way to a consistent utility differential expression. Although our particular CVM question dealt with changes in the marginal cost of the last trip taken, we present a utility differential setup that could be generalized to all types of price changes and a corresponding limit to the number of trips available to users.

The results obtained support the idea that imposing a consistent utility structure results in greater gains in terms of both consistency of surplus measures between models and the efficiency of the parameters estimated. The gain in efficiency, although expected, is encouraging because it considerably enhances the ability to compare benefits and costs with a high confidence level. In our particular application, a $184 window between the maximum and the minimum possible values would not say much about how much people are really benefiting from the sites they visit. In fact if we consider the estimated 600,000 yearly visitors that El Yunque National forest receives every year this interval would produce benefit estimates that would range between $14.6 millions and $125.6 millions for users in a particular season. If instead we consider the $27 interval obtained with our joint estimation aggregate seasonal benefits to users would lie somewhere between $57.3 millions and $73.8 millions.

When using these results for benefit transfer purposes, the size of our confidence interval will have a big impact over the values we choose to transfer over to a policy site. It is in this type of applications (benefit transfer studies) that future research should take place to determine the worth of reducing variability in such a significant way through the use of models like this.

Finally, another area in which this model can be used is to get at compensated values through the derivation of Hicksian demands and the use of the consistent expenditure functions presented here. This type of application may be useful to compare surpluses in situations where it is believed that consumer surplus differs greatly from the compensated measures and income plays a big role in the changes proposed.

References

Adamowicz W., J. Louviere and M. Williams "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities." *Journal of Environmental Economics and Management* 26 (May 1994): 271-292.

Azevedo, C. D., J. A. Herriges and C. L Kling "Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations." *American Journal of Agricultural Economics* 85 (August 2003): 525-537.

Ben-Akiva M. and T. Morikawa "Estimation of Switching Models from Revealed Preferences and Stated Intentions" *Transportation Research* 24 (1990): 485-495.

Blundell, R. and J.M. Robin "Latent Separability: Grouping Goods Without Weak Separability." *Econometrica* 168 (Jan 2000): 53-84.

Boyle, K.J. "Contingent Valuation in Practice" *A primer on Nonmarket Valuation*, Champ, P., K.J. Boyle and T.C. Brown, eds. Netherlands: Kluwer Academic Publishers, 2003.

Cameron, T. "Combining Contingent Valuation and Travel Cost Data for Valuation of Nonmarket Goods." *Land Economics* 68 (Aug. 1992): 302-317.

Englin, J. and T.A. Cameron "Augmenting Travel Cost Models with Contingent Behavior Data: Poisson Regression Analyses with Individual Panel Data." *Environmental and Resource Economics* 7 (Sept. 1996): 133-147.

Englin, J. and J.S. Shonkwiler "Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand Under Conditions of Endogenous Stratification and Truncation." *Review of Economics and Statistics* 77 (Feb. 1995): 104-112.

Eom, Young-Sook and D. Larson "Improving environmental valuation estimates through consistent use of revealed and stated preference information." *Journal of Environmental and Economic Management* 52 (July 2006): 501- 516.

González-Sepúlveda, Juan M., J. Loomis and A. González-Cabán "A Joint Estimation Method to Combine Dichotomous Choice CVM Models with Count Data TCM Models Corrected for Truncation and Endogenous Stratification" *Journal of Agricultural and Applied Economics* (August 2008) In press.

Haspel, A.E. and F.R. Johnson "Multiple Destination Trip Bias in Recreation Benefit Estimation." *Land Economics* 58 (Aug. 1982): 364-372.

Larson, D. and S. Shikh "Empirical Specification Considerations for Two-Constraint Models of Recreation Demand" *American Journal of Agricultural Economics* 83 (May 2001): 428-440.

Loomis, John "Correcting for On-Site Visitor Sampling Bias When Estimating the Regional Economic Effects of Tourism." *Tourism Economics* 13 (March 2007): 41-47

Smith, V.K. and R.J. Kopp "The Spatial Limits of the Travel Cost Recreational Demand Model." *Land Economics* 56 (1980), 64-72

McConnell, K., Q. Weninger and I.E. Strand "Joint Estimation of Contingent Valuation and Truncated Recreational Demands." *Valuing recreation and the environment: Revealed preference methods in theory and practice*. Herriges, J. A. and C. L. Kling, eds. Cheltenham, Glos, England: Edlward Elgar Publishing Limited, 1999.

Table 4.1 Single and joint estimation parameters and standard errors for models uncorrected and corrected for on-site sampling.

	CVM[a]	TCM	
		Individual	Joint
Constant	1.1648***	1.423507***	1.3813***
	(0.203)	(0.142)	(0.132)
Bid	-0.0103***		
	(0.001)		
Gas Cost		-0.0229**	-0.0093***
		(0.011)	(0.001)
Travel Time		-0.0033***	-0.0035***
		(0.001)	(0.001)
Income	-1.0500E-06	-1.1000E-05	-8.4568E-06
	(2.580E-06)	(1.70E-05)	(1.47E-05)
Mean Annual Discharge	0.0165	-0.1011	-0.0937
	(0.180)	(0.084)	(0.078)
Dist. Bridge to Pool	-0.0013	0.0008	0.0009
	(0.002)	(0.002)	(0.001)
Median Grain Size	-1.3000E-05	9.8000E-05	9.6640E-05
	(1.264E-04)	(6.60E-05)	(6.43E-05)
Pool Volume	0.0001		
	(1.969E-04)		
Overdispersion		1.2992***	1.3071***
		(0.223)	(0.225)
Rho			-0.0164
			(0.043)
σ			98.8720***
			(12.431)
Log Likelihood	-243.74991	-923.7718	-1155.0234
WTP	$ 110.50	$ 43.71	$ 107.51

* Significant beyond 90% confidence level, ** Significant beyond 95% confidence level, *** Significant beyond 99% confidence level

a. CVM results where obtained estimating a typical linear utility differential, not the one derived here consistent with count data models.

Table 4.2 90% confidence intervals for each model's mean surpluses (consumer and compensated).

Model	Lower 90%	Mean	Upper 90%	Difference
Consumer Surplus				
Individual	\$24.40	\$43.71	\$209.28	\$184.88
Joint	\$95.49	\$107.51	\$122.99	\$27.49
Compensated Surplus				
Individual	\$23.46	\$43.60	\$156.76	\$133.30
Joint	\$93.26	\$107.51	\$124.98	\$31.72

Table 4.3 Comparison of model's consumer and compensating surplus measures.

	TCM	
	Individual	**Joint**
E(q)	5.2762	5.2896
$\bar{q}$	5.4407	5.4407
Median Compensating surplus (with *E(q)*)	\$43.61	\$107.05
Median Compensating Surplus (with $\bar{q}$)	\$43.60	\$107.03
Mean Consumer Surplus	\$43.71	\$107.51

E(q) is the expected number of trips as calculated by each estimated trip demand function.

$\bar{q}$ refers to the sample average number of trips.

CHAPTER FIVE

Concluding Remarks

Sir, I have found you an argument;
but I am not obliged to find you an understanding.

SAMUEL JOHNSON
BRITISH AUTHOR, LEXICOGRAPHER

In the process of explaining the difference between willingness to pay values obtained with TCM and CVM, the work here uncovers some interesting limitations that each valuation method has. In our particular application, jointly estimating these models without imposing consistency provides little gain in efficiency and highlights, as traditional economic intuition would suggest, the importance of a common utility structure to close the gap between the two.

However, we focus on more than fixing this consistency issue and how it relates to willingness to pay values. This work seeks to look deeper into the nature of TCM and CVM and show that they can be complementary. In doing so, we identify some relevant issues that go beyond the traditional problems of uncertainty about the definition of the travel cost variable or the hypothetical nature of the CVM questions and responses. Even without imposing a common theoretical framework, joint use of these models can help us correct for endogenous stratification in the CVM and avoid biases by recognizing spatial limitations in the TCM. Finally, when we decide to impose a common utility framework, we observe a substantial gain in the efficiency of our price parameters and willingness to pay measures.

The general lesson here is to learn as much as you can from the data you have before jumping into imposing quick fixes that may deliver a consistent economic valuation but may overlook important issues. This dissertation is an example that such path can lead to important discoveries about your models that are useful in applications beyond yours.

General Appendix A

Visitor Survey Script

Hello, my name is ______________________. I work for University of Puerto Rico, and I am doing this survey to learn more about what visitors to this river do for recreation. This will just take a few minutes of your time, but what you tell me will be very helpful for improving the management of this recreation site. Your answers are completely **confidential** and **anonymous**.
(Note to Interviewer: Please keep a tally of number of visitors contacted and the number that refused to be interviewed).

0. Record whether respondent is ___Male ___ Female
(Please try and alternate between male and female respondents)

1. During your stay here at the river, what kind of activities have you or will you participate in?
   (Check all that they mention or closest category)

___ Picnicking/eating/drinking	___Visiting with family & friends
___ Sun bathing	___ Relaxing
___ Enjoying nature	___ Spiritual renewal/Therapy
___ fishing/shrimping	___ kayak/canoe/belly boards, rafts
___ Listening to music	___ Swimming/Wading in River/cooling off in River
___ Other Please list	

2. Including yourself, how many people are in your group?
   ___# Adults ____# Teenagers ___# children ___ # Total

3. About how long do you expect to be here at this spot on the river today?
   ____ Minutes (convert hours to minutes: half hour = 30 minutes, 1 hour = 60 minutes, 1 ½ hours = 90 minutes, 2 hours – 120, 3 hours = 180 minutes, etc.)

4. How enjoyable would you say your visit has been?

1	2	3	4	5	6	7	8	9
Not very enjoyable		Somewhat enjoyable			Moderately enjoyable		Very enjoyable	

5. How crowded did you think the river segment was where you were visiting? Please circle a number representing how crowded it was.

1	2	3	4	5	6	7	8	9
Not at all crowded		Slightly crowded			Moderately crowded		Extremely crowded	

6. Was this visit to the river today (check one):
   ___ The sole or principal reason for your trip from home?
   ___ One of several equally important places you visited today on your trip from home?
   ___ Just a spur of the moment or incidental stop on this trip?

   WHAT ANIMALS LIVE IN THE RIVER WATER
   NOTHING FISHES SHRIMP EELS SNAILS

7. About how long did it take you to travel from home to this spot on the river?
   ____ Minutes (convert hours to minutes: half hour = 30 minutes, 1 hour = 60 minutes, 1 ½ hours = 90 minutes, 2 hours – 120, 3 hours = 180 minutes, etc.)

8. About how far is it (one-way) from where you live to this spot on the river?
   ____ One-way Miles

9. About how much was the gasoline cost for this trip to you? $_____
   IF CANNOT SAY $ SPENT OR SAID ZERO, ASK HOW MANY GALLONS OF GAS THEY THINK THEY USED.

10. As you know the price of gasoline often goes up. Taking into consideration that there are other rivers as well as beaches nearby where you could go visit, if the cost of this visit to this river was $____ more than what you have already spent, would you still have come today?

    ____ Yes ____ No

11. About how many trips from home did you make in the last 12 months to this particular spot or area of the river?
    _______# of Annual of trips to this river

12. If this river area was not available, where else would you go today?
    Name of other site/area ________________________________

13. Do you visit other rivers in Puerto Rico?
    13a. Yes-→ Name of the other river ____________________ → Annual # of Trips___
    Name of the other river ____________________ → Annual # of Trips___
    Name of the other river ____________________ → Annual # of Trips___
    ____ No

    If Yes→ 13b. About how many total trips did you make to **all** rivers in Puerto Rico during that same 12 months?
    _______ # of Annual number of trips to all other rivers in PR

14. We are also interested in people that go shrimping.

    14a. Do you go shrimping? ____ Yes→ We would like to pay you to attend a short discussion group with other shrimpers to learn more about where you go shrimping.

    14b. Would you be interested in participating?

    ____ No ____ Yes→ Please write down your name and phone number on this card, which will be kept separate from the survey. Thank you for your help on this matter.

    14c. Do you know others that go shrimping that you think would be willing to attend this same discussion group? ____ No ____ Yes→ Please write down their name and phone number below yours on this card. Thank you for your help on this matter.

DEMOGRAPHICS

Now we want to ask some questions that will help us to check to see if our sample of visitors reflects the population of Puerto Rico. Your answers are **confidential** and **anonymous** and you will not be linked to your answers.

15. How old are you? ______

16. What is your zip code? ______

17. What is your education level? (number of years of formal schooling) ____# years Where Elementary school = 6, junior high = 9, High School =12, junior college = 14, college graduate = 16, graduate school = 18.

18. We would like to have an idea of the family annual income (This includes the income of everyone living in your household). Please tell us which of the following income categories your family income falls into (Hand then card):

a. Less than $10,000	b. $10,000 - $19,999	c. $20,000 - $29,999
d. $30,000 - $39,999	e. $40,000 - $49,999	f. $50,000 - $59,999
g. $60,000 - $74,999	h. more than $75,000	

Thank you for your participation. Your cooperation will help to improve the future management of these rivers and recreation areas.

Can I answer any questions for you?

General Appendix B

Overview of Data and Site Descriptions

The data used in this dissertation was obtained from an on-site survey performed in El Yunque National Forest in Puerto Rico. The analysis presented here is part of a broader project that seeks to evaluate interactions between hydrological characteristics, human activities and biological processes that take place on these sites. The following is information about El Yunque National Forest and a summary of the data obtained with the surveys collected.

El Yunque National Forest is located in the northeastern part of the Caribbean island of Puerto Rico. It is the only tropical rainforest in the U.S. National Forest System. Although it is relatively small (28,000 acres), it has great importance for the recreational opportunities it provides to locals and visitors alike. The forest is also “home” to thousands of native plants including 150 fern species, 240 tree species (23 of them only found in this forest). Even though there are no large wildlife species in the forest, thousands of smaller creatures can be found.

The forest encloses the Luquillo Mountains that rise to 3,533 ft. above sea level. It can receive over 200 inches (508 centimeters) of rain and has an average temperature of 73° F (21° C) all year long. With many trails and recreation areas, El Yunque offers unique opportunities for eco-tourism activities and passive enjoyment of waterfalls and rivers. Roughly 600,000 visitors enjoy these activities every year (USDA Forest Service 2003).

Something unique about El Yunque National Forest is its proximity to a large metropolitan area. The greater San Juan area (capital of Puerto Rico) is only 45 minutes away and represents a market of more than two million people that can easily access its trails, rivers and impressive waterfalls. This makes the forest a popular destination for recreators. Particularly, because there is no other protected rainforest with the same accessibility, both in the sense of proximity and developed infrastructure for their enjoyment.

The data gathered for this project were obtained from 10 different sites during random dates and times in the months of May and September of 2005. This is considered by most locals to be the span of the season to visit the sites studied. Why these period is the most active in terms of visits to the forest is directly related to the beginning of summer vacation for most families (May) and the intensification of the hurricane season (September). Collection dates included weekdays and weekends as well as holidays. The 10 sites chosen are points that are either accessed through or right at intersections between rivers and roads in two of the main watersheds in the forest, Mameyes and Espíritu Santo.

The río Mameyes (Mameyes River) basin starts from atop of El Yunque Mountains and extends 7.5 miles to the Atlantic Ocean. The Mameyes watershed is of great importance due to its scenic values and research and biological values (USDA Forest Service 2003). It covers roughly 10 square miles, which is 16.5% percent of the forest surface. Some of the most visited sites for water recreation in the island are found in this watershed due to numerous trails that give access to remote waterfalls and pools (González-Cabán and Loomis 1999).

On the other hand, the río Espíritu Santo (Espíritu Santo River) watershed extends over the western part of the forest. The tributaries in it run opposite to the Mameyes watershed for 11.9 miles to the Atlantic Ocean. The watershed covers roughly 35 square miles or 57% of the forest area. The Espíritu Santo River is not as well known and it does not have an extensive infrastructure in place as the Mameyes.

The following are maps of the study area and the sites included in the data collection process.

Figure 1. Location of Study Area, El Yunque National Forest.

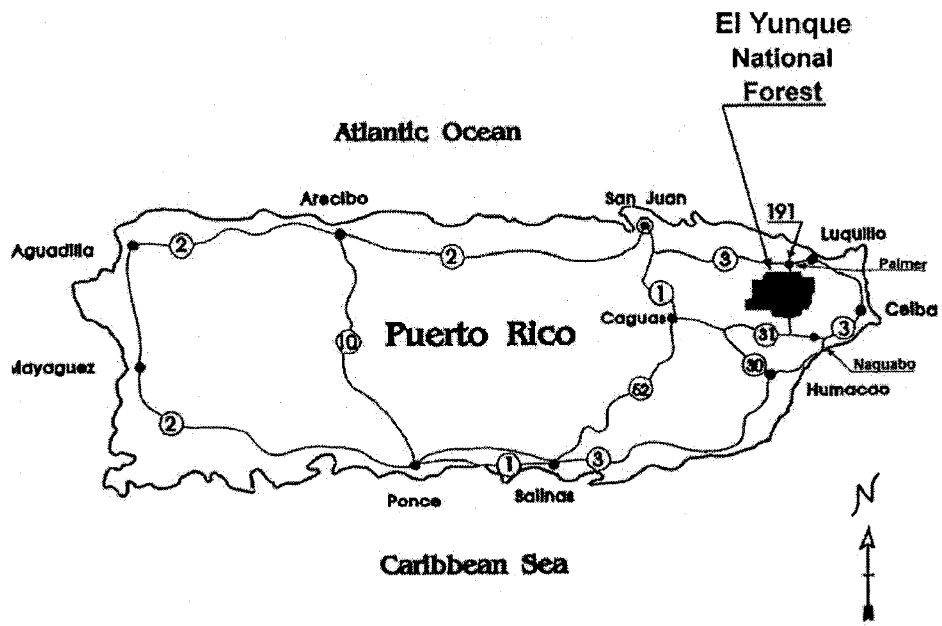
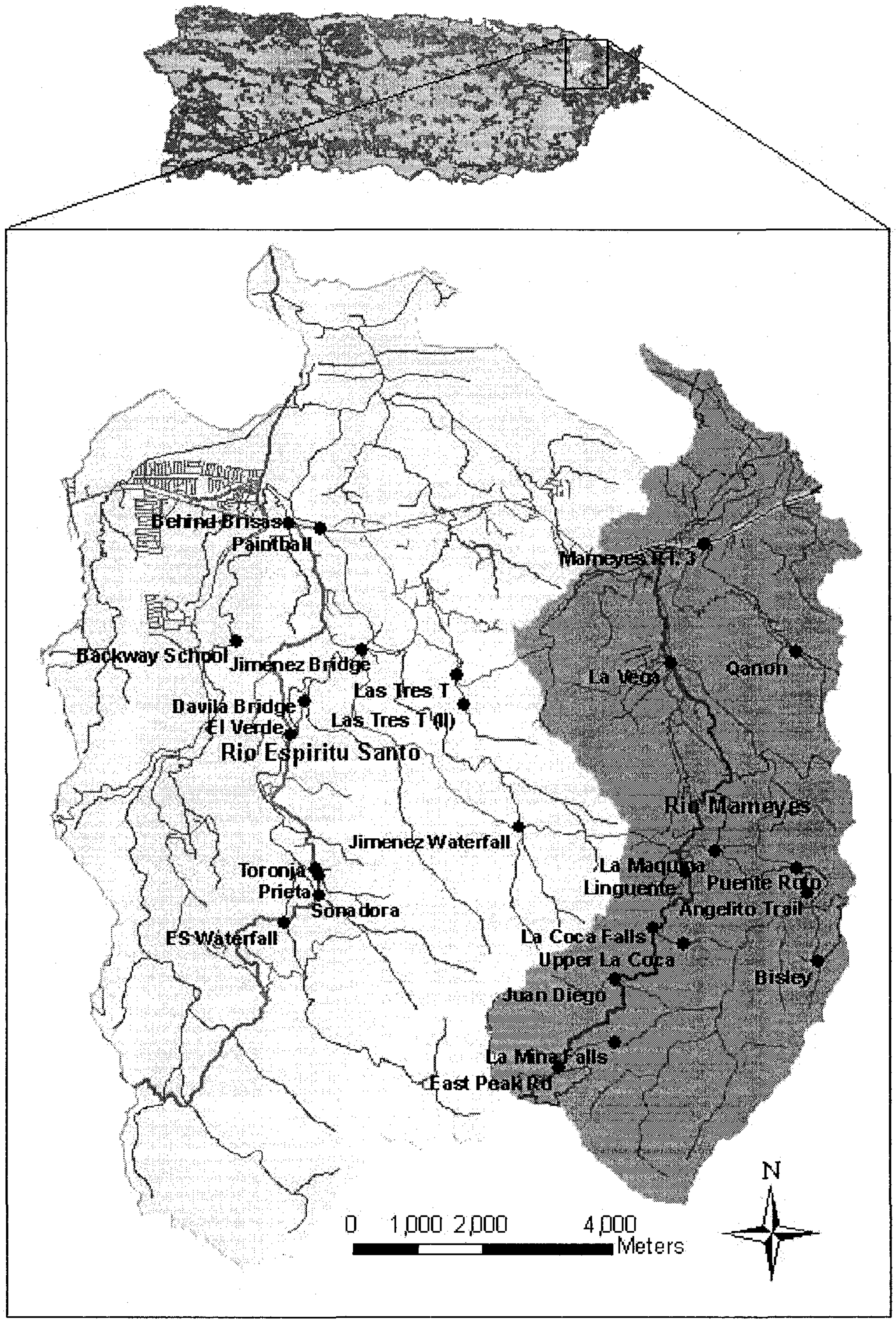
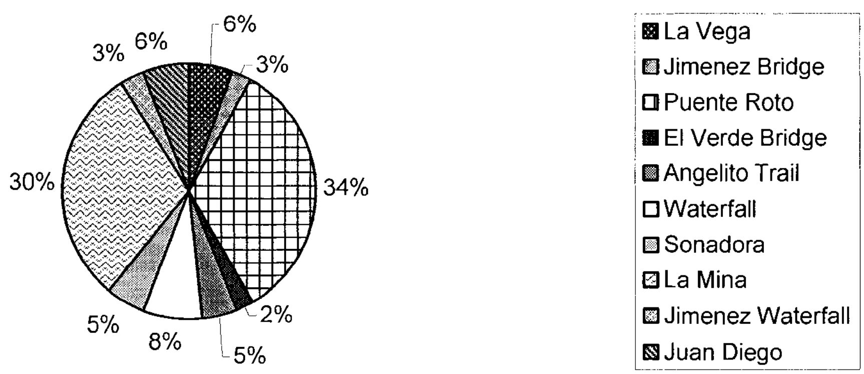


Figure 2. River Watersheds and Recreation Points considered in the Study, Mameyes (right) and Espíritu Santo (left).



Although figure 1 presents a large number of nodes where data were collected, only 10 of those presented were used for recreation surveys. Over 700 observations were collected in La Vega, Jimenez Bridge, Puente Roto, El Verde Bridge, Angelito Trail, Waterfall, Sonadora, La Mina, Jimenez Waterfall and Juan Diego. Out of these observations, only 430 observations were used to avoid multiple destination bias. That is, only respondents that reported their visit was the main purpose of their trip were included in the dataset used. No information was used on visiting expenses that went beyond the gas cost associated with driving from a Puerto Rico address. Also, observations that had more than 100 seasonal visits were excluded because they seem to be generated by a different process. Figure 3 summarizes the distribution per site of the observations used.

Figure 3. Distribution of Observations by Site



Tables 1 to 10 present the summary statistics per site of the variables used in the model. These variables include: annual trips, gas cost, travel time, bid price (CVM hypothetical increase in marginal price of visit), family annual income, distance from bridge to pool (as a proxy for accessibility), median grain size (measure of substrate sand size), pool volume (proxy for pool size), number of adults in the group education level and gender.

Table 1. Summary Statistics: La Vega

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	28	3.035714	5.540973	1	30
Gas Cost	27	6.037037	3.877762	0	15
Bid	27	65.37037	53.11319	5	200
Travel Time	28	61.42857	39.67127	10	150
Fam. Annual Income	27	31481.48	22427.87	5000	75000
Dist. Bridge to Pool	28	0	0	0	0
Median Grain Size	28	102	0	102	102
Pool Volume	28	428.3	0	428.3	428.3
Adults	28	2.928571	1.274495	2	8
Education	28	12.67857	3.356073	4	18
Gender	28	0.535714	0.507875	0	1

Table 2. Summary Statistics: Jimenez Bridge

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	13	9.846154	16.05639	1	48
Gas Cost	13	4.461538	2.258886	2	10
Bid	13	77.30769	73.64372	5	200
Travel Time	13	50.76923	35.46396	15	120
Fam. Annual Income	13	35384.62	27402.22	5000	75000
Dist. Bridge to Pool	13	145	0	145	145
Median Grain Size	13	180	0	180	180
Pool Volume	13	877	0	877	877
Adults	13	4.076923	2.691392	1	10
Education	13	14.61538	2.399252	9	18
Gender	12	0.416667	0.514929	0	1

Table 3. Summary Statistics: Puente Roto

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	169	3.828402	7.560086	1	52
Gas Cost	165	9.586364	9.468443	0	60
Bid	169	56.56805	58.14501	1	200
Travel Time	168	72.44048	88.06127	5	990
Fam. Annual Income	169	26997.04	26165.65	5000	250000
Dist. Bridge to Pool	169	0	0	0	0
Median Grain Size	169	159	0	159	159
Pool Volume	169	764.3	0	764.3	764.3
Adults	168	3.666667	3.282446	1	30
Education	169	13.40828	3.014675	4	18
Gender	169	0.502959	0.501477	0	1

Table 4. Summary Statistics: El Verde Bridge

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	11	4.818182	8.459529	1	30
Gas Cost	11	5.909091	3.505839	1	10
Bid	11	55.90909	51.61483	10	160
Travel Time	11	60.90909	35.83421	10	120
Fam. Annual Income	10	40500	27507.57	5000	75000
Dist. Bridge to Pool	11	0	0	0	0
Median Grain Size	11	241	0	241	241
Pool Volume	11	342.9	0	342.9	342.9
Adults	11	3.272727	2.148996	1	9
Education	11	14.90909	4.826066	9	28
Gender	11	0.545455	0.522233	0	1

Table 5. Summary Statistics: Angelito Trail

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	23	3.434783	8.049992	1	40
Gas Cost	18	10.44444	7.617489	5	38
Bid	23	81.30435	57.54874	5	180
Travel Time	23	78.69565	28.45321	20	120
Fam. Annual Income	23	30000	23584.95	5000	75000
Dist. Bridge to Pool	23	0	0	0	0
Median Grain Size	23	114	0	114	114
Pool Volume	23	1868.4	0	1868.4	1868.4
Adults	23	5.391304	3.499859	1	12
Education	23	13.43478	4.272811	2	23
Gender	23	0.478261	0.510754	0	1

Table 6. Summary Statistics: Waterfall

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	38	4.973684	6.491103	1	30
Gas Cost	37	8.486486	7.286295	2	40
Bid	37	94.32432	59.3853	5	200
Travel Time	38	64.23684	48.05652	1	180
Fam. Annual Income	38	32026.32	24559.9	5000	75000
Dist. Bridge to Pool	38	30	0	30	30
Median Grain Size	38	2337	0	2337	2337
Pool Volume	38	455	0	455	455
Adults	38	3.710526	1.929961	1	9
Education	38	13.94737	3.35267	6	19
Gender	38	0.631579	0.488852	0	1

Table 7. Summary Statistics: Sonadora

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	25	3.28	4.782956	1	24
Gas Cost	22	13.13636	12.20966	2	60
Bid	25	68.2	58.11053	5	180
Travel Time	25	68	35.56098	20	180
Fam. Annual Income	25	43000	24769.77	5000	75000
Dist. Bridge to Pool	25	5	0	5	5
Median Grain Size	25	711	0	711	711
Pool Volume	25	90.5	0	90.5	90.5
Adults	24	4	2.813013	1	12
Education	25	16.16	2.134635	10	20
Gender	25	0.56	0.506623	0	1

Table 8. Summary Statistics: La Mina

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	152	7.098684	11.96726	1	60
Gas Cost	146	6.630137	5.926069	0.25	38
Bid	150	56.13333	56.61641	5	200
Travel Time	152	50.14474	41.53279	5	360
Fam. Annual Income	151	23658.94	18878.5	5000	75000
Dist. Bridge to Pool	152	35	0	35	35
Median Grain Size	152	508	0	508	508
Pool Volume	152	71	0	71	71
Adults	149	3.09396	2.20056	1	15
Education	152	12.99342	2.957753	6	18
Gender	152	0.546053	0.499521	0	1

Table 9. Summary Statistics: Jimenez Waterfall

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	15	10.46667	22.61121	1	80
Gas Cost	15	7.2	3.509172	3	15
Bid	15	55.66667	64.19464	5	200
Travel Time	15	57.33333	30.87224	30	120
Fam. Annual Income	15	28000	24168.31	5000	75000
Dist. Bridge to Pool	15	52	0	52	52
Median Grain Size	15	457	0	457	457
Pool Volume	15	42	0	42	42
Adults	14	2.928571	1.77436	1	8
Education	15	13.6	4.484895	0	18
Gender	15	0.4	0.507093	0	1

Table 10. Summary Statistics: Juan Diego

Variable	Obs	Mean	Std. Dev.	Min	Max
Annual Trips	30	3.266667	5.25182	1	30
Gas Cost	27	10.55556	7.667781	3	30
Bid	30	89	66.86837	5	200
Travel Time	30	86.16667	68.22482	25	360
Fam. Annual Income	30	21583.33	19348.03	5000	75000
Dist. Bridge to Pool	30	93	0	93	93
Median Grain Size	30	178	0	178	178
Pool Volume	30	60.1	0	60.1	60.1
Adults	28	5.107143	3.224206	2	16
Education	30	13.3	2.394678	9	16
Gender	30	0.5	0.508548	0	1

References

González-Cabán, A. and J. Loomis. "Measuring the Economic Benefit of Maintaining the Ecological Intergrity of the Río Mameyes in Puerto Rico" *USDA Forest Service Research Paper PSW-RP-240* (1999)

USDA Forest Service Web Site, El Yunque National Forest Information Page. http://www.fs.fed.us/r8/caribbean/index.shtml