

9-12-2014

# The Value of Wildlife Watching Recreation and Techniques for Addressing Nonresponse in Nonmarket Valuation Surveys

Jee Hwang

Follow this and additional works at: [https://digitalrepository.unm.edu/econ\\_etds](https://digitalrepository.unm.edu/econ_etds)

---

## Recommended Citation

Hwang, Jee. "The Value of Wildlife Watching Recreation and Techniques for Addressing Nonresponse in Nonmarket Valuation Surveys." (2014). [https://digitalrepository.unm.edu/econ\\_etds/15](https://digitalrepository.unm.edu/econ_etds/15)

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Economics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

**Jee Woong Hwang**

*Candidate*

**Economics**

*Department*

This dissertation is approved, and it is acceptable in quality and form for publication:

*Approved by the Dissertation Committee:*

**Alok K. Bohara, Chairperson**

**Jennifer A. Thacher**

**Brady P. Horn**

**Wendy L. Hansen**

**THE VALUE OF WILDLIFE WATCHING RECREATION  
AND TECHNIQUES FOR ADDRESSING  
NONRESPONSE IN NONMARKET VALUATION  
SURVEYS**

**by**

**JEE WOONG HWANG**

B.S., Economics, East Carolina University, 2003  
M.S., Economics, East Carolina University, 2005

DISSERTATION

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

**Doctor of Philosophy  
Economics**

The University of New Mexico  
Albuquerque, New Mexico

**July, 2014**

## **DEDICATION**

*To my Father, Mother, Han, and Scarlett for all their love and support*

## **ACKNOWLEDGEMENTS**

I am truly grateful for the wonderful guidance I received from my advisor and dissertation committee chairman Professor Alok K. Bohara. This work would not have been possible without his patience and dedication. I would also like to thank the other committee members: Professors Jennifer A. Thacher, Brady P. Horn, and Wendy L. Hansen, for their interest in my research and for their suggestions and insights that have been invaluable in improving the quality of this work.

**THE VALUE OF WILDLIFE WATCHING RECREATION AND  
TECHNIQUES FOR ADDRESSING NONRESPONSE IN NONMARKET  
VALUATION SURVEYS**

**by**

**Jee Woong Hwang**

**B.S., Economics, East Carolina University, 2003**

**M.S., Economics, East Carolina University, 2005**

**Ph.D., Economics, University of New Mexico, 2014**

**ABSTRACT**

In this dissertation, the value of wildlife watching recreation is used as the context for exploring different modeling approaches available for combining data from nonmarket valuation surveys such as the travel cost method (TCM) and contingent valuation method (CVM). Another topic explored in this dissertation is the issue of nonresponse in nonmarket valuation surveys. The results of this dissertation are useful for wildlife and land managers interested in obtaining theoretically consistent values of wildlife watching recreation using a combination of TCM and CVM data. The practical solutions to nonresponse in nonmarket valuation surveys are useful for researchers who wish to implement

relatively small-scale surveys due to limited budgets and are concerned about maintaining the sample size.

In the first segment of this dissertation (chapter 2), differences in the preference structure of wildlife watchers is studied by estimating a finite mixture (FM) model of wildlife watching recreation using data from a national survey. This model is unique in that the multiple imputation (MI) method was applied to the FM model framework to address missing travel cost data. When compared to a truncated FM model, it was found that using the MI approach led to different consumer surplus estimates, but also greater performance in terms of goodness of fit.

Combining TCM and CVM data has many advantages for recreation demand modeling in that it helps overcome some of the shortfalls where only one type of data is used. An area less explored is the use of a utility-consistent modeling framework that unifies the TCM and CVM components to arrive at a single estimate of welfare. However, the scenarios reflected in CVM questions may not always have a TCM counterpart, therefore leading to a discrepancy in the corresponding welfare measures. Chapter 3 presents a utility-consistent joint model of wildlife watching recreation where the scenarios and the welfare measures from the TCM and CVM components are the same. An advantage of this model is that welfare effects can be interpreted in terms of net benefits or willingness to pay.

Nonresponse in CVM surveys can lead to a loss in statistical efficiency and bias. In the final segment of this dissertation (chapter 4), a simple recoding procedure is introduced to address missing follow-up responses from a double-

bounded dichotomous choice CVM survey. A Monte Carlo simulation was used to examine the performance of recoding compared to list-wise deletion across three scenarios that included randomly arising nonresponse and systematically arising nonresponse. The results of the simulation showed recoding led to lower losses in statistical efficiency.



## TABLE OF CONTENTS

<b>LIST OF FIGURES</b> .....	xii
<b>LIST OF TABLES</b> .....	xiii
<b>LIST OF ACRONYMS</b> .....	xv
<b>CHAPTER 1 Introduction</b> .....	1
<b>CHAPTER 2: Finite Mixture Model of Wildlife Watching Recreation with Multiply-imputed Travel Costs</b> .....	7
2.1. Introduction.....	7
2.2. Literature Review.....	10
2.2.1. Methods for Valuing Wildlife Watching Recreation.....	10
2.2.2. Preference Heterogeneity in Recreation.....	12
2.2.3. Imputation of Nonparticipant Travel Costs.....	13
2.3. Methodology.....	15
2.3.1. Finite Mixture Model.....	15
2.3.2. Estimation of Benefits from Access to Recreation.....	17
2.3.3. Predictive Mean Matching Multiple Imputation.....	19
2.4. Application to Wildlife Watching Recreation Data.....	23
2.4.1. Data.....	23
2.4.2. PMM Results.....	25
2.4.3. Criteria for the Number of Mixture Components.....	27
2.4.4. MI FM Model Results.....	28
2.4.5. Expected Consumer Surplus Estimates.....	31

2.4.6. Zero-truncated FM Model Results.....	32
2.4.7. Comparison of Goodness of Fit.....	34
2.5. Discussion and Conclusion.....	35
<b>CHAPTER 3: Utility-consistent Joint Estimation of the Value of Wildlife Watching Recreation.....</b>	<b>55</b>
3.1. Introduction.....	55
3.2. Literature Review.....	57
3.2.1. Combining Revealed and Stated Preference Data.....	57
3.2.2. Utility-consistent Joint Estimation.....	59
3.3. Methodology.....	62
3.3.1. Derivation of Utility-difference Model using TCM Parameters.....	62
3.3.2. Estimation of Utility-consistent Model Parameters.....	66
3.2.3. Econometric Equivalence between Compensating Variation and Consumer Surplus.....	68
3.4. Data.....	70
3.5. Results.....	73
3.6. Discussion and Conclusion.....	75
<b>CHAPTER 4: Recoding Double-bounded Dichotomous Choice Contingent Valuation Method Data with Nonresponse.....</b>	<b>82</b>
4.1. Introduction.....	82
4.2. Literature Review.....	84
4.2.1. Nonresponse in Contingent Valuation Surveys.....	84
4.2.1.1. Unit versus Item Nonresponse.....	84

4.2.1.2. Sources of Nonresponse.....	85
4.2.1.3. Consequences of Nonresponse.....	86
4.2.2. Conventional Methods for Addressing Nonresponse.....	87
4.2.2.1. Weight Adjustments.....	87
4.2.2.2. Imputation and Nonparametric Methods.....	87
4.2.2.3. Heckman’s (1979) Two-Step Procedure.....	88
4.3. Methodology.....	89
4.3.1. Econometric Model for Double-bounded Dichotomous Choice Data.....	89
4.3.2. Monte Carlo Simulations.....	92
4.4. Results.....	95
4.4.1. Random Nonresponse.....	95
4.4.2. Systematic Nonresponse in Upper $X$ .....	95
4.4.3. Systematic Nonresponse in Upper $Z$ .....	96
4.5. Discussion and Conclusion.....	97
<b>CHAPTER 5: Conclusion.....</b>	<b>104</b>
<b>APPENDICES.....</b>	<b>108</b>
Appendix A: Contingency Table of Average Posterior Probabilities.....	109
Appendix B: Table of Average Information Criterion and Entropy Values for Truncated FM Models.....	110
Appendix C: Zero-truncated Finite Mixture Count Model.....	111
Appendix D: Derivation of Expected Consumer Surplus with Finite Upper Limit.....	113

Appendix E: Variance Function of Expected Consumer Surplus with Finite Upper Limit.....	114
Appendix F: Goodness of Fit Measures for Count Data Models.....	116
Appendix G: Joint estimation of TCM and CVM Model Parameters.....	117
Appendix H: Stata Codes.....	119
<b>REFERENCES</b> .....	165

## **LIST OF FIGURES**

2.1	Distributions of Expected Trip Counts by Components.....	52
2.2	Consumer Surplus with Finite Upper Limit.....	53
2.3	Cumulative Densities of Observed and Predicted MI and Truncated Trips.....	54
3.1	Effects of Complete Substitution from Trips to AOG.....	81

## LIST OF TABLES

2.1	Summary Statistics of Variables and Descriptions.....	37
2.2	PMM Results for Travel Costs.....	41
2.3	Average Information Criterion and Entropy Values.....	42
2.4	MI Parameter Estimates of Single and 3-Component FM Models.....	43
2.5	Component-level Mean Values.....	45
2.6	Expected Consumer Surplus Estimates from MI Models.....	46
2.7	Parameter Estimates of Single and 3-Component Zero-truncated FM Model.....	47
2.8	Expected Consumer Surplus Estimates from Zero-truncated Models.....	49
2.9	Summary Statistics of Observed Trips Counts and Predicted Values.....	50
2.10	R <sup>2</sup> Measures for MI and Zero-truncated FM Models.....	51
3.1	Summary Statistics of Variables and Descriptions .....	77
3.2	Individual and Joint Models, and Utility-consistent Model Results.....	79
4.1	Double-bounded Data Regression Simulation with Random Nonresponse.....	98
4.2	Double-bounded Data Regression Simulation with Systematic Nonresponse in Upper X.....	100

4.3	Double-bounded Data Regression Simulation with Systematic Nonresponse in Upper $Z$ .....	102
-----	---	-----

## **LIST OF ACRONYMS**

AIC	Akaike Information Criterion
AOG	All Other Goods
BIC	Bayesian Information Criterion
CS	Consumer Surplus
CV	Compensating Variation
CVM	Contingent Valuation Method
DB	Double-bounded
FHWAR	Fishing, Hunting, and Wildlife-Associated Recreation
FM	Finite Mixture
FMI	Fraction of Missing Information
HPM	Hedonic Price Model
INR	Item Nonresponse
MAR	Missing at Random
MI	Multiple Imputation
NB1	Negative Binomial Model with Constant Dispersion
NB2	Negative Binomial Model with Mean Dispersion
PMM	Predicted Mean Matching
RMSE	Root Mean Squared Error
RE	Relative Efficiency
RP	Revealed Preference
RVI	Relative Variance Increase



SB	Single-bounded
SEI	Standard Error Increase
SP	Stated Preference
TCM	Travel Cost Method
UNR	Unit Nonresponse
WTP	Willingness to Pay
ZTNB	Zero-truncated Negative Binomial

## **Chapter 1: Introduction**

Wildlife watching recreation has become one of the most popular forms of outdoor recreation in the United States. Nearly a quarter of the population enjoys observing, feeding, or photographing wildlife every year and there are substantial benefits to local economies (USFWS, 2012). However, the growing interest by domestic and international visitors has prompted concerns over the sustainability of the wildlife watching industry. Studies have shown excessive visitation may disturb animal populations and adversely affect the long-term health and viability of some species (Tappe, 2006). For decades, data from national surveys such as the National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (FHWAR) has provided researchers with a plethora of data to model the behavior of wildlife watchers. However, there has been little effort to better understand the disparate preference structures of wildlife watchers which can provide important insights on visitation patterns and help ensure sustainable wildlife populations without stifling local economies. For example, a better understanding of the disparate behavioral patterns of wildlife watching groups may help recreational sites managers curb crowding at these sites by introducing different price schedules and through adjustments in visitation hours. One of the key contributions of this dissertation is the investigation of the presence of preference heterogeneity in the U.S. wildlife watching population while employing the latest recreation demand modeling techniques to estimate the value wildlife watching activities.

The inclusion of contingent valuation method (CVM) questions in FHWAR surveys has provided researchers with the opportunity to take advantage of a methodological advancement in recreation demand modeling—the combining of CVM data with travel cost method (TCM) data. While TCM data has been the traditional choice for recreational demand modeling, combining it with CVM data into a joint model has shown to have many advantages. In particular, TCM data provides information on actual behavior; for example, the number of recreation trips taken over the course of a season and the costs associated with each trip. On the other hand, CVM data provides information on hypothetical behavior which provides insights on how an individual might behave under different circumstances. Combining TCM and CVM data has been shown to increase statistical efficiency (González-Sepúlveda, Loomis, & González-Cabán, 2008) as well as attenuate some problems that are particular to each data type (Cameron, 1992). Another key contribution of this dissertation is that it presents a utility-consistent joint TCM-CVM model of wildlife watching recreation using FHWAR data. The unique feature of this model is that it unifies the TCM and CVM modeling frameworks so that the preference structures, the scenarios under analysis, and the welfare measures used to assess the impact of the scenarios, are the same. In particular, the equivalence between the welfare measures (consumer surplus and compensating variation) is explicitly incorporated into the econometric framework which has not been done in previous studies.

A common problem faced by nearly all researchers employing survey data is when a respondent fails or refuses to answer a survey question, which can lead to a loss in statistical efficiency and bias. This is an important problem when

estimates are used to determine a course of action that can affect a large population. With missing data points, statistical inference may not accurately reflect, for example, the preference structure of wildlife watchers, and policy recommendations can be severely misleading because certain groups in the target population may be under or over-represented. This dissertation contributes to the literature on nonresponse in nonmarket valuation surveys by: 1) introducing a new way of applying the multiple imputation imputation to missing TCM data, and 2) developing simple recoding procedure for missing CVM data.

The substantive parts of this dissertation are contained in three chapters. In chapter 2, the demand for wildlife watching recreation is estimated using a finite mixture (FM) model to examine preference heterogeneity. FM models are especially useful for examining and modeling preference heterogeneity as it allows the grouping of distinct user classes within a TCM regression framework. The FM and other latent-class modeling techniques are considered to be at the forefront of recent innovation in recreation demand modeling. Using data from the 2006 FHWAR survey, the application of the FM model will aid in examining the composition of the wildlife watching population as well as recovering the estimated net benefits received by the differing user classes. However, estimating TCM models requires data on the number of trips and the associated travel costs for each respondent. One issue addressed in this chapter is missing travel cost data for nonparticipants—respondents who had not taken any trips during the survey period. Missing travel cost data is problematic because these observations or respondents will be excluded from the analysis by way of list-wise deletion.

In addition to data on trip-related activities, the 2006 FHWAR survey included responses to CVM data that elicited the respondents' *choke price*—the amount of cost that would have prevented the respondent from taking a trip. The conceptual innovation of this chapter is that the CVM choke price data obtained from respondents who had taken at least one trip, can be used as a reference to fill-in missing travel cost values for nonparticipants. This allows the analyst to avoid list-wise deletion and the TCM analysis and inference can be carried out at the population level. The multiple imputation (MI) method was employed to carry out the data imputation process. This technique has received very little attention in the nonmarket valuation literature despite the vast amounts of studies supporting its use to address nonresponse in large scale survey data. In fact, this is the first study where the MI method was applied to a TCM framework and also the first time CVM data was used for the purpose of filling-in missing travel cost values.

In chapter 3, TCM count data and CVM dichotomous choice data from the 1996 FHWAR survey are combined into a unified joint model of wildlife watching recreation. Much of the discussion focuses on the alignment of TCM and CVM theoretical and econometric frameworks. The purpose of this chapter is to advocate for survey designs that can yield a higher degree of cohesion between the TCM and CVM frameworks both theoretically and econometrically. In particular, the language of the CVM questions may not always be compatible with the TCM framework and therefore the two sets of information may not contribute equally towards understanding the scenario under consideration. In addition, the benefits of combining data such as attenuating hypothetical bias and

obtaining nonuse values, may not be fully realized. Although the survey employed in this chapter is outdated, it is still useful for demonstrating a case where the potential level of cohesion of the TCM and CVM information is high. Several econometric models are estimated to demonstrate not only the advantages of combining TCM and CVM data, but also to show that different approaches to joint modeling also can influence levels information cohesion. Finally, the chapter looks at the special case where the welfare measures arising from the TCM and CVM models (compensating variation and consumer surplus) are equal for the TCM and CVM scenarios reflected in the data.

The third and final segment of this dissertation focuses on addressing nonresponse in double-bounded dichotomous choice (DB) CVM data. The DB format in CVM surveys was introduced as a way of obtaining greater statistical efficiency compared to a single-bounded (SB) format. As in a SB format, DB asks respondents if they would be willing to pay a certain amount of money, or bid value, contingent upon a given scenario, where the bid value is the basis for measuring willingness to pay (WTP). The difference is that DB includes a *follow-up* question that depends on the response to the *initial* question. For example, suppose the respondent answers *yes* to the initial question. Then a follow-up question asks if they (respondent) would be willing to pay a higher amount, keeping the scenario constant. If the respondent answers *no*, then the follow-up question asks if they would be willing to pay a lower amount, again keeping the scenario constant.

Since its initial introduction, researchers have noticed certain tendencies of respondents to answer the follow-up question in a way that is not consistent

with the way they answered the initial question. For different reasons, the respondent may behave strategically in order to obfuscate their true preference structure. The case examined in this chapter is when the respondent is either unable or refuses to answer the follow-up question. Similar to other types of nonresponse, partial responses leads to the observation being excluded from the analysis through list-wise deletion. The purpose of this chapter is to introduce a simple recoding procedure for DB data with nonresponse that allows the analyst to preserve the observation with the missing data and curtail the loss of statistical efficiency and potential bias from list-wise deletion. The main objective is to compare the statistical efficiencies and biases from recoding and list-wise deletion using Monte Carlo simulations.

The overarching aim of this dissertation is to provide a better understanding of the different modeling approaches available for combining TCM and CVM data within the context of wildlife watching recreation. The results of this dissertation provides important insights for wildlife and land managers interested in obtaining theoretically consistent measures of the welfare effects from recreation and also options for combining disparate types of data. In addition, the discussions concerning nonresponse are useful for researchers interested in implementing small-scale nonmarket valuation surveys due to limited budgets and concerned about maintaining the sample size.

## **Chapter 2: Finite Mixture Model of Wildlife Watching with Multiply-imputed Travel Costs**

### **2.1. Introduction**

Wildlife watching recreation continues to be one of the most popular forms of outdoor activities in the United States. In 2011, an estimated 71.8 million U.S. residents enjoyed observing, feeding, or photographing wildlife (USFWS, 2012a). Steady growth in the number of wildlife watchers since 1996 (14 percent) has contributed substantially to the health of the nation's, states', and local economies. For example, spending on wildlife-watching activities generated \$18.2 billion in tax revenues and over a million jobs (Leonard, 2008). Total industry output is estimated to be \$122.6 billion with major U.S. sectors including retail trade, manufacturing, and hospitality and food services. While wildlife watching has many benefits to tourism, there are also important implications for wildlife management.

The growing interest by domestic and international visitors has prompted concerns over the sustainability of the wildlife watching industry. Excessive visitation may disturb local animal populations and cause them to deviate from their normal behavior. These changes could have serious physiological effects on wildlife with long-term implications for the survival of some species (Tapper, 2006). A better understanding of the disparate preference structures of wildlife watchers, which has been linked to their behavior, can provide important insights



on how to help ensure sustainable wildlife populations while not stifling the growth of local economies.

Data from national surveys such as the National Survey of Fishing, Hunting, and Wildlife-Associated Recreation have been used extensively for studying the demand for wildlife watching (Hay & McConnell, 1979; Hay & McConnell, 1984; Rockel & Kealy, 1991; Zawacki, Marsinko, & Bowker, 2000) but have not yet been examined for preference heterogeneity. In this paper, it is hypothesized that the wildlife watching population can be segmented into heterogeneous user classes that differ systematically in their visitation patterns. Consequently, group-level responses to policy changes may vary considerably. For example, site managers seeking to curtail excess visitation may find increasing access fees effective for some visitors, while adjusting visitation hours works better on others. Hence, knowing the composition of wildlife watchers can lead to a more accurate analysis for policy assessment. A finite mixture (FM) model was employed to estimate the composition of the wildlife watching population and the group-level demand functions using data from the 2006 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. The FM and other latent-class modeling techniques are considered to be at the forefront of recent innovations in recreation demand modeling (Moeltner & von Haefen, 2011).

In addition to data on trip-related activities, the survey included responses to contingent valuation method (CVM) questions that elicited the respondents' *choke price* for participation—that is, the amount of cost that would have prevented a current wildlife watcher from taking a trip (See Aiken, 2009 and

USFWS, 2012b). The most direct application of the CVM responses (choke prices) is to assume a linear demand function so that choke price can be combined with travel cost and trip frequency to calculate the respondent's consumer surplus or net benefits from wildlife watching recreation activities (Aiken & Rouche, 2003; Aiken, 2009; Hwang, 2012; USFWS, 2012b). One limitation of this approach is that the analysis is exclusive to respondents who had reported taking at least one trip. Incidentally, the survey did not collect data on travel cost from respondents who did not take any trips during the survey period; thus, the usable sample is truncated to observations with positive trips.

One approach to address this type of truncation is to employ a zero-truncated count data model (Cameron & Trivedi, 1998). For zero-truncated count data models (in particular, the negative binomial model), an important consideration is that the theoretical frequency of zeros predicted by the model depends on the distribution of the positive counts (Hilbe, 2011). For instance, the theoretical probability of observing a zero value will be lower when the mean of the positive counts is large compared to when the mean is small. This mechanism however, can misrepresent the data if there are outliers or extreme values in the distribution accompanied by a high frequency of zeros. These values have the effect of inflating the conditional mean and in turn, the frequency of zero values predicted by the model underrepresents the actual number of zeros in the data. In the context of recreation demand, the coefficient for the travel cost variable is biased upwards, hence a flatter inverse demand curve, and the estimated consumer surplus is biased downwards.

In this paper, the multiple imputation (MI) method (Rubin, 1987) is employed to impute the missing travel cost data so that data truncation is no longer an issue. With the imputed values incorporated into the FM model framework, the analysis revealed the presence of three distinct sub-populations or user classes in the wildlife watching sample. The groups can be characterized by their intensity of participation and the amount of spending. The performance of the MI FM model is compared to that of a zero-truncated FM model and found to provide a better fit to the data.

To the best of my knowledge, this study is the first to estimate a FM model of wildlife watching recreation, and also the first to apply the MI method within a FM model framework. The remainder of the paper is organized as follows. The next section provides a review of the relevant literature. Section 2.3 presents the theoretical framework of the FM model, the benefits formulas, and the MI method. Section 2.4 presents the data and results of the analysis. Finally, section 2.5 concludes the paper with a discussion of the results and avenues for future research.

## **2.2. Literature Review**

### *2.2.1. Methods for Valuing Wildlife Watching Recreation*

Much of the existing research on wildlife watching recreation used data from national surveys on outdoor recreation. Cicchetti (1973) used bird watching and wildlife photography data from the 1965 national survey of wildlife-associated recreation to estimate a simultaneous equations model. Hays and McConnell (1979) used the 1975 national survey to model participation in wildlife

watching and photography. The authors found the amount of forested acres and the number of species were greater in states with higher participation rates. The authors later extended the analysis to a joint decision model of wildlife watching and hunting participation and found the two activities may be complementary (Hay & McConnell, 1984). In Rockel and Kealy (1991), wildlife watching was modeled as a two-stage decision process, treating the participation decision as being influenced by different factors than what might motivate the intensity of participation. Using data from the 1980 national survey, the authors proposed using Heckman's selection bias model (Heckman, 1979), and Cragg's hurdle model (Cragg, 1971). Similar to Hays and McConnell (1979; 1984), the authors found the amount of forested acres to positively influence participation and also intensity. Aggregate annual willingness to pay for access to wildlife watching recreation were between \$8.7 billion and \$164.5 billion for the 1980 sample, which was comparable to those of the 1991 sample studied by Zawacki, Marsinko, and Bowker (2000) who applied similar techniques. Beginning 2001, the U.S. Fish and Wildlife Service's National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (FHWAR) began collecting contingent valuation method (CVM) data on choke prices. Aiken and Rouche (2003), and later Aiken (2009), report consumer surplus estimates using the direct calculation method mentioned earlier. Hwang (2012) obtained consumer surplus values the same way but used the data to test for temporal reliability across the 2001 and 2006 surveys. However, all the abovementioned studies assumed homogeneous preferences for the wildlife watching population.

### *2.2.2. Preference Heterogeneity in Recreation*

Wildlife watching recreation is a natural candidate for exploring preference heterogeneity using the finite mixture approach. Participants can vary greatly in skill level, experience, and economic constraints, which gives rise to different visitation patterns and perceived values of participation. The FM model is especially useful for when there are different user classes or groups of individuals who are like-minded and share certain behavioral traits, which may be particularly useful for recreational site managers (Hynes, Hanley, & Scarpa, 2008)<sup>i</sup>. In a study of Lake Michigan anglers, Provencher, Baerenklau, and Bishop (2002) identified three angler classes distinguished by their time costs and trip frequencies. Interestingly, their finding showed anglers with high time costs (and low trip frequencies) were comparably more sensitive to changes in fishing conditions. In another study on anglers, Morey, Thacher, and Breffle (2006), used responses to attitudinal questions on topics such as boating fees, catch rates, and fish consumption advisories to identify angler classes. Similar to Provencher, Baerenklau, and Bishop (2002), the authors found three to be the optimal number of angler classes. Evidence of finite mixing in other forms of recreation, for example: rock climbing (Scarpa & Thiene, 2005; Scarpa, Thiene, & Tempesta, 2007); kayaking (Hynes, Hanley, & Scarpa, 2008); backcountry hiking (Baerenklau, 2010); and beaching (Landry & Liu, 2009; Kuriyama, Hanemann, & Hilger, 2010), seems to suggest that preference heterogeneity is now a common

---

<sup>i</sup> Because the focus here is on finite mixture models, we omitted the literature that addresses preference heterogeneity using a random parameters model. While there are no theoretical reasons to use one over the other, the finite mixture model has been viewed as a nice alternative to the random parameters approach (Provencher, Baerenklau, & Bishop, 2002). The reader is referred to Train (1998), Chen & Cosselett (1998), Breffle and Morey (2000), Holmes and Englin (2010).

feature in recreation demand models and is generally well characterized by finite mixtures. FM models have also been used in the past for handling count data with excess zeros (See Gurmu & Trivedi, 1996; McLachlan & Peel, 2000).

### *2.2.3. Imputation of Nonparticipant Travel Costs*

For many recreation demand studies, the survey focuses on a specific site so that distance-based travel costs are imputed using information on the respondent's place of origin regardless of whether the site was actually visited by the respondent or not (Hellerstein, 1991; Yen & Adamowicz, 1993; Haab & McConnell, 1996; Scarpa, Thiene, & Tempesta, 2007; Landry & Liu, 2009; Kuriyama, Hanemann, & Hilger, 2010; Baerenklau, 2010; Holmes & Englin, 2010; and others)<sup>ii</sup>. However national surveys, in particular FHWAR, will routinely censor information pertaining to respondents' place of residence as well as their destinations in the public-use data in order to maintain confidentiality. FHWAR surveys collect travel cost data from respondents but only from those reported to have taken at least one trip during the survey period. Some previous studies imputed missing data with average travel cost values in the respondent's resident state (Zawacki, Marsinko, & Bowker, 2000), or with fitted or predicted values following a regression on the observed data (Hellerstein, 1998). Mean and regression imputations have also been suggested for imputing missing data from CVM surveys (Whitehead, Groothuis, & Blomquist, 1993; Whitehead, 1994; Brox,

---

<sup>ii</sup> English and Bowker (1996) compared several distance-based measures with imputed values and noted the various discrepancies and effects on welfare values. But the paper's main focus was on exploring different price definitions and much less was said about the limitation of the imputation procedure used.

Kumar, & Stollery, 2003). Here, the multiple imputation (MI) method (Rubin, 1987) is proposed for imputing the missing data.

MI is a technique specifically designed for handling nonresponse in large public-use data such as population surveys and censuses (Rubin, 1987; 1996)<sup>iii</sup>. MI produces multiple simulated-datasets (imputations) using a model designed to capture the missingness (imputation model). Each of the imputations are analyzed using a method (for example, regression model) chosen by the analyst, and the results are combined using Rubin's (1987) combination rules to produce estimates and confidence intervals that incorporate the missing-data uncertainty. One advantage of using MI is that it can incorporate the data collector's knowledge concerning the missingness of values (Rubin, 1987; pp.15-16). The information could also be related to sampling or interview methodology. MI also overcomes some major problems from using mean and regression imputations such as the tendency to distort final estimates and statistical associations between variables (Lessler & Kalsbeek, 1992; Little & Rubin, 2002).

One important consideration concerning the results is that imputed values may not always reflect a nonparticipant's choke price but may sometimes correspond to corner solutions. Smaller values may be more reflective of corner solutions, while larger values may suggest nonparticipation from realizing a choke price. The treatment of nonparticipation in this paper is very different

---

<sup>iii</sup> MI has been used extensively to impute missing income and wage data (Schenker, et al., 2006; Zarnoch, Cordell, Betz, & Bergstrom, 2010; Durrant & Skinner, 2006), and to improve financial data with severe levels nonresponse (Kofman & Sharpe, 2003). In simulation studies, using MI provided estimates that were less sensitive to model misspecification compared to top coded data, while adequately maintaining disclosure protection of confidential data (An & Little, 2007). Moreover, MI has been suggested for merging data from two surveys where one of the surveys is missing a key variable that is imputed using MI (Brown, 2002). Kohnen and Reiter (2009) show how disclosure protection can still be maintained through two stages of MI during merging. Shafer (1999) notes the great virtues of MI are its simplicity in execution and generality of application.

from, for example the Kuhn-Tucker method (Wales & Woodland, 1983; Phaneuf & Herriges, 2000), which assumes all nonparticipation arise from corner solutions. The idea is more similar to that applied in the Double Hurdle model proposed by Shonkwiler and Shaw (1996), but only in that it attempts to differentiate ‘potential’ participants, who may have realized their choke price, from nonparticipants at corner solutions.

### **2.3. Methodology**

#### *2.3.1. Finite Mixture Model*

The finite mixture model methodology is briefly explained in this section in general form. Let  $y_i$   $\{i = 1, 2, \dots, n\}$  be an outcome variable for observation  $i$  in a sample with size  $n$ , and probability density function  $f(y_i)$ . Suppose  $y_i$  is drawn from a population comprising an additive mixture of  $G$  distinct subpopulations in proportions  $\pi_1, \pi_2, \dots, \pi_G$   $\{g = 1, 2, \dots, G\}$ . Given a vector of covariates  $x_i$  and respective parameters  $\varphi_g$ , a  $G$ -component finite mixture model can be written in the form

$$(1) \quad f(y_i; \Phi) = \sum_{g=1}^G \pi_g f_g(y_i; \varphi_g)$$



where  $f_g(\cdot)$  is the density function for component  $g$ , and  $\pi_g$  is a mixing probability with  $\sum \pi_g = 1$ , and  $0 \leq \pi_g \leq 1$ <sup>iv</sup>. Although not required, it is convenient and computationally simpler to restrict the densities to be the same<sup>v</sup>. The log likelihood function for  $\Phi$  using the observed data has the form  $\ell(\Phi) = \sum \ln f(y_i; \Phi)$  where the maximum likelihood estimate  $\hat{\Phi}$  can be obtained by solving the likelihood equation  $\partial \ell(\Phi) / \partial \Phi = 0$  using numerical methods, or via the Expectation Maximization (EM) algorithm (Dempster, Laird & Rubin 1977). Incidentally,  $\hat{\Phi}$  satisfies  $\sum_g \sum_i \pi_{g,i}(y_i; \hat{\Phi}) \partial f_g(y_i; \hat{\varphi}_g) / \partial \Phi = 0$ , and  $\hat{\pi}_g = \sum_i \pi_{g,i}(y_i; \hat{\Phi}) / n$  is the estimated posterior probability that  $y_i$  belongs to component  $g$  with the usual conditions  $\sum_g \pi_{g,i}(\cdot) = 1$ ,  $0 \leq \pi_{g,i}(\cdot) \leq 1$ , and  $\pi_{g,i}(y_i; \hat{\Phi}) = \hat{\pi}_g f_g(y_i; \hat{\varphi}_g) / \sum_h \hat{\pi}_h f_h(y_i; \hat{\varphi}_h)$  (McLachlan & Peel 2000; p.47).

A negative binomial distribution is used to model the integer nature of recreation trips and data over-dispersion. Following Cameron and Trivedi (1998), the densities of a  $G$ -component mixture negative binomial model can be expressed as

$$(2) \quad f_g(y_i; \varphi_g) = \frac{\Gamma(y_i + \eta_{g,i})}{\Gamma(y_i + 1)\Gamma(\eta_{g,i})} \left( \frac{\eta_{g,i}}{\eta_{g,i} + \mu_{g,i}} \right)^{\eta_{g,i}} \left( \frac{\mu_{g,i}}{\eta_{g,i} + \mu_{g,i}} \right)^{y_i}$$

where  $y \in \mathbb{Z}_+$ ,  $\mu_{g,i} = \exp(x_i' \beta_g)$  and  $\eta_{g,i} = (1/\alpha_g) \mu_{g,i}^t$ , and  $t = 1$  gives the constant dispersion negative binomial model (NB1) and  $t = 0$  gives the mean dispersion

---

<sup>iv</sup> As mentioned in Cameron and Trivedi (1998; p.128), it is useful to parameterize the grouping probabilities using, for instance, a logit function:  $\pi_g = \exp(\gamma_g) / (1 + \sum \exp(\gamma_g))$  to help ensure  $0 \leq \pi_g \leq 1$ , where  $\gamma_g$  may be further parameterized in terms of observable covariates

<sup>v</sup> In the present formulation, the number of mixture components  $G$  is fixed and set by the analyst prior to estimation, but this need not be the case (See McLachlan and Peel, 2000 and references therein).

negative binomial model (NB2). The corresponding conditional mean and variance functions of the mixture model are  $E[y_i|x_i; \pi_g] = \sum_g \pi_g \mu_{g,i}$ ; and  $\text{var}[y_i|x_i; \pi_g] = \sum_g (\pi_g \mu_{g,i} [1 + \alpha_g \mu_{g,i}^{-t}]) + \bar{\mu}_i - (\bar{\mu}_i)^2$ .

### 2.3.2. Estimation of Net Benefits from Access to Recreation

Hellerstein and Mendelsohn (1993) showed if the outcome variable  $y^*$  follows a count distribution (Negative Binomial with  $k = 0$  in this case) with mean  $\bar{y}^* = \mu$ , the expected consumer surplus for the sample can be calculated by integrating under the estimated mean function  $\hat{\mu}$  from an initial travel cost level  $c^0$  to a choke price  $c^1 > c^0$ . In the present context, the expected consumer surplus reflects the net economic benefits from access to recreational opportunities. An exponential functional form is used to model the mean count function; hence  $\hat{\mu} = n^{-1} \sum \exp(x_i' \hat{\beta}) = n^{-1} \sum \exp(x_i' \hat{\beta} | c^0)$ , where  $n$  is the sample size,  $x_i'$  is the vector of covariates for observation  $i$ ,  $\hat{\beta}$  is the vector of coefficient estimates, and  $c^0$  and  $c^1$  are nonnegative constants. The expected consumer surplus can be expressed in the form

$$(3) \quad E[CS] = \hat{\beta}_c^{-1} \{ \exp(\hat{\beta}_c (c^1 - c^0)) - 1 \} \cdot \hat{\mu}$$

where the estimated travel cost coefficient  $\hat{\beta}_c < 0$  is assumed<sup>vi</sup>. Hellerstein and Mendelsohn (1993) argued that (3) is an appropriate measure of expected

---

<sup>vi</sup> The per-trip consumer surplus is found by dividing (1) by  $\hat{\mu}$ , which gives  $E[cs] = \hat{\beta}_c^{-1} \{ \exp(\hat{\beta}_c (c^1 - c^0)) - 1 \}$ .

consumer surplus because the random errors in a count data model are incorporated in a parametric fashion rather than as a residual<sup>vii</sup>. Furthermore, the authors showed that (3) approximates the expected compensating variation measure  $E[CV]$  by demonstrating how count models can be derived from repeated discrete choices similar to those in a Random Utility model framework. The standard practice is to evaluate the limit of (3) as  $c^1 \rightarrow \infty$  which reduces to

$$(4) \quad E[CS|c^1 = \infty] = -\hat{\beta}_c^{-1} \cdot \hat{\mu}$$

This is a useful assumption as it guarantees correspondence with  $\hat{\mu}|_{\infty} = 0$  in the limit. In addition, (4) will equal  $E[CV]$  since the Marshallian and Hicksian demands will overlap for price increases that force  $y^*$  to zero. This equivalence is more formally demonstrated in Hellerstein and Mendelsohn (1993). However, one might ask about the implications of allowing  $c^1 \ll \infty$  and if doing so could help obtain more accurate estimates. In this light, (4) would provide an upper bound measure of (3). It is clear from (3) that restricting  $c^1$  to be finite will in general yield estimates that are more conservative, especially when the magnitude of  $\hat{\beta}_c$  is small. If  $|\hat{\beta}_c|$  is large, the difference between using  $c^1 = \infty$  and  $c^1 \ll \infty$  will be smaller. Given the availability of choke price data, estimates for both scenarios were obtained. In the FM framework, the wildlife watching sample was partitioned into  $g$  components. The overall expected consumer

---

<sup>vii</sup> See Bockstael and Strand (1987); Hellerstein (1992); Yen and Adamowicz (1993); Habb and McConnell (2002); and Beatty, Brozovic, and Ward (2005) for a comparison of discussions on this topic.

surplus is obtained by computing the weighted average of the component-level estimates.

$$(5) \quad \overline{CS} = \sum_{g=1}^G \hat{\pi}_g E[CS]_g$$

where  $\hat{\pi}_g$  is the estimated mixing probability for component  $g$ . For comparing models with  $c^1 = \infty$  and  $c^1 \ll \infty$ , the sample mean of the CVM responses and travel costs for  $c^0$  and  $c^1$  respectively was used in (3).

### 2.3.3. Predictive Mean Matching Multiple Imputation

The predictive mean matching (PMM) multiple imputation method (Rubin, 1987; Little, 1988) was employed to obtain imputed travel cost values. Based on Rubin's (1986) ideas for statistical file matching, PMM combines the linear regression imputation and the nearest-neighbor imputation methods to fill in missing data points. Let  $x_o \in x_o$  be an observed data point and  $x_m \in x_m$  be a missing data point, where  $x = (x_o | x_m)'$  is a partitioned column vector of the analysis model predictor matrix  $X$ . The first step is to fit a linear regression model to the observed data following  $x_o | Z_o \sim \mathcal{N}(Z_o' \gamma, \sigma^2)$ , where  $Z_o$  is a matrix of covariates. Using the estimates  $(\hat{\gamma}, \hat{\sigma}^2)$ , fitted values  $\hat{x}_o$  are obtained and a new set of parameters  $(\tilde{\gamma}, \tilde{\sigma}^2)$  are simulated from their joint posterior distribution  $\mathcal{P}(\tilde{\gamma}, \tilde{\sigma}^2 | x_o)$  having the standard uninformative improper reference prior  $P(\gamma, \sigma^2) \propto 1/\sigma^2$ . In particular, the parameter  $\tilde{\sigma}^2$  is drawn from the  $\chi^2$

distribution with  $n_o - q$  degrees freedom, where  $n_o$  is the sample size of the observed data, and  $q$  is the dimension of  $\hat{\gamma}$ . Using  $\tilde{\sigma}^2 \sim \hat{\sigma}^2(n_o - q)\chi_{(n_o - q)}^2$ ,  $\tilde{\gamma}$  is drawn following  $\mathcal{N}(\hat{\gamma}, \tilde{\sigma}^2(Z_o'Z_o)^{-1})$  giving the predicted values  $\tilde{x}_m$  which are drawn from the posterior predictive distribution  $\mathcal{P}(\tilde{x}_m|x_o)$  (Zhang, 2003)<sup>viii</sup>. Missing values  $x_m$  are imputed by matching  $\tilde{x}_m$  with  $\hat{x}_o$  that minimizes the distance metric  $D(\hat{x}_o, \tilde{x}_m) = |\hat{x}_o - \tilde{x}_m|$ , then taking the corresponding  $x_o$  as its imputation. If more than one candidate value is allowed, say the first  $k$  minimums, one value is randomly chosen for imputation. The process is usually repeated several times giving multiple imputations.

A useful feature of PMM is that only actual values are used for imputing. This allows the analyst to preserve the distribution of variables in the filled-in data and avoid extrapolating beyond the original data range. This also allows the model to be less sensitive to misspecification (Little, 1988; Schenker & Taylor, 1996). It was assumed that the data missingness mechanism is *ignorable* or the data are missing at random (MAR). MAR refers to the situation where the missingness is independent of the missing data given the observed values (See Zhang, 2003 for a theoretical depiction)<sup>ix</sup>.

The observed data from which the missing values are imputed included rows of both travel cost values and CVM responses which initially increase the

---

<sup>viii</sup> Zhang (2003, p.584) explains if the values of the parameters are drawn from their posterior predictive distribution  $\mathcal{P}(\tilde{\gamma}, \tilde{\sigma}^2 | \mathbf{x}_o)$ , the corresponding draws from the conditional predictive distribution  $\mathcal{P}(\tilde{x}_m | \mathbf{x}_o, \tilde{\gamma}, \tilde{\sigma}^2)$  are equivalent to drawing from the posterior predictive distribution  $\mathcal{P}(\tilde{x}_m | \mathbf{x}_o)$ .

<sup>ix</sup> Schafer (1999) notes however, that the MI paradigm does not require the MAR assumption and will still produce valid results as demonstrated by Glynn, Laird, and Rubin (1993). In the present case, missingness of travel costs can be conditioned on trip counts and other factors so the MAR assumption is expected to hold.

usable sample size<sup>x</sup>. This was done to incorporate as much information into the imputation model as possible. It was assumed that there are travel cost values corresponding to all levels of participation which seems necessary since travel costs do not always carry the usual marginal adjustments in quantity-demanded interpretation when no trips are taken. Thus, imputing based only on CVM responses seems too unrealistic because the language of the CVM questions requires a response where cost is the only reason for not participating. Although the survey was designed to screen out individuals suspected to never participate, interested individuals may still fail to participate for reasons unrelated to costs—for example, health ailments. Lastly, all information associated with when a trip is actually taken would be excluded, hindering the imputation model from fully capturing the relationship between cost and its determinants.

Referencing only travel cost observations seems limiting also. The imputation model may inadvertently dismiss the possibility that nonparticipation arose due to choke prices. Again, the imputation model would ignore pertinent information on behavior. By utilizing both types of information, the imputations will be based on the full range of trip values. Therefore, the imputed-travel costs should differentiate nonparticipants by what they may consider to be their *perceived* benefits of having access to wildlife watching recreation. Because inferences are based on current or observed behavior, data on CVM responses were excluded during the estimation of the parameters of the FM model.

Given  $M > 1$  imputed data sets,  $M$  sets of parameter estimates are obtained independently. For instance, let  $\hat{\Phi}_m \{m = 1, 2, \dots, M\}$  be the vector of parameter

---

<sup>x</sup> Similar situation concerning nonignorable missing data can be found in Glynn, Laird, and Rubin (1993).

estimates using the  $m^{\text{th}}$  imputed data. Then, the MI estimate of  $\Phi$  is obtained by computing the average of the  $M$  estimates as follows.

$$(6) \quad \bar{\Phi} = \frac{1}{M} \sum_{m=1}^M \hat{\Phi}_m$$

The estimated total variance of  $(\Phi - \bar{\Phi})$  is given by

$$(7) \quad T = \bar{U} + (1 + M^{-1})B$$

where  $\bar{U} = M^{-1} \sum \hat{U}_m$  is the within-imputation variance, and  $B = \sum (\hat{\Phi}_m - \bar{\Phi})^2 / (M - 1)$  is the between-imputation variance (Rubin, 1987; p.76). Hypothesis tests and confidence intervals are based on a Student's  $t$ -approximation  $(\bar{\Phi} - \Phi) / \sqrt{T} \sim t_\nu$  with  $\nu$  degrees of freedom. Under the large-sample assumption  $\nu = (M - 1)(1 + 1/r)^2$ , where  $r = (1 + M^{-1})B / \bar{U}$  measures the relative increase in variance (RVI) due to missing data. Alternatively, the percentage increase in standard error due to missing data (SEI) can be obtained using the formula  $\% \Delta SE = \left\{ \sqrt{T / \bar{U}} - 1 \right\} \times 100\%$ . Finally, the relative efficiency (RE) of an estimate is approximated by  $(1 + \gamma / M)^{-1}$ <sup>xi</sup>, where  $\gamma = (r + 2 / (\nu + 3)) / (1 + r)$  is the estimated fraction of missing information about  $\Phi$  due to nonresponse (FMI). For brevity, only the RVI and FMI values are reported.

---

<sup>xi</sup> This measure provides the efficiency of the current MI estimates compared to ones that used an infinitely large number of imputations.

An important feature of MI is that the imputation step is operationally distinct from the subsequent analysis; that is, the imputation model used to create the imputed datasets need not be compatible with the analysis model (Shafer, 1999). When the imputation model is more general than the analysis model, MI gives valid inferences but with some loss of power due to the added variation from the imputed values. When the imputation model is more detailed, the imputations may be characterized as what Rubin (1996) calls ‘superefficient’. Superefficient imputations contain information pertaining to  $\Phi$  that goes beyond the complete-data estimate  $\hat{\Phi}$ , if for example, the researcher possess *a priori* knowledge of the true distribution of the subject variable. Superefficient imputations can give  $\Phi$  estimates that are more precise than could be achieved using the observed data and the analysis model alone. Some authors caution against introducing too much detail in the imputation model as illegitimate assumptions can lead to biased estimates (Shafer, 1999). Still, imputation models should include all variables believed to help predict the subject variable, and all variables used in the analysis model including non-exogenous variables to ensure that a multiple imputation is *proper*<sup>xii</sup> (Little, 1988; Rubin, 1996).

## **2.4. Application to Wildlife Watching Recreation Data**

### *2.4.1. Data*

Wildlife watching data was obtained from the 2006 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. The survey is conducted every 5

---

<sup>xii</sup> A multiple imputation is said to be *proper* if it satisfies certain frequentist properties such as preserving the distributional relationships between the missing values and the observed values (See Zhang, 2003 for a summary of the consequences from improper imputations). However, confidence-valid repeated-imputation can still be obtained even in the case of improper multiple imputations (Rubin, 1987; 1996).



years by the U.S. Fish and Wildlife Service in conjunction with the U.S. Census Bureau. The data includes an array of information on wildlife watching activities by residents around their homes and at least one mile away. Data collection was carried out in two phases: an initial screening survey to identify current and potential wildlife watchers, and a series of follow-up interviews on the selected sample to collect detailed data on their wildlife watching activities. The final sample consisted of 11,285 unique records that reported the total number of trips to a wildlife watching destination at least one mile from their home and the respective expenditures. CVM questions were included for all respondents who had taken at least one trip. More specifically, the participants were asked: *what is the cost that would have prevented you from taking even one such trip? In other words, if the trip cost was below this amount, you would have gone observing, photographing, or feeding wildlife..., but if the trip cost was above this amount, you would not have gone.* Two sets of CVM responses were collected: one pertaining to trips in their resident state, and one for trips in another state randomly chosen by the survey.

Data points that were irregular or had unresolvable inconsistencies were deleted but these were relatively small in numbers (70 records). For example, some respondents reported having not taken any trips but also reported having observed, photographed, or fed fish during a trip. Also, 45 records were deleted because of missing values in one of more of the other model variables. The total number of unique records that remained was 11,161 (98.9%), where 8,453 records were nonparticipants with missing travel costs. Similar to previous users of this survey (Rockel & Kealy, 1991; Zawacki, Marsinko, & Bowker, 2000), the analysis

model treats data on trips to multiple locations by the same respondent as additional independent records. The sample size with this addition was 12,224 records. A list of the variables, their descriptions, and summary statistics are provided in table 2.1.

The imputation model included an additional 3,091 records of CVM responses (choke prices) by participants which correspond to a hypothetical version of nonparticipation where the number of trips equals zero. Here, CVM responses showing \$10,000 or higher (33 records) were deleted as outliers giving a total of 15,315 records. Therefore, 6,862 observed data points were used to impute 8,453 missing travel cost values for nonparticipants. As mentioned before, the additional data rows were excluded from the FM procedure so estimates are based on current behavior. The sampling weight provided in the survey was applied to all estimation procedures and the calculation of subsequent statistics. This weight is a combination of several adjustment weights that brings the wildlife watching sample up to the screening sample so as to allow inference about the general population (See USFWS, 2007). This weight was normalized to sum to the sample size of the analysis model<sup>xiii</sup>. Additional variables used elsewhere such as measures of state-level forest and rangeland compositions were constructed using data from the National Resources Inventory (USDA, 2009; 2012).

#### *2.4.2. PMM Results*

---

<sup>xiii</sup> Wedel, ter Hofstede, and Steenkamp (1998) demonstrated how estimates of the number of class types, class probabilities, and class-level parameters can be severely biased when sampling weights are not used with FM models in complex surveys. The authors noted measures of information criterion (AIC, BIC) tends to overshoot the optimal number of class types.

This section briefly explains the PMM procedures used and reports the imputation results. PMM was used to generate  $m = 10$  imputations with  $k = 3$  candidate donors for each missing value<sup>xiv</sup>. Simulations by Glynn, Laird, and Rubin (1993) showed their fully parametric mixture model performed well with  $m = 10$  even when the fraction of missing information (FMI) was large. To account for positive skewing, the natural log transform of the travel cost variable was used in the imputation phase. Following the suggestions by Rubin (1987, 1996) and others, PMM included all variables used in the trip demand model, and other factors believed to have influenced missingness (or nonparticipation) as regressors.

The mean estimates of the CVM responses as well as the observed travel costs at different trip counts are reported for comparison in table 2.2. The CVM mean is more than two times the mean of the imputed values. In addition, although the imputed values are on average larger than the overall travel cost mean, it is still smaller than the mean travel cost for a single trip. These initial findings suggest the presence of structural differences between the preferences of some respondents who have taken wildlife watching trips, and some who have not. The average relative increase in variance (RVI) was 0.461 for the imputed values and 0.428 for all values. With  $m = 10$ , this implies the between-imputations variance of the estimated mean was about 40 to 42 percent of the within-imputation variance. Finally, the largest fraction of information missing

---

<sup>xiv</sup> STATA<sup>®</sup> module `mi impute` was used in the multiple imputation stage, and modules `frm` and `frm1c` were used to estimate the coefficients and obtain summary tables following the `frm` procedures (Stata Data Analysis and Statistical Software, StataCorp LP, College Station, Texas). MI estimates of the parameters and associated statistics were computed by the author. A zero-truncated version of `frm` was programed by the author.

due to nonresponse (FMI) was 0.331 and 0.313, which implies about one-third of the information in obtaining the mean estimates was missing due to missing travel cost data.

### 2.4.3. Criteria for the number of mixture components

Two information criterion measures were employed to assess the optimal number of components: *Akaike Information Criterion*,  $AIC = -2 \ln L + 2p$  (Akaike, 1974) and the *Schwarz-Bayesian Information Criterion*,  $BIC = -2 \ln L + p \ln(n)$  (Schwarz, 1978). Information criteria are used to compare a model's goodness-of-fit for sample size  $n$  as additional parameters  $p$  are included in the model. In general, the model exhibiting the smallest criterion value is chosen. Another criterion measure is the *relative entropy index*,  $\varepsilon = 1 - \{(n \ln(G))^{-1} \sum_g \sum_i -\pi_{g,i} \ln(\pi_{g,i})\}$  (See Wedel & Kamakura, 2000). Although the  $\varepsilon$  measure does not provide an explicit guide to choosing the optimal number of components, it can help assess the performance of the mixture model by gauging how well the groups are separated (Celeux & Soromenho, 1996). More specifically,  $\varepsilon$  uses the estimated posterior probabilities to arrive at a measure of the degree of group distinctiveness. Values of  $\varepsilon$  lie between 0 and 1, where values close to 1 indicate greater distinctiveness and values close to 0 imply less distinctiveness<sup>xv</sup>. Table 2.3 reports the average values of the entropy and criterion measures from the 10 sets of MI estimates.

---

<sup>xv</sup> The relative entropy index  $\varepsilon$  is a scaled version of the entropy of fuzzy classification  $EN = \sum_{i=1}^n \sum_{g=1}^G -\pi_{ig} \ln(\pi_{ig})$ , where  $EN \in [0, \infty)$ . As mentioned by Celeux and Soromenho (1996),  $EN$  cannot be used directly to assess the number of components. The authors suggested a normalized entropy criterion but found some limitations in its applicability (p.202). The Wedel and

Four models were estimated for comparison and the optimal number of components indicated by the entropy and criterion measures is highlighted with a box in table 2.3. The AIC (25977.01) favored the 4-component model and possibly a 5-component model. The BIC (26472.60) however, reached its minimum with the 3-component model, and  $\varepsilon$  (0.596) was greatest for the 3-component model as well. Therefore the optimal number of components lies somewhere between 3 and 4 components with slightly more evidence supporting 3. It should be noted that the AIC has been observed to sometimes overestimate the optimal number of components (McLachlan & Peel, 2000; Wedel & Kamakura, 2000), so preference was given to the 3-component model suggested by the BIC and  $\varepsilon$ <sup>xvi</sup>.

#### 2.4.4. MI FM Model Results

The results of the 3-component FM model are reported in table 2.4. The single-component model results are reported also for comparison. The *travel cost* coefficient was statistically significant at the 1% level for all models while the  $\ln[\text{hincome}]$  variable was not significant for any except in component 3 of the 3-component FM model. Other demographic variables showed varying significance across the models, while the wildlife watching variables (*photo*, *feedbird*, *trip05*, *hunto6*, *fish06*) were all significant at the 1% level but this was so only in the

---

Kamakura (2000) relative measure has seen wider acceptance for use in latent class studies (See e.g., Morey, Thacher and Breffle 2006; Dias and Vermunt 2006; Dias and Vermunt 2008).

<sup>xvi</sup> Other measures similar to the AIC and BIC are the Consistent Akaike Information Criterion (CAIC) (See Bozdogan 1987), Modified AIC (MAIC), and sample size adjusted BIC (SBIC). The CAIC gives very similar values as the BIC especially with a large sample size, while the MAIC and SBIC gave values that more closely resembled the AIC. A 5-component model was attempted but the estimates could not be obtained due to difficulties with model convergence.

single-component model. Both the overall average RVI and largest FMI were smaller in the 3-component FM model, which favor the use of the 3-component model. Additionally, the overall estimated mean number of trips from the 3-component model  $\sum \pi_g E[\text{trips}|x'\beta]_g$  was closer in value (2.05) to the sample mean (2.00) compared to that of the single-component model (2.69).

The estimated probabilities of membership  $\pi_g$  from the 3-component model showed group representation was highly skewed. Nearly 79% of the wildlife watching population was designated to component 1, 3.4% to component 2, and 18.1% to component 3. From the estimated trip means  $E[\text{trips}|x'\beta]_g$ , it can be seen that component 1 was composed mostly of infrequent trip takers while component 2 included more frequent trip takers. Component 3 included a mixture of the first two components. The histograms in figure 2.1 illustrate the component distributions. The distribution of component 2 can be distinguished by its less pronounced rightward skew and more mass in the positive direction.

A closer examination of the results revealed that all nonparticipants were designated to component 1. Table 2.5 provides the mean values of model variables across the components<sup>xvii</sup>. The frequency count in the bottom row shows that nonparticipants ( $\text{trips} = 0$ ) represented nearly 77 percent (8453/11042) of component 1. The imputed travel cost (188.45) was slightly greater than the overall average (177.81) but less than the average for non-resident trip takers (222.02). Nonparticipants were on average older and included more retirees (0.24). The lower levels for the income and schooling

---

<sup>xvii</sup> To check for premature convergence of the FM model, 10 different starting values were used to help ensure convergence to a global maximum. Component designation was based on the maximum probability of component membership.

variables are consistent with this demographic, while there were no remarkable differences across the gender and marital status variables. Nonparticipants had substantially lower values for the around-the-home variables (*photo*, *feedbird*, *parks*), as well as for contributions (*wildcontr*), previous trip experience (*trip05*), and participation in other forms of outdoor recreation (*hunto6*, *fisho6*). Finally, a measure for the abundance of state forest and rangeland was lowest (8.49 and 0.43) for nonparticipants<sup>xviii</sup>.

The remaining 23 percent of component 1 spent an average of 3 to 4 days per trip. However, there was a stark difference in the forest and rangeland values of *out of state* trips and *in state* trips (18.58 versus 9.50). So, out-of-state travel for this group may be to locations with greater wildlife watching opportunities than in their resident state. Age, rate of retirement, income, level of schooling, and travel cost were slightly higher for out-of-state travel compared to in-state. There was also a lower rate of participation around their homes and higher rates of contribution to wildlife organizations. This could mean out-of-state trips are taken by more experienced and highly skilled wildlife watchers. In addition, it's been found that wildlife watchers may engage in different economic activities locally compared to when they travel (Davis, Tisdell, & Hardy, 2001).

Forest and rangeland values for component 2 (22.66 and 0.49) are similar to those for out-of-state travelers in component 1 (18.58 and 0.48). One explanation is that wildlife watchers in component 1 were traveling to the same

---

<sup>xviii</sup> The state forest and rangeland variables were excluded from the FM model because it was not clear how they should be interpreted for nonparticipants. Specifically, it was unclear if the assigned values should correspond to their resident state or another state since it is possible they may have travelled to other states given the right circumstances. Furthermore, it was believed the FM model sufficiently accounted for the effects of these variables through the mixture probabilities.

locations as wildlife watchers in component 2. The higher trip frequencies and lower travel costs for component 2 may indicate people are choosing to reside close to their frequented sites in order to reduce their travel costs. This behavior has been studied by Parsons (1991), and also by Baerenklau (2010) for backcountry hikers. The relatively low value of the *non-resident* variable (0.20) and high values for the forest and rangeland variables suggests component 2 members prefer participating locally because of more opportunities. Also, member of component 2 spent an average of \$60.74 per trip and stayed about 2.5 days each outing which were more than for component 3. This suggests component 3, which had even lower rates of out-of-state travel (0.14), consist of novice or casual wildlife watchers. In contrast, out-of-state-travelers in component 1 and members of component 2 resembled what could be considered skilled and experienced wildlife watchers. Overall, the post estimation analysis revealed there may actually have been as many as five classes of wildlife watchers.

#### *2.4.5. Expected Consumer Surplus Estimates*

The expected consumer surplus estimates derived from the MI model results are reported in table 2.6. The first two rows are estimates with infinity as the upper limit  $c^1 = \infty$ , and the second set of results used the mean of CVM response as the upper limit  $c^1 < \infty$  and the mean of travel cost as the lower limit  $c^0$ . Across the 3-component models, one pattern that emerges is that component 2 places the highest value for access to wildlife watching recreation, but not at the per-trip level. The highest per-trip values (shown in brackets) were found for component 1 which is consistent with the finding from earlier that these trips tend to be



longer and more expensive. As anticipated, consumer surplus estimates were larger when  $c^1 = \infty$ ; more than 1.5 times larger in the single-component model, and 4 times larger in the 3-component model. For the 3-component model, notice the overall differences mainly arise from component 1. Figure 2.2 illustrates why this might occur. This difference is shown as the striped area below the demand curve  $D_1$  which is excluded when a finite limit  $c^1$  is imposed. The surplus estimates for components 2 and 3 showed smaller differences perhaps because the demand curves lie farther below  $c^1$  as depicted by  $D_2$ . The overall estimates were also smaller for the 3-component model but the difference in estimates between the single-component model and the 3-component model was smaller when  $c^1 = \infty$  was used. In both cases, the single-component model may have given too much weight to individuals designated to component 2, which was estimated here to be only 3.4 percent of the wildlife watching population. The standard error estimates were smaller when  $c^1 = \infty$  was used, but this difference was less pronounced across components 2 and 3. The discrepancy is in part caused by the smaller magnitude of the coefficients in component 1 and its influence on the calculation of the variance function which was approximated using the Delta method. Overall, the difference in consumer surplus between models with  $c^1 = \infty$  and  $c^1 < \infty$  is quite large.

#### *2.4.6. Zero-truncated FM Model Results*

The parameter estimates of the zero-truncated models are reported in table 2.7. Similar to before, estimates from a single-component model are reported alongside the 3-component FM model. An analysis for the optimal number of

components showed the 3-component model was best but is suppressed here for brevity. The log-likelihood values indicate the 3-component model was more efficient than the single-component model. However, the conditional mean estimate  $\sum \pi_g E[\text{trips} | \text{trips} > 0]_g$  for the single-component model is relatively closer in value to the truncated sample mean. Compared to the MI model, the relative entropy index  $\varepsilon$  is very small (0.22 versus 0.596). This indicates the zero-truncated FM model performed somewhat poorly in identifying distinct groups in the truncated sample. One possibility is the data on the respondents that were omitted from the analysis contained information that was pertinent in identifying distinct groups of wildlife watchers. It was seen from earlier that 23 percent of component 1 in the MI model consisted of respondents who had taken one or more trips. The poorer performance of the zero-truncated FM model seems to provide further evidence that there is no clear divide between participants and nonparticipants. Given the relatively low estimate of  $\varepsilon$ , we did not pursue a closer examination of the component memberships. Some of the coefficients such as for *age*, *agesq*, *male*,  $\ln[\text{hincome}]$ , and *wildcontr* were similar to those from the MI models, but estimates for *school*, *trip05*, and *hunto6* showed different signs. Estimates for other variables showed the same sign but were different in magnitudes. Most notably, the travel cost coefficients were on average larger in magnitude in the zero-truncated models. As shown in table 2.8, this resulted in smaller per-trip expected consumer surplus estimates compared to the MI models. The patterns are very similar to those of the MI models. It can be seen that the largest difference in values from using  $c^1 = \infty$  versus  $c^1 < \infty$  arose from

only one of the components. Therefore a similar situation to that depicted in figure 2.2 may be present here as well.

#### *2.4.7. Comparison of Goodness of Fit*

To more formally compare the performances of the MI and zero-truncated FM models, several measures of goodness of fit were obtained. In order to simplify the analysis, a weighted average of the predicted trip values for each observation using the component level probabilities was calculated. Also, the observations were restricted to the truncated sample ( $n = 3771$ ) for comparability. Table 2.9 reports the summary statistics of the observed trip counts and predicted values from each of the models. From this table, it is clear that the zero-truncated model failed to predict a significant number of large values as the maximum was 55.23. Table 2.10 reports five  $R^2$  measures to more formally compare the performance of each model. A detailed discussion of these measures can be found in Cameron and Windmeijer (1996). Table 2.10 clearly shows the MI model outperformed the zero-truncated model. While the negative value of the deviance residual based  $R^2$  measure is unusual, the deviance residual value from which this  $R^2$  measure was based is smaller for the MI model (15239.91 versus 29138.36) suggesting greater performance. Figure 2.3 illustrates the differences in performances further using graphs of the cumulative distribution functions of observed counts and the model predictions. Notice how the MI Trips line (solid line) is closer to the Observed Trips line (dotted line) nearly throughout the entire range of values. Notice also the Truncated Trips line (dashed line) ceases at 55.23 trips as previously shown in the summary statistics.

## **2.5. Discussion and Conclusion**

This study applied a finite mixture model to wildlife watching recreation data from a national survey. The predicted mean matching multiple imputation method was used to fill in missing travel cost values. A unique feature of this approach is that contingent valuation method (CVM) responses were included to help obtain more accurate imputations for current nonparticipants, as the questions elicited respondents' choke prices.

The results provided strong evidence in support of heterogeneous preferences in the wildlife watching population, which was argued to have important implications for sustainable tourism and wildlife management. The analysis identified three subgroups of wildlife watchers roughly characterized as: local enthusiasts, traveling enthusiasts, and average participants. The groups exhibited very different visitation and spending patterns, as well as different benefits from access to wildlife watching. An important policy implication from the results for wildlife and recreation site managers is that excess visitation can be curbed by spreading out the demand by using a mix of prices and visiting hours since some visitor were found to be traveling while others local. Higher prices during the weekends and holidays may encourage local visitors to take more trips during times with less crowding. For visitors that are less sensitive to prices, restricting visitation during important periods during the year such as breeding season may be required to help sustain a healthy stock of wildlife. The post estimation analysis revealed further segmentation was possible differentiating skill and experience levels. There were also indications that

wildlife watchers may consciously reside closer to their frequented sites in order to reduce travel costs, and their economic activities may differ when participating locally compared to when they travel. The use of choke price data in the calculation of consumer surplus revealed some interesting discrepancies that may be an avenue for further study. However, an important consideration in regards to the results is the difficulty in obtaining credible open-ended CVM responses. Therefore, an extension to these findings might study the mechanisms people use in deriving their choke prices more closely. Finally, future studies could investigate the performance of models when the multiple imputation (MI) method is used to impute other types of CVM data, as the goodness of fit of a model using MI shown here was substantially higher than that of the traditional zero-truncated model.

Table 2.1: Summary Statistics of variables and descriptions

Variable	Mean	Std. Dev.	Min	Max	Description
<u>FM model variables used in PMM</u>					
<i>trips</i>	2.46	12.44	0.0	365.00	Total number of trips taken in 2006 by respondent
<i>travel cost (n = 3771)</i>	106.91	235.10	0.71	3248.46	Defined as: (transportation & access costs) + opportunity cost
<i>opportunity cost</i>	17.27	42.35	0.40	1473.15	Defined as: 0.04earnings (days/trips) <sup>a</sup>
<i>age</i>	50.12	16.39	16.00	90.00	Age of respondent
<i>agesq</i>	2780.99	1687.98	256	8100	Age of respondent variable squared
<i>male (=1)</i>	0.47	0.50	0	1	Respondent is male
<i>married (=1)</i>	0.69	0.46	0	1	Respondent is married
$\ln[\textit{hincome}]$	10.91	0.74	8.36	12.19	Natural logarithm of respondent's household income <sup>b</sup>
<i>hincome</i>	69978.15	46989.27	4278.10	196934.80	Household income

<sup>a</sup>To calculate *opportunity cost* of time, we set the total number of potential work days at 250 and used the household income variable to obtain daily earnings. Daily earnings are useful here because the time cost, which is measured by the amount of time spent wildlife watching, is recorded in days. We divided the commonly used wage factor (.333) by 8 to arrive at the earnings factor 0.04.

<sup>b</sup>The household income *hincome* is a predicted value obtained using interval regression. Model results are available upon request.

Table 2.1: Summary Statistics of variables and descriptions continued

Variable	Mean	Std. Dev.	Min	Max	Description
<u>FM model variables used in PMM</u>					
<i>school</i>	13.92	2.70	0	18	Total number of years of schooling obtained by respondent
<i>photo</i> (=1)	0.24	0.43	0	1	Respondent photographed wildlife around their home
<i>feedbird</i> (=1)	0.59	0.49	0	1	Respondent fed wildlife around their home
<i>parks</i> (=1)	0.17	0.37	0	1	Respondent visited a public park within a mile of their home
<i>wildcontr</i> (=1)	0.13	0.34	0	1	Respondent paid dues to a wildlife watching organization
<i>trip05</i> (=1)	0.30	0.46	0	1	Respondent took wildlife watching trip(s) in 2005
<i>hunt06</i> (=1)	0.05	0.21	0	1	Respondent took hunting trip(s) in 2006
<i>fish06</i> (=1)	0.15	0.35	0	1	Respondent took fishing trip(s) in 2006

Table 2.1 Summary Statistics of variables and descriptions continued

Variable	Mean	Std. Dev.	Min	Max	Description
<u>Other variables used in PMM</u>					
<i>white</i> (=1)	0.92	0.27	0	1	Respondent's race is White
<i>urban</i> (=1)	0.59	0.49	0	1	Respondent resided in a urban area
<i>feedfish</i> (=1)	0.21	0.41	0	1	Respondent fed fish around their home
<i>wildbooks</i> (=1)	0.13	0.34	0	1	Respondent purchased books on wildlife watching
<i>private</i> (=1)	0.14	0.35	0	1	Respondent took wildlife watching trips to privately owned land
<i>wildlife</i> (=1)	0.54	0.50	0	1	Respondent observed wildlife around their home in 2006
<i>observe05</i> (=1)	0.62	0.48	0	1	Respondent observed wildlife in 2005
<i>feed05</i> (=1)	0.56	0.50	0	1	Respondent fed wildlife in 2005
<i>photo05</i> (=1)	0.25	0.44	0	1	Respondent photographed wildlife in 2005
<i>planto5</i> (=1)	0.31	0.46	0	1	Respondent planted vegetation for wildlife in 2005
<i>occupation</i> (=1,2,...)					Respondent's occupation (categorical)
<i>census division</i> (1,2,...)					Census division designation of respondent residence (categorical)
<i>trip_spend05</i> (1,2,...)					Trip spending in 2005 (categorical)



Table 2.1 Summary Statistics of variables and descriptions continued

Variable	Mean	Std. Dev.	Min	Max	Description	
<u>Other variables not used in PMM but used elsewhere</u>						
<i>days</i>	3.52	15.84	0.0	365	Total number of trip days in 2006	
<i>days/trips (n = 3771)</i>	2.60	6.85	0.07	200	Total number of days per trip in 2006	
<i>non-resident (=1) (n = 3771)</i>	0.39	0.49	0	1	Trip was to a location outside of respondent's resident state	
<i>forest &amp; rangeland proportion</i>	0.44	0.24	0.0	2.03	'roportion of land acres in state designated as forest or rangeland	
<i>forest &amp; rangeland per capita</i>	10.11	48.69	0.0	452.87	Per capita forest and rangeland acres in state	
<hr/>						
<i>occupation</i>	Freq.	Percent	<i>census division</i>	Freq.	Percent	
Retired	2,641	21.61	New England	1,918	15.69	
Going to School	384	3.14	Middle Atlantic	1,001	8.19	
Keeping House	841	6.88	East North Central	1,243	10.17	
Something else	505	4.13	West North Central	1,369	11.20	
Job/Business	7,853	64.24	South Atlantic	2,101	17.19	
Total	12,224	100	East South Central	739	6.05	
			West South Central	632	5.17	
			Mountain	1,643	13.44	
			Pacific	1,578	12.91	
			Total	12,224	100	
				<i>trip_spend05</i>	Freq.	Percent
				\$25 or less	9,770	79.92
				\$26 to \$99	653	5.34
				\$100 to \$199	478	3.91
				\$200 to \$299	246	2.01
				\$300 or more	1,077	8.81
				Total	12,224	100

Table 2.2: PMM results for travel cost

	Mean	95% Conf Interval	Average RVI	Largest FMI
<u>All: PMM imputed and observed values<sup>a</sup></u>				
<i>n</i> = 12224	169.44	[156.23, 182.64]	0.428	0.313
<u>PMM imputed values</u>				
<i>trips</i> = 0				
<i>n</i> = 8453	189.67	[172.41, 206.94]	0.461	0.331
<u>Observed values</u>				
<i>trips</i> = 1				
<i>n</i> = 1385	211.75	[183.98, 239.52]	0.0	0.0
<i>trips</i> = 2				
<i>n</i> = 546	89.42	[69.53, 109.30]	0.0	0.0
<i>trips</i> = 3				
<i>n</i> = 395	54.87	[45.30, 64.45]	0.0	0.0
<i>trips</i> = 4+				
<i>n</i> = 1445	36.34	[33.08, 39.59]	0.0	0.0
<i>trips</i> > 0				
<i>n</i> = 3771	110.05	[98.88, 121.22]	0.0	0.0
<u>CVM Response</u>				
<i>n</i> = 3091	408.96	[372.47, 445.44]	0.0	0.0

Note: Estimated means were obtained using the wildlife watching sample survey weight provided in the data. RVI is the relative increase in the variance due to nonresponse. FMI is the fraction of missing information due to nonresponse.

<sup>a</sup>Imputed values and observed values are in reference to the *travel cost* variable defined in table 2.1.

Table 2.3: Average information criterion and entropy values

Components	$n$	$\ln L$	$p$	AIC	BIC	$\varepsilon$
1	12224	13988.27	16	28008.54	28127.12	—
2	12224	13117.74	33	26301.49	26546.06	0.483
3	12224	13001.02	50	26102.05	26472.60	0.596
4	12224	12921.51	67	25977.01	26473.56	0.441

Note: Values are averages taken over MI estimates.  $p$  is the number of parameters.

Table 2.4: MI parameter estimates of single and 3-component FM models

Variable	Single-Component		3-Component	
	—	Comp-1	Comp-2	Comp-3
<i>age</i>	0.0304 (0.0203)	0.0745** (0.0298)	-0.0291 (0.0433)	0.1063*** (0.0299)
<i>agesq</i>	-0.00032 (0.00021)	-0.00081*** (0.00027)	0.00035 (0.00042)	-0.00113*** (0.00032)
<i>male</i> (=1)	0.4103*** (0.1245)	-0.0164 (0.1245)	0.7202*** (0.2387)	0.3817 (0.2338)
<i>married</i> (=1)	-0.2544 (0.1399)	-0.1716 (0.1419)	-0.0495 (0.4454)	-0.5101** (0.2270)
$\ln[\textit{hincome}]$	0.0654 (0.0918)	0.2065 (0.1160)	0.02060 (0.1851)	0.6876* (0.3962)
<i>school</i>	0.0556** (0.0221)	0.0714** (0.0285)	0.0161 (0.0446)	0.0579 (0.0367)
<i>travel cost</i> (MI)	-0.0029*** (0.00037)	-0.00057*** (0.00017)	-0.0097*** (0.0021)	-0.0351*** (0.0121)
<i>photo</i> (=1)	0.8917*** (0.1307)	0.8349*** (0.1415)	0.5403** (0.2575)	0.7643 (0.312)
<i>feedbird</i> (=1)	0.8151*** (0.1203)	0.4978*** (0.1337)	0.6859** (0.3434)	0.6857*** (0.2507)
<i>parks</i> (=1)	1.071*** (0.1240)	1.203*** (0.1290)	0.5205 (0.4556)	0.9192*** (0.2799)
<i>wildcontr</i> (=1)	0.4063*** (0.1159)	0.6971*** (0.1427)	0.4557 (0.3007)	0.4947** (0.2111)
<i>trip05</i> (=1)	1.023*** (.0986)	1.6887*** (0.1320)	0.3540 (0.2461)	0.8345*** (0.1670)
<i>hunt06</i> (=1)	0.9739*** (0.2456)	0.5749** (0.2776)	-0.0495 (1.1906)	1.036** (0.4728)
<i>fish06</i> (=1)	0.8632*** (0.2438)	0.6951*** (0.1718)	1.0225 (0.6319)	0.6020** (0.2743)
<i>constant</i>	-3.076*** (1.069)	-7.578*** (1.207)	1.866 (2.453)	-9.172*** (4.457)
$\alpha_g^a$	6.643*** (0.3458)	1.452*** (0.2369)	1.952** (0.5397)	1.244 (0.3234)
$\pi_g = \text{prob}[i \in g]^b$ <i>entrophy</i>	1.0 —	0.785	0.034 0.596	0.181

Table 2.4: MI parameter estimates of single and 3-component FM models continued

Variable	Single-Component	3-Component		
	–	Comp-1	Comp-2	Comp-3
Average RVI	0.0148	0.0149	0.0034	0.0163
Overall Average RVI	0.0148		0.0115	
Largest FMI	0.0893	0.0767	0.0079	0.0596
Overall Largest FMI	0.0893		0.0767	
$E[\text{trips}   x'\beta]_g^c$	2.69	0.58	16.07	5.85
$\sum \pi_g E[\text{trips}   x'\beta]_g^c$	2.69		2.05	
sample mean <sup>c</sup>			2.0	
$n$			12224	

Robust standard errors in parentheses

\* significant at 10% level, \*\* significant at 5% level, \*\*\* significant at 1% level

<sup>a</sup>Significance are based on the t-statistics of the estimates of  $\ln[\alpha]_g$  which were obtained using Rubin's (1987) rules. The standard errors were obtained using the Delta method.

<sup>b</sup>Significance are based on the t-statistics of the estimates of the auxiliary parameter  $\gamma_g$ , where

$\pi_g = \exp(\gamma_g)/(1 + \sum_g \exp(\gamma_g))$ , which were obtained using Rubin's (1987) rules.

Estimates of  $\gamma_g$  for components 1 and 2 were significant at  $\leq 5\%$  level.

<sup>c</sup>Parameter and mean estimates were obtained using the available survey weights.

Note: RVI is the relative increase in variance due to nonresponse, and FMI is the fraction of missing information due to nonresponse.

Table 2.5: Component-level mean values

Variable	Multiple Imputation 3-Component Finite Mixture Model						
	<i>trip</i> ≥ 0	Comp-1			<i>trip</i> = 0	Comp-2	Comp-3
		<i>trip</i> > 0	<i>trip</i> > 0	<i>trip</i> > 0		<i>trip</i> > 0	<i>trip</i> > 0
		<i>all</i>	<i>out of state</i>	<i>in state</i>			
<i>travel cost</i>	177.81	143.04	222.02	66.17	188.45	60.74	23.00
<i>age</i>	50.39	48.82	50.06	47.62	50.87	49.80	47.34
<i>male</i> (=1)	0.46	0.47	0.45	0.49	0.46	0.48	0.52
<i>married</i> (=1)	0.69	0.74	0.75	0.73	0.68	0.66	0.67
ln[ <i>hincome</i> ]	10.92	11.14	11.20	11.08	10.85	10.68	10.90
<i>school</i>	13.93	15.02	15.19	14.86	13.59	13.30	13.90
<i>retired</i> (=1)	0.22	0.17	0.19	0.14	0.24	0.23	0.16
<i>photo</i> (=1)	0.23	0.49	0.47	0.50	0.16	0.26	0.29
<i>feedbird</i> (=1)	0.58	0.72	0.68	0.75	0.54	0.54	0.65
<i>parks</i> (=1)	0.16	0.42	0.41	0.44	0.08	0.12	0.21
<i>wildcontr</i> (=1)	0.13	0.31	0.36	0.27	0.08	0.11	0.15
<i>trip05</i> (=1)	0.29	0.68	0.67	0.69	0.17	0.32	0.39
<i>hunto6</i> (=1)	0.04	0.07	0.05	0.09	0.03	0.05	0.08
<i>fisho6</i> (=1)	0.14	0.23	0.19	0.26	0.12	0.14	0.19
<i>days</i>	1.44	6.13	5.20	7.02	0	63.88	17.08
<i>days/trips</i>	—	3.17	3.79	2.57	0	2.47	1.19
<i>non-resident</i> (=1)	—	0.49	1	0	—	0.20	0.14
<i>forrest &amp; rangeland proportion</i>	9.78	13.98	18.58	9.50	8.49	22.66	11.87
<i>forrest &amp; rangeland per capita</i>	0.44	0.45	0.48	0.43	0.43	0.49	0.45
<i>n</i>	11042	2589	1277	1312	8453	149	1033

Table 2.6: Expected Consumer surplus estimates from MI models

	Single-Component	3-Component		
	—	Comp-1	Comp-2	Comp-3
$E[CS   c^1 = \infty]_g$	[342.38] 919.97 (111.13)	[1756.85] 1018.36 (310.27)	[103.22] 1658.84 (362.21)	[28.50] 166.65 (57.25)
95% Conf. Interval	{702.15,1137.78}	{410.23,1626.49}	{948.91,2368.77}	{54.44,278.86}
$\sum \pi_g E[CS   c^1 = \infty]_g$	[342.38] 919.97 (111.13)		[1388.60] 886.06 (283.97)	
95% Conf. Interval	{702.15,1137.78}		{329.48,1442.64}	
$c^1 = 408.96^a$				
$c^o = 90.89^a$				
$E[CS   c^1 < \infty]_g$	[207.16] 556.62 (195.79)	[290.94] 168.64 (616.02)	[98.49] 1582.70 (430.06)	[27.19] 159.00 (57.26)
95% Conf. Interval	{172.87,940.37}	{-1038.80,1376.04}	{739.78,2425.62}	{46.77,271.23}
$\sum \pi_g E[CS   c^1 < \infty]_g$	[207.16] 556.62 (195.79)		[236.77] 214.67 (552.19)	
95% Conf. Interval	{172.87,940.37}		{-867.62,1296.96}	

Note:  $E[CS | c^1 = \infty]_g$  is the expected consumer surplus with upper limit set to infinity,  $E[CS | c^1 < \infty]_g$  is the consumer surplus with upper limit set to the mean of CVM response (408.96) and lower limit set to the sample mean (90.89), and  $\pi_g$  is the component-level estimated probability of membership.

Per-trips expected consumer surplus estimates in brackets and standard errors using the Delta Method in parentheses.

<sup>a</sup> The estimated mean of CVM response and travel cost were obtained using the available survey weights.

Table 2.7: Parameter estimates of single and 3-component zero-truncated FM model

Variable	Single-Component		3-Component	
	—	Comp-1	Comp-2	Comp-3
<i>age</i>	0.0334 (0.0227)	0.0823 (0.0571)	0.0342 (0.0257)	0.0627** (0.0268)
<i>agesq</i>	-0.00027 (0.00023)	-0.00084 (0.00056)	-0.00029 (0.00025)	-0.00068** (0.00029)
<i>male</i> (=1)	0.5263*** (0.1413)	0.7905*** (0.2680)	0.4734*** (0.1712)	0.3374** (0.1401)
<i>married</i> (=1)	-0.4408** (0.1733)	-0.9980*** (0.3138)	-0.5059** (0.1978)	-0.0298 (0.1877)
$\ln[\textit{hincome}]$	0.0232 (0.1091)	0.6657*** (0.2441)	0.1494 (0.1252)	0.4924*** (0.1591)
<i>school</i>	-0.0404 (0.0304)	-0.1205** (0.0532)	-0.0248 (0.0351)	-0.0683* (0.0365)
<i>travel cost</i>	-0.00867*** (0.00092)	-0.0037*** (0.00080)	-0.0285*** (0.0023)	-0.0294*** (0.0042)
<i>photo</i> (=1)	0.3687** (0.1439)	0.5341* (0.2858)	0.4687*** (0.1720)	0.1537 (0.1427)
<i>feedbird</i> (=1)	0.3759*** (0.1378)	0.5125* (0.2974)	0.1572 (0.1585)	0.2718 (0.1730)
<i>parks</i> (=1)	0.4542*** (0.1385)	-0.4136 (0.2938)	0.5780*** (0.1693)	0.1792 (0.1591)
<i>wildcontr</i> (=1)	0.1860 (0.1565)	0.4543 (0.3024)	0.4091** (0.2070)	0.2020 (0.1701)
<i>trip05</i> (=1)	-0.0345 (0.1291)	-0.00079 (0.2963)	-0.0018 (0.1910)	0.1424 (0.1485)
<i>hunt06</i> (=1)	0.6595*** (0.2433)	1.334** (0.5551)	0.3306 (0.3300)	0.8194*** (0.2486)
<i>fish06</i> (=1)	0.3472* (0.1943)	0.0969 (0.3340)	0.4136* (0.2405)	0.2422 (0.1618)
<i>constant</i>	-8.884 (7.663)	-15.386*** (2.189)	0.5146 (1.269)	-4.475*** (1.703)
$\alpha_g^a$	65512.746 (490690.468)	10055.087*** (11853.058)	4.118*** (2.044)	1.867*** (0.4053)
$\pi_g = \text{prob}[i \in g   \textit{trips} > 0]^b$	1.0	0.184	0.209	0.607
<i>entropy</i>	—		0.220	



Table 2.7: Parameter estimates of single and 3-component zero-truncated FM model continued

Variable	Single-Component		3-Component	
	—	Comp-1	Comp-2	Comp-3
$E[\text{trips}   \text{trips} > 0]_g^c$	7.28	3.00	18.41	4.39
$\sum \pi_g E[\text{trips}   \text{trips} > 0]_g^c$	7.28		7.06	
$\ln(L)$	-7590.718		-7333.147	
sample mean <sup>c</sup>		7.88		
$n$		3771		

Robust standard errors in parentheses

\* significant at 10% level, \*\* significant at 5% level, \*\*\* significant at 1% level

<sup>a</sup>Significance are based on the t-statistics of the estimates of  $\ln[\alpha]_g$ . The standard errors were obtained using the Delta method.

<sup>b</sup>Significance are based on the t-statistics of the estimates of the auxiliary parameter  $\gamma_g$ , where  $\pi_g = \exp(\gamma_g) / (1 + \sum_g \exp(\gamma_g))$ .

Estimates of  $\gamma_g$  for components 1 and 2 were significant at  $\leq 5\%$  level.

<sup>c</sup>Parameter and mean estimates were obtained using the available survey weights.

Table 2.8: Expected Consumer surplus estimates from zero-truncated models

	Single-Component	3-Component		
	—	Comp-1	Comp-2	Comp-3
$E[CS   c^1 = \infty]_g$	[115.34] 839.68 (89.10)	[271.85] 815.55 (176.57)	[35.12] 646.57 (52.66)	[33.96] 149.10 (21.45)
95% Conf. Interval	{665.04,1014.32}	{469.47,1161.63}	{543.36,749.78}	{17.06,191.42}
$\sum \pi_g E[CS   c^1 = \infty]_g$	[115.34] 839.68 (89.10)		[77.98] 375.70 (81.21)	
95% Conf. Interval	{665.04,1014.32}		{216.53,534.87}	
$c^1 = 408.96^a$				
$c^0 = 90.89^a$				
$E[CS   c^1 < \infty]_g$	[108.03] 786.41 (110.34)	[187.48] 562.43 (340.76)	[35.12] 646.50 (1059.69)	[33.96] 149.08 (500.88)
95% Conf. Interval	{570.14,1002.68}	{105.46,1230.32}	{-1430.50,2723.49}	{-832.64,1130.80}
$\sum \pi_g E[CS   c^1 < \infty]_g$	[108.03] 786.41 (110.34)		[62.45] 329.10 (639.02)	
95% Conf. Interval	{570.14,1002.68}		{-923.38,1581.58}	

Note:  $E[CS | c^1 = \infty]_g$  is the expected consumer surplus with upper limit set to infinity,  $E[CS | c^1 < \infty]_g$  is the consumer surplus with upper limit set to the mean of CVM response (408.96) and lower limit set to the sample mean (90.89), and  $\pi_g$  is the component-level estimated probability of membership.

Per-trips expected consumer surplus estimates in brackets and standard errors using the Delta Method in parentheses.

<sup>a</sup> The estimated mean of CVM response and travel cost were obtained using the available survey weights.

Table 2.9: Summary statistics of observed trips counts and predicted values

	Mean	Std. Dev.	Min	Max
Observed	7.98	21.39	1	365
MI	7.58	11.91	0.013	219.77
Zero-truncated	7.59	5.86	0.028	55.23
$n$		3771		

Predicted values for MI and zero-truncated models were calculated using the formula  $\xi_i = \sum \pi_{gi} \mu_{gi}$ , where  $\mu_{gi}$  is the predicted trip value for the  $i^{\text{th}}$  observation in component  $g$ .

Table 2.10:  $R^2$  measures for MI and zero-truncated FM models

$R^2$ measure	Model	
	MI	Truncated
Residual Sum of Squares	0.378	0.072
Explained Sum of Squares	0.310	0.075
Squared Sample Correlation	0.382	0.073
Pearson Residuals	0.756	0.459
Deviance Residual [Deviance Residual Value]	0.698 [15239.91]	-0.500 [29138.36]

Note: MI values reflect the truncated sample, i.e.,  $trips > 0$ , for comparability ( $n = 3771$ ).

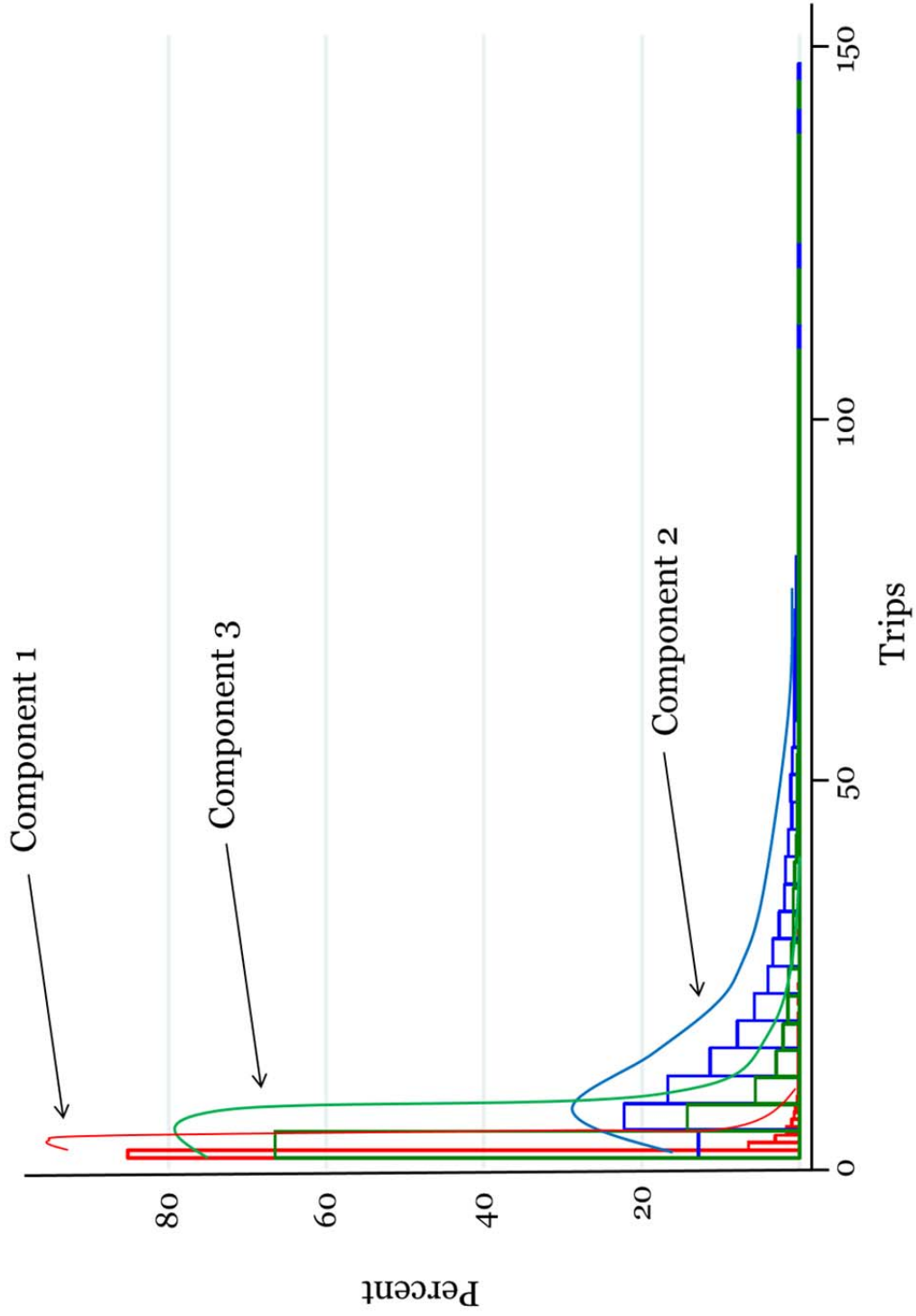


Figure 2.1: Distributions of expected trip counts by components

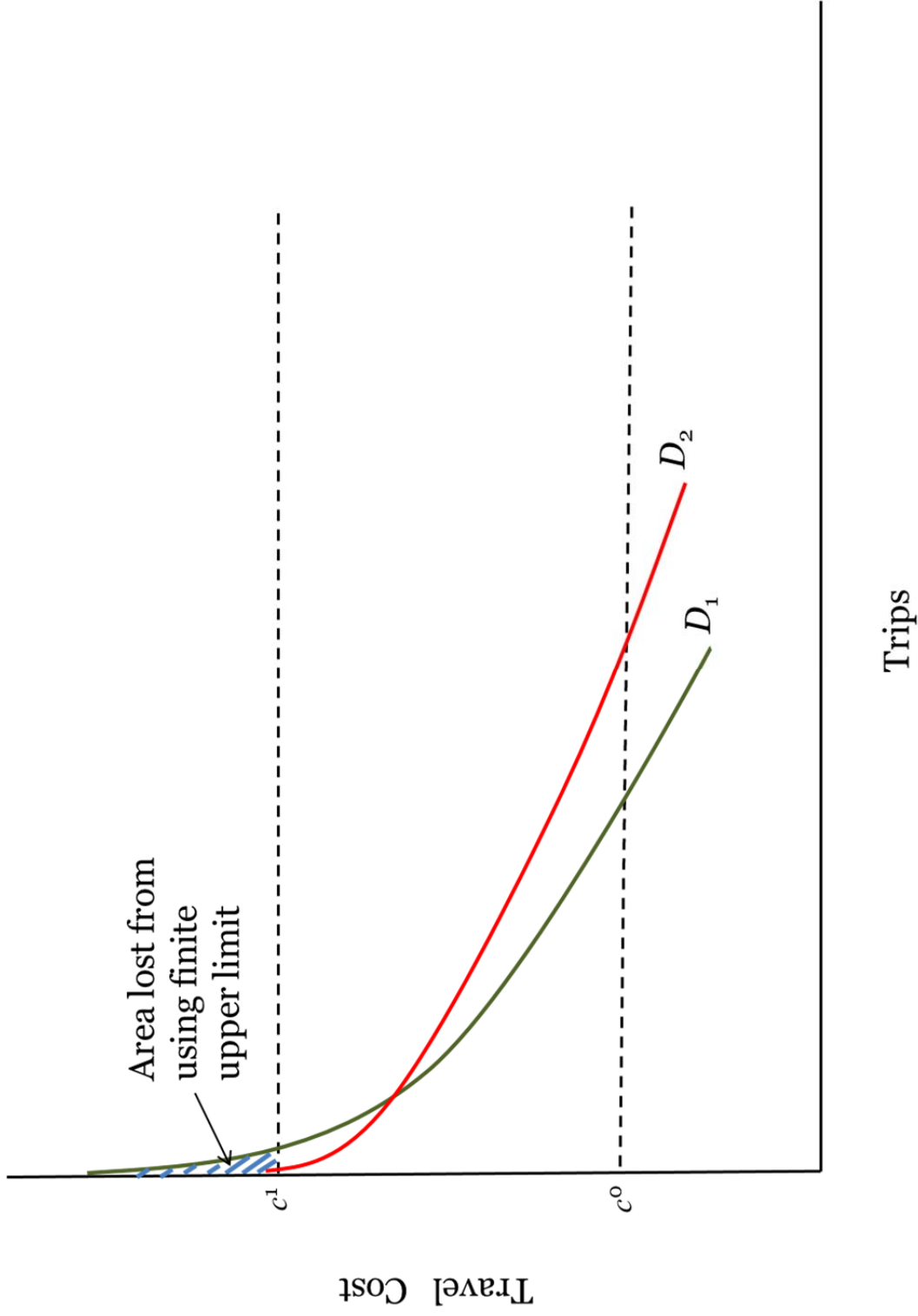


Figure 2.2: Consumer surplus with finite upper limit

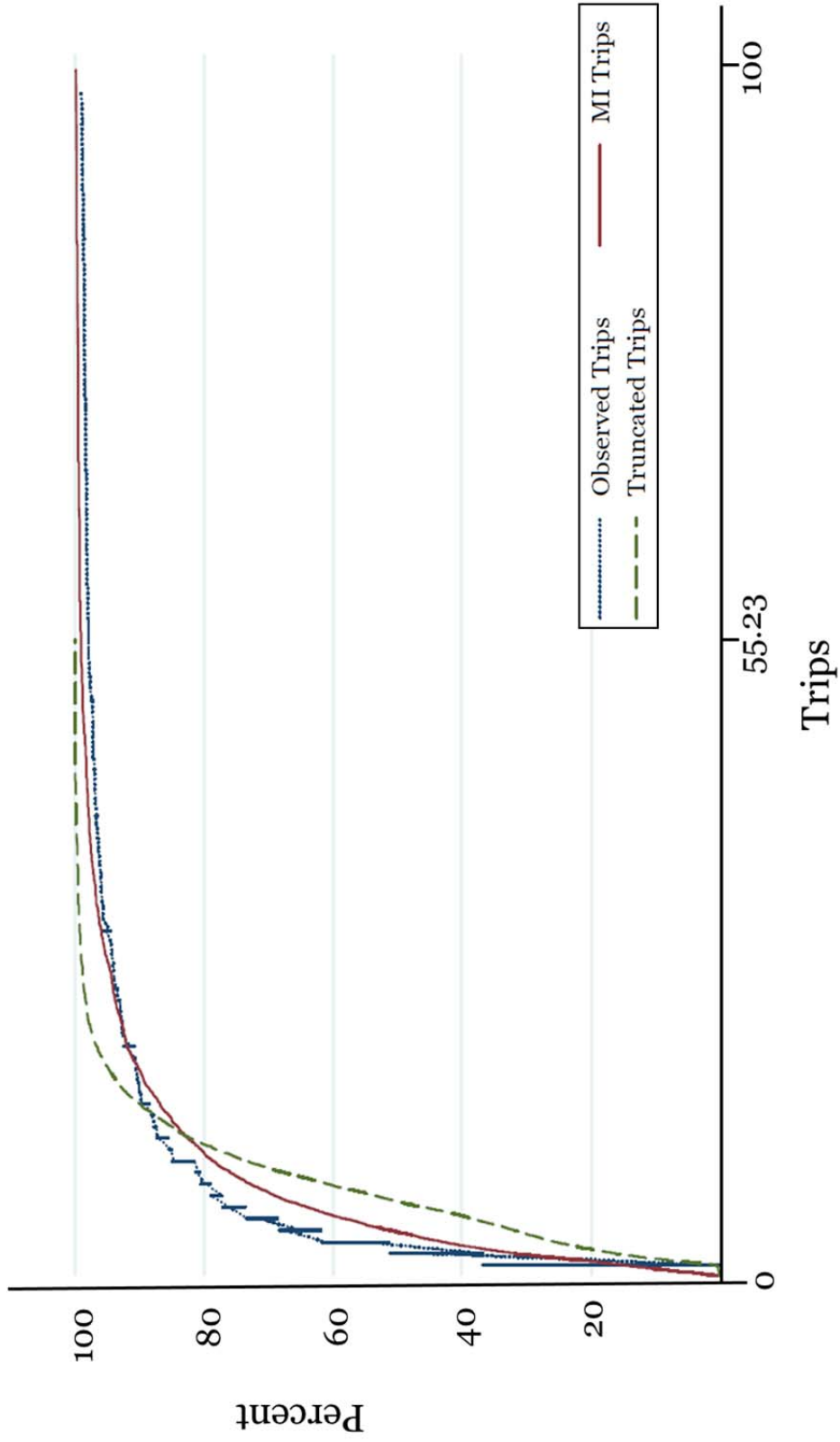


Figure 2.3: Cumulative Densities of Observed and Predicted MI and Truncated Trips

## **CHAPTER 3: Utility-consistent Joint Estimation of the Value of Wildlife Watching Recreation**

### **3.1. Introduction**

Combining revealed preference (RP) data with stated preference (SP) data has been shown to have many advantages and applications in nonmarket valuation (Cameron, 1992; Dhazn, Woodward, Ozuna Jr., & Griffin, 2003; Eom & Larson, 2006; González-Sepúlveda, 2008; González-Sepúlveda, Loomis, & González-Cabán, 2008; Whitehead, Dumas, Hertine, Hill, & Buerger, 2008). RP data provides information on actual behavior which can be used to reveal an individual's preference for a nonmarket good given the current circumstances (Herriges & Kling, 1999). SP data provides information on hypothetical behavior which can also be used to reveal preferences, but within a hypothetical context where the circumstances are changed (Mitchell & Carson, 1989). Combining the two types has advantages such as increased statistical efficiency (González-Sepúlveda, Loomis, & González-Cabán, 2008) and an ability to attenuate some problems that are particular to each data type (Cameron, 1992; Azevedo, Herriges, & Kling, 2003).

However, when joining RP information such as travel cost method (TCM) data with SP information such as contingent valuation method (CVM) data, different approaches may yield different degrees of data integreibility and cohesion of information. Furthermore, the modeling frameworks that arise from TCM and CVM may not always reflect the same preference structure nor are they



guaranteed to describe the same scenario. This paper presents a joint model that combines TCM count data with CVM dichotomous choice data in a way so that the preference structure reflected in the two data types as well as the scenarios being analyzed and welfare measures are consistent.

The integration of TCM and CVM data into a joint model can be accomplished in several ways (For example, Loomis, 1997; González-Sepúlveda, Loomis, & González-Cabán, 2008). The framework used in this paper follows Cameron (1992) who proposed a modeling approach that constrains the utility function reflected in the TCM data to be the same as the utility function reflected in the CVM data. Cameron (1992) argued the decisions of a utility maximizing economic agent, be they actual or hypothetical, should reflect the same underlying preference structure. Moreover, by requiring the hypothetical behaviors from CVM to be consistent with actual TCM behaviors, the problem of hypothetical bias attributed to CVM questions may be attenuated. In a similar manner, the CVM data contributes information about demand that is difficult to capture from using TCM data alone such as nonuse values. Cameron's (1992) *utility-consistent* joint model combines the TCM and CVM modeling frameworks so that a single set of estimates incorporating both types of information are obtained. However, success in integrating the models depends on the cohesiveness of the two types of information.

Combining TCM and CVM data can be more fruitful if they can jointly be used to investigate a common scenario. A scenario described in the CVM survey to elicit compensating/equivalent variation (*CV*) (or willingness to pay) should align with the scenario reflected by a (change in) consumer surplus (*CS*) (or net

benefits) from TCM. However, scenarios that can easily be depicted in CVM surveys may not be as easy to analyze using a TCM model. Compared to CVM, a TCM analysis is less direct since it involves first estimating a demand function and then calculating the change in consumer surplus. Even with a common scenario however, the welfare measures of TCM and CVM (*CV* and *CS*) will generally differ due to their theoretical underpinnings. In this paper, we present a utility-consistent joint model where the TCM and CVM welfare measures are equal. The advantage of having this equivalence is that welfare effects can be interpreted from either a Hicksian and Marshallian perspective—that is, a single estimate can be interpreted in terms of willingness to pay or net benefits.

The remainder of the paper is as follows. The next section reviews the existing literature on combining RP and SP data and the utility-consistent joint estimation approach. Section 3.3 presents the methodology used to estimate the utility-consistent model, and the formulas underlying the econometric estimation and calculation of the welfare measures. The data used in the analysis is discussed in section 3.4 and the results are reported in section 3.5. Section 3.6 provides the discussion and conclusion, and avenues for future research to end the paper.

## **3.2. Literature Review**

### *3.2.1. Combining Revealed and Stated Preference Data*

Nonmarket valuation data has taken two major forms over the past several decades; revealed preference (RP) and stated preference (SP). RP data consists of information on observed behavior, such as the number of recreation trips

taken over the course of a year and the amount of money spent on these trips. SP data consists of responses or statements concerning hypothetical behavior. Here, the respondent may be asked to consider a hypothetical rise in the cost of access to a recreational facility or the quality of the facility. They would then be asked whether they would change their behavior in response to the scenario.

RP data are analyzed using the Travel Cost Method (TCM) or Hedonic Price Models (HPM). TCM is used to estimate demand functions for recreation trips, where the cost of travel acts as the price. The coefficient estimate for the travel cost can be used to derive an estimate of the consumer surplus. In HPM, benefits from nonmarket goods are estimated by observing its effect on the prices of market goods such as residential real estate. For SP data, the Contingent Valuation Method (CVM) is the most common means of collecting information while other methods such as the choice experiment method is popular as well. With CVM, individuals may be asked to provide a *yes* or *no* response to a proposed hypothetical scenario whereby the data are analyzed using discrete choice models.

Criticisms surrounding the use of CVM data concern the hypothetical nature of the responses and its credibility as a reliable source of information on preferences (Bishop & Heberlein, 1979, 1986), although there is little consensus on the matter as subsequent validation studies have refuted these aspects as a problem (Loomis, 1989; Carson, Flores, Martin, & Wright, 1996). The use of TCM data has been criticized for its inability to capture passive or nonuse values and sensitivity to how the travel cost variable is defined. By combining the data types, it is hoped that they can complement each other and overcome some of the

weaknesses that have been found when they are used individually (Azevedo, Herriges, & Kling, 2003; Whitehead, Pattanayak, Houtven, & Gelso, 2008).

Loomis (1997) merged TCM and CVM data within the framework of the dichotomous choice format initiated by Hanemann (1984) and Bishop and Heberlein (1979). Loomis (1997) used the notion that positive trip records reflect an implicit *yes* response with respect to current costs. These were matched with responses to CVM questions which served as an additional choice observation with a relatively higher hypothetical cost.

González-Sepúlveda, Loomis, and González-Cabán (2008) combined CVM dichotomous choice data with TCM count data in a joint model. Their CVM question asked respondents if they would have taken their last trip if the cost was increased by a certain amount chosen by the survey. Recreation trips in the TCM component were modeled as count data and adjusted for truncation and endogenous stratification. The CVM dichotomous choice responses were modeled as a Probit distributed discrete variable. Their modeling approach allowed the parameters of the TCM model to be independent of the parameters of the CVM model.

### *3.2.2. Utility-consistent Joint Estimation*

Cameron (1992) developed a way of combining stated and revealed preference data into a model where a single set of parameters are shared by the TCM and CVM models. In a study about recreational fishermen, the CVM questions asked whether the respondent would entirely cease taking salt water fishing trips given a increase in total cost. A quadratic functional form was adopted for the utility

function which was then used to derive a corresponding demand function. In the econometric model, recreation trips in the TCM component were modeled as a continuous and normally distributed variable while the CVM dichotomous choice responses were modeled as a Probit discrete variable. In order to impose the requirement that the CVM and TCM decisions reflect the same utility function, the models were estimated simultaneously in a joint maximum likelihood function. Compensating variation and equivalent variation estimates were obtained using Hanemann's (1984) utility difference approach.

Dhazn, Woodward, Ozuna Jr., and Griffin (2003) extended Cameron's (1992) utility-consistent framework of combining CVM dichotomous choice data with TCM data by modeling the TCM data as a truncated normal distribution. The CVM portion of the survey presented respondents with a randomly chosen price for a program to maintain the current red snapper catch rates, and then were asked if they would be willing to pay that price or consider the alternative of an eventual elimination of the red snapper population. Similar to Cameron (1992), the authors used the utility-difference approach outlined in Hannemann (1984) to derive a compensating variation measure to estimate respondents' willingness to pay for maintaining red snapper catch rates. One caveat acknowledged by the authors however is that the responses to their CVM questions reflect discrete changes in fishing stock while the trip responses may be affected by marginal changes in fishery. Although the data may capture essentially the same preferences, the authors still found the TCM welfare estimates to be different from the CVM estimates.

Eom and Larson (2006) also combined data from TCM and CVM data in a utility-consistent unified model. Their goal however was to use this framework to develop a way to more easily obtain estimates of use, nonuse, and total values of environmental quality changes. In their approach they were also able to test whether preferences satisfied weak complementarity. The utility-consistent modeling approach was slightly different from Cameron (1992) in that the authors began with the demand function which took on a semi-log functional form and then derived a quasi-expenditure function to finally arrive at an indirect utility function. The advantage of using Eom and Larson's (2006) approach is that a semi-log functional form of the demand function is used which conforms to the functional form of the count data regression model. A count data regression model has been shown to be more appropriate for modeling recreation trip data (Hellerstein, 1991; Hellerstein & Mendelsohn, 1993; Cameron & Trivedi, 1998; Hilbe, 2011). This extension on Cameron's (1992) model was also adopted by González-Sepúlveda (2008) to study recreators at the El Yunque National Forest in Puerto Rico but the author here used a truncated count data distribution to model the recreation trip data.

González-Sepúlveda (2008) developed a utility-consistent version of the joint model used in González-Sepúlveda, Loomis, and González-Cabán (2008). González-Sepúlveda's (2008) extension of Cameron (1992) followed closely with that of Eom and Larson (2006) in deriving the utility-difference model but adopted the econometric framework used in González-Sepúlveda, Loomis, and González-Cabán (2008). The author also reported estimates of consumer surplus (CS) and median compensating surplus. The CS measure captures the net

benefits from access to recreation trips. The CVM question however, asked respondents about their ‘last’ or most recent trip, and not ‘all’ trips which would be more reflective of access. Therefore, the CVM scenario is different from the TCM scenario and also the welfare measures: *CS* from TCM, and the compensating variation (*CV*) from the CVM, do not equal.

The model presented in this study uses the utility-consistent frameworks of Cameron (1992), Eom and Larson (2006), and González-Sepúlveda (2008), but the scenarios reflected by the *CS* measure from TCM is the same for *CV* measure from CVM. The CVM questions follows a similar scenario as in Cameron (1992) while the TCM scenario is similar to that in González-Sepúlveda (2008) resulting in a  $CS = CV$ . The model presented in this paper demonstrates how understanding the language used in the CVM questions and the limitations of the TCM analysis can improve the integration of TCM and CVM information. The next section derives the utility-consistent framework used in this study.

### **3.3. Methodology**

#### *3.3.1. Derivation of Utility-difference Model using TCM Parameters*

Similar to Cameron (1992) and others, we express the utility maximization objective function with recreation trips  $x$ , a composite of all-other-goods  $AOG$ , travel cost  $p$ , cost of all-other-goods  $A$ , and income  $y$  as

$$(1) \quad \max_x u(y - px, x) \quad \text{subject to} \quad y = A + px$$

Here, the Marshallian demand function for recreational trips follows a semi-log form as given by

$$(2) \quad x(p, y) = \exp(\alpha + \beta p + \delta y + \mathbf{z}'\xi)$$

where  $\mathbf{z}$  is a vector of other factors, and  $\beta < 0$  and  $\delta > 0$  are assumed for theoretical consistency. The advantage of using the semi-log form is its consistency with the count data regression model. Due to the integer or count nature of recreation trip data, count data distributions such as the Poisson and negative binomial have been recommended (Hellerstein, 1991; Hellerstein & Mendelsohn, 1993; Cameron & Trivedi, 1998; Hilbe, 2011). Following Eom and Larson (2006) and González-Sepúlveda (2008), a quasi-expenditure function corresponding to the demand function in (2) is given by

$$(3) \quad \tilde{e}(p, u) = -\frac{1}{\delta} \ln \left( -\frac{\delta}{\beta} \exp(\alpha + \beta p + \mathbf{z}'\xi) - \delta \mathcal{C}(u) \right)$$

where  $\mathcal{C}(u)$  is a constant of integration. The quasi-expenditure function (3) is defined under  $0 \leq x < \beta/\delta$  (See Eom & Larson, 2006). For simplicity, we allow  $\mathcal{C}(u)$  to possess the same functional properties as the theoretical reference utility level  $u$  so that an indirect utility function corresponding to (3) can be expressed as



$$\begin{aligned}
(4) \quad \tilde{v}(p, y) &= -\frac{1}{\delta} \exp(-\delta y) - \frac{1}{\beta} \exp(\alpha + \beta p + \mathbf{z}'\xi) \\
&= \frac{-[\beta \exp(-\delta y) + \delta \exp(\alpha + \beta p + \mathbf{z}'\xi)]}{\delta\beta}
\end{aligned}$$

The validity of the above expression can be verified using *Roy's Identity*. In addition, the corresponding direct utility function can be obtained by multiplying (4) by the unit factor  $\exp(\delta y)/\exp(\delta y)$  and rearranging as follows.

$$\begin{aligned}
(5) \quad \tilde{u}(x_0^*) &= \frac{-[\beta + \delta \exp(\alpha + \beta p + \delta y + \mathbf{z}'\xi)]}{\delta\beta \exp(\delta y)} \\
&= \frac{-[\beta + \delta x_0^*]}{\delta\beta \exp(\delta y)}
\end{aligned}$$

where  $x_0^* > 0$  is the observed number of trips optimally chosen by the respondent. In this study, the utility-difference equation was model after the CVM question found in the 1996 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. The CVM question asked respondents: *Would you have taken any trips...if your total cost were \$B more than the amount you just reported?* (Boyle & Roach, 1998), where  $\$B$  is randomly chosen by the survey. In responding, we assume the respondent decides whether their utility of paying the additional  $\$B$  and continuing to take positive trips would have been greater than the utility of not paying and completely forgoing taking any trips. This “all or nothing” style of dichotomous choice questions have been used by many CVM practitioners (See Bishop, Heberlein, & Kealy, 1983; Cameron, 1988; Hanemann,

1984; McConnell, 1990; Cameron, 1992). Therefore, the utility of paying the additional \$ $B$  and taking positive trips can be expressed as

$$\begin{aligned}
 (6) \quad \tilde{u}(x_1^* > 0 | B > 0) &= \tilde{u}_0 + d\tilde{u} = \frac{-[\beta + \delta x_0^*]}{\delta \beta \exp(\delta y)} + \left( \frac{\partial \tilde{u}}{\partial p} dp \right) \\
 &= \frac{-[\beta + \delta x_0^*]}{\delta \beta \exp(\delta y)} - \frac{x_0^* dp}{\exp(\delta y)} \\
 &= \frac{-[\beta + \delta x_0^*] - \delta \beta x_0^* dp}{\delta \beta \exp(\delta y)} \\
 &= \frac{-[\beta + \delta x_0^*(1 + \beta dp)]}{\delta \beta \exp(\delta y)}
 \end{aligned}$$

where  $\partial \tilde{u} / \partial p = -\exp(\alpha + \beta p + \mathbf{z}'\xi)$ , and for present purposes  $dp = B/x_0^*$ . The utility of not paying \$ $B$  and foregoing participation is

$$(7) \quad \tilde{u}(x_1^* = 0 | B = 0) = \frac{-1}{\delta \exp(\delta y)}$$

Therefore given (6) and (7), the respondent is expected to answer *yes* to the CVM question if and only if

$$\begin{aligned}
 (8) \quad \Delta \tilde{u} &= \tilde{u}(x_1^* > 0 | B > 0) - \tilde{u}(x_1^* = 0 | B = 0) > 0 \\
 &= \frac{-[\beta + \delta x_0^*(1 + \beta dp)]}{\delta \beta \exp(\delta y)} - \left( \frac{-1}{\delta \exp(\delta y)} \right) > 0 \\
 &= \frac{-[\beta + \delta x_0^*(1 + \beta dp)] + \beta}{\delta \beta \exp(\delta y)} > 0
 \end{aligned}$$

$$\begin{aligned}
&= \frac{-[x_0^*(1 + \beta dp)]}{\beta \exp(\delta y)} > 0 \\
&= \frac{x_0^*(1 - |\beta| dp)}{|\beta| \exp(\delta y)} > 0
\end{aligned}$$

As noted in Cameron (1992), survey data leading to TCM expressions such as (2) reflect the respondent's utility maximizing quantity decisions given the current conditions. Thus, we expect the CVM model expressed in (8) to be based on the same preference structure as the TCM model in (2).

### 3.3.2. Estimation of Utility-consistent Model Parameters

Following previous studies that applied the utility-consistent framework, the utility-difference model was combined with the TCM count data model in a maximum likelihood function and the two models were estimated simultaneously. Parameter estimates of the utility-consistent model was found by maximizing the log-likelihood function

$$\begin{aligned}
\mathcal{L} = & \sum_{i=1}^n \{y_{cvm,i} \ln[\Phi(W_i)] + (1 - y_{cvm,i}) \ln[1 - \Phi(W_i)]\} \\
& + \sum_{i=1}^n \left\{ y_{tcm,i} \ln \left( \frac{\alpha \exp(\mathbf{x}'_i \theta)}{1 + \alpha \exp(\mathbf{x}'_i \theta)} \right) - \left( \frac{1}{\alpha} \right) \ln(1 + \alpha \exp(\mathbf{x}'_i \theta)) + \ln \Gamma \left( y_{tcm,i} + \frac{1}{\alpha} \right) \right. \\
& \left. - \ln \Gamma(y_{tcm,i} + 1) - \ln \Gamma \left( \frac{1}{\alpha} \right) - \ln[1 - (1 + \alpha \exp(\mathbf{x}'_i \theta))^{-1/\alpha}] \right\}
\end{aligned}$$

where  $y_{cvm,i}$  is the CVM dichotomous response variable,  $\Phi(\cdot)$  is the standard normal cumulative distribution function,  $y_{tcm,i}$  is the TCM count variable that follows a zero-truncated negative binomial distribution with mean function  $\exp(\mathbf{x}'_i\theta) = \exp(\alpha + \beta p_i + \delta y_i + \mathbf{z}'_i\xi)$ , where  $\mathbf{x}_i$  is a vector of covariates and  $\theta$  is a vector of model parameters. For the utility-consistent joint model,

$$W_i = \left( \left( \frac{\Delta\tilde{u}}{\sigma} \right) + \rho Z_{tcm,i} \right) / \sqrt{(1 - \rho^2)}$$

where  $\Delta\tilde{u}$  is the utility-difference expression shown in equation (8),  $\sigma > 0$  is the standard normal standard deviation, and  $\rho$  is the correlation coefficient. Also,  $Z_{tcm,i}$  is a standardized fitted error in the demand function (Cameron, 1992) defined as

$$Z_{tcm,i} = \left( y_{tcm,i} - \mathbf{E}(y_{tcm,i}) \right) / \sqrt{\text{var}(y_{tcm,i})}, \text{ with conditional expected value function}$$

$$\mathbf{E}(y_{tcm,i}) = \frac{\exp(\mathbf{x}'_i\theta)}{1 - (1 + \alpha \exp(\mathbf{x}'_i\theta))^{-1/\alpha}}$$

The conditional variance function as shown in Grogger and Carson (1991) is

$$\text{var}(y_{tcm,i}) = \frac{\mathbf{E}(y_{tcm,i})}{(1 + \alpha \exp(\mathbf{x}'_i\theta))^{-1}} \left\{ 1 - [(1 + \alpha \exp(\mathbf{x}'_i\theta))^{-(1+\alpha)/\alpha}] \mathbf{E}(y_{tcm,i}) \right\}$$

It should be mentioned that the utility difference formulation in (8) is evaluated with travel cost set to infinity because of the semi-log specification of demand. As explained by Hellerstein and Mendelsohn (1993), the expected compensating variation measure  $E[CV]$  is equivalent to the expected consumer surplus measure  $E[CS]$  as a consequence of evaluating the welfare measures with the travel cost in (7) set to infinity.

### *3.2.3. Econometric Equivalence between Compensating Variation and Consumer Surplus*

Hellerstein and Mendelsohn (1993) showed that because count data models can be derived from repeated discrete choices, the standard expected consumer surplus  $E[CS]$  formula provides an approximation to a measure of the expected compensating variation<sup>xix</sup>. While the derivation of this result is shown in Hellerstein and Mendelsohn (1993), we provide the resulting expression below.

$$(9) \quad E[CV] = \int_{p^a}^{p^b} \sum_t^T \pi_t(p, y) dp \approx \int_{p^a}^{p^b} \lambda(p, y) dp = E[CS]$$

where  $\pi_t(p, y)$  is the probability that the good will be chosen on day  $t$ ,  $\lambda(p, y)$  is the expected value of the count variable with  $p^a$  and  $p^b$  as the lower and upper price limits of integration, and income  $y$ . The implication of (9) used in this

---

<sup>xix</sup> This conclusion has also been reach by Willig(1976) in his support for the use of consumer surplus as an acceptable approximation for the compensating variation.

paper and mentioned by Hellerstein and Mendelsohn (1993) however is that when  $\lambda(p, y) = \exp(\mathbf{x}'_i\theta)$  and  $p^b = \infty$  in (1),  $E[CV] = E[CS] = -\lambda/\beta_p$ , where  $\beta_p$  is the coefficient for  $p$ . This equivalence is convenient in that welfare changes can be interpreted in terms of net benefits or willingness to pay. However, it is unclear whether the  $E[CV] = E[CS]$  result can be assumed to hold in general and was not explained in Hellerstein and Mendelsohn (1993) whether this equivalence should be true even when TCM data is used alone. As mentioned earlier, the utility-consistent model presented in this paper imposes this equivalence within the modeling framework. The CVM questions in the survey used in this study presented respondents with an ‘all or nothing’ scenario as reflected in the utility-difference equation in (8). In fact, it could be argued that both CVM and TCM must be evaluated in terms of the ‘all or nothing’ scenario for this relation to be realized. A graphical depiction of this scenario is shown in figure 3.1.

Figure 3.1 illustrates the scenario where the consumer is completely substituting away from recreation trips (Trips) towards All Other Goods (AOG) because their current travel cost increased from  $P$  to  $P^*$ . At  $P^*$ , the consumer does not take any recreation trips and allocates their entire budget towards AOG. This is illustrated as a movement from point A to B, representing a decrease in utility from  $U_0$  to  $U_1$ , and a pivot of the budget line B to where it aligns with the y-axis. In this case, the indifference curves must intersect the y-axis because a choke price  $P^*$  exists, thus suggesting a quasi-linear function form for utility. The compensating variation (CV), defined as the difference in expenditure (or utility) functions when the individual faces the choke price (Whitehead & Aiken, 2007),

can be measured as the difference between points C and D. Notice this is exactly equal to the consumer surplus (*CS*), as measured by the area underneath the inverse demand function *D* and above price *P* since the Marshallian ( $D^M$ ) and Hicksian ( $D^H$ ) demands are equal here. This is easily shown by setting (8) to zero in order to isolate *B*, which serves as a measure of *CV* (Hanemann, 1984). Taking the expected value of the above expression gives the expected compensating variation measure which is exactly equal to the expected consumer surplus measure derived from the travel cost model.

$$(10) \quad E[CV] = E[B] = \frac{E(x_0^*)}{|\beta|} = \beta^{-1} \cdot E(x_0^*) = E[CS]$$

where the truncated mean  $E(y_{tcm,i})$  is used in place of  $E(x_0^*)$ . Expression (10) is an econometric expression of the theoretical equivalence between *CS* and *CV*.

### 3.4. Data

Wildlife watching recreation in the United States is one of the most popular forms of outdoor recreation. An estimated 71.8 million U.S. residents enjoyed observing, feeding, or photographing wildlife in 2011 (USFWS, 2012). The number of wildlife watchers has grown steadily since 1996 (14 percent) and has contributed in various ways to the health of the nation's, states', and local economies. Spending on wildlife-watching activities generated \$18.2 billion in tax revenues and 1.1 million jobs (Leonard, 2008). Total industry output is estimated to be \$122.6 billion in major sectors in the U.S. such as retail trade,

manufacturing, and hospitality and food services. There are also important implications for wildlife management which has led to some pioneering studies on the relationship between wildlife watching recreation and management programs (Cicchetti, 1973; Bishop, 1978; Hay & McConnell, 1979; Hay & McConnell, 1984; Rockel & Kealy, 1991).

Wildlife watching data was obtained from the 1996 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation which is conducted every 5 years by the U.S. Fish and Wildlife Service in conjunction with the U.S. Census Bureau. Although the analysis from this data is outdated, it is still useful for examining and comparing different econometric approaches. The data includes an array of information on wildlife watching activities by residents around their homes and at least one mile away. Data collection was carried out in two phases: an initial screening survey to identify current and potential wildlife watchers, and a series of follow-up interviews on the selected sample to collect detailed data on their wildlife watching activities. The final sample consisted of 11,759 unique records (individuals) that reported the total number of trips for the purpose of observing, feeding, or photographing wildlife at least one mile from their home and the respective expenditures. Nonparticipants, who accounted for about 62 percent of the sample, were excluded from the analysis because of missing travel cost data. The final sample size used in this study was 4,482 records. Descriptions of the variables and their summary statistics are reported in table 3.1.

Contingent valuation method (CVM) questions were included for all persons who had taken at least one trip during 1996. CVM questions were



presented using the dichotomous-choice format. Respondents were asked to provide a *yes* or *no* response to the question:

*...Would you have taken any trips during 1996 for the PRIMARY PURPOSE of observing, photographing, or feeding wildlife in [state] if your total costs were \$[bid value] more than the amount you just reported?*

The travel cost variable was constructed using the respondent's total transportation costs, access costs, and their opportunity cost of time. The cost of time can have a significant influence on the estimation of recreation demand. As shown in previous studies, ignoring time costs can lead to an overestimate of the travel cost parameter, and therefore an underestimate of benefits (Cesario & Knetsch, 1976; McConnell & Strand, 1981; Bockstael, Strand, & Hannemann, 1987). The opportunity cost of time was calculated using the reported total number of potential work days (250 days) and an estimate of their personal income in order to arrive at an estimate of their daily earnings. Daily earnings were used instead of hourly wage because the data on the amount of time spent on each trip was recorded in days. Following González-Sepúlveda, Loomis, and González-Cabán (2008), the wage factor 0.33 was used but divided by 8 (estimated daily work hours) to arrive at the factor 0.04<sup>xx</sup>.

Other variables included here are an estimate of personal income, age, marital status, gender, and indicators of wildlife watching activities, other outdoor recreation activities, and measures of the amount of forest and

---

<sup>xx</sup> There is much variation in the literature regarding which wage factor should be used. For instance, McConnell and Strand (1981) suggests the wage factor 0.612 for anglers which was adopted by Dhazn, Woodward, Ozuna Jr., and Griffin (2003). Rockel and Kealy (1991) however, applied 0.30 and 0.60 for wildlife watchers, while Zawacki, Marsink, and Bowker (2000) provided estimates that used 0.0, 0.25, and 0.50 for wildlife watchers.

rangeland at their destinations. Table 3.2 provides the summary statistics of the variables used in the analysis.

### 3.5. Results

The results of the utility-consistent model are reported in table 3.2. For comparison, the results of individual TCM and CVM models (INDIVIDUAL) and a joint model of the TCM and CVM data (JOINT TCM-CVM) are also reported in table 3.2. Similar to the utility-consistent model, the zero-truncated negative binomial regression model (ZTNB) was used to estimate the TCM portion of the individual and joint models, while Cameron and Englin's (1997) censored-normal (CENSORED) model was used to estimate the CVM portions. The framework of the joint model follows closely with that used in González-Sepúlveda, Loomis, and González-Cabán (2008). Similar to González-Sepúlveda, Loomis, and González-Cabán (2008), the results for the joint model shows there was a slight improvement in statistical performance compared to the individual models as indicated by comparing the log-likelihood value (-15450.08) to the sum of the values from the individual ZTNB and CENSORED models (-15460.03). This improvement is also reflected in the Akaike Information Criterion and the Bayesian Information Criterion measures. The travel cost coefficient in the ZTNB portion of the joint model is slightly larger than in the individual model leading to a smaller  $E[CS]$  estimate. The differences across the other ZTNB variables are also small, and no substantial differences across the predicted values  $E[\text{trips}|x'\theta, \text{trips}>0]$ . The differences in the estimates across the CVM portions are also unremarkable. The  $E[CV]$  value is estimated to be smaller in the joint model

compared to the individual model and the same is true for the  $\sigma$  estimate. The  $\rho$  estimate is positive and significant indicating that the CVM and TCM portions of the joint model are positively correlated. However, the magnitude of  $\rho$  (0.071) suggests the association between the two models is not very strong. Reported also are the per-trip values of  $E[CV]$  for the INDIVIDUAL and JOINT TCM-CVM models to compare with  $E[CS]$ . It is clear that they are not equal in these cases<sup>xxi</sup>.

The utility-consistent joint model results were obtained using the formulas presented in the methodology section. Most of the coefficient estimates are in the range of the other two ZTNB models but the estimated  $\sigma$  is quite large. One reason for this may be due to the fact that the utility-difference model in the CVM portion is not a regression model as is the case in the individual and joint models. Therefore, it may have not been able to explain much of the variation in the CVM responses which could lead to an inflated  $\sigma$ . Also, the log-likelihood values are lower, and the information criterion values higher than the joint model. However, it is unclear whether we can accurately compare the performance of the utility consistent model with the individual and joint models using these measures since the modeling frameworks are very different. The  $\rho$  estimate is also positive and significant so the joint and utility-consistent models both agree that the CVM and TCM portions move in the same direction. The welfare measures for the CVM and TCM portions are the same in this model, that is  $E[CS] = \text{per-trip } E[CV]$ . The value (63.51) is lower than that from the other ZTNB

---

<sup>xxi</sup> It is possible that scaling the variances of the two model components, where one is normalized and the other a proportion of the first might provide a better comparison of the results. However, it was not obvious how this could be accomplished since one model uses a binary variable and the other a count variable.

models and much lower than the CENSORED models still. This is in contrast to the results found in González-Sepúlveda (2008) where the utility-consistent model provided  $E[CS]$  estimates that lied in between those of the individual ZTNB and CENSORED models.

### **3.6. Discussion and Conclusion**

This paper estimated a utility-consistent joint model of wildlife watching recreation using TCM count data and CVM dichotomous choice data. It was argued in this paper that consistency across the TCM and CVM scenarios when combining data allowed for better cohesion in the two types of information. Also, certain types of scenarios such as the “all or nothing” scenario, led to the TCM welfare measure being equal to the CVM welfare measure but it was shown that this equivalence needed to be explicitly incorporated into the modeling framework to hold. Thus, the novelty of the utility-consistent model presented here was that it was consistent across the welfare measures as well. Although it was unclear how its statistical performance fared against other modeling frameworks, its adoption may be argued purely from a conceptual standpoint. Thus, the argument for combining RP and SP data is still largely supported in this paper, although justifying the method by which it should be carried out may require further study. The utility-consistent model presented in this paper demonstrated how greater cohesion between the RP and SP information can be obtained by taking care at matching the modeling frameworks and scenarios at the design phase of the survey. While the focus of this paper was on combining TCM and CVM data, expanding the integration exercise to data from other

nonmarket valuation methods such as the hedonic price model and choice experiment surveys, could prove fruitful.

Table 3.1: Summary statistics of variables and descriptions

Variable	Mean	Std. Dev.	Min	Max	Description
<i>n</i> = 4482					
<i>CVM 'Yes' Response (=1)</i>	0.39	0.49	0	1	Dichotomous choice CVM response
<i>Bid</i>	333.79	126.38	22	1067	Bid value randomly selected by survey
<i>trips</i>	8.98	19.35	1	250	Total number of trips taken in 2006 by respondent
<i>travel cost</i>	17.73	40.07	0.06	486.63	Defined as: (transportation & access costs) + <i>opportunity cost</i>
<i>opportunity cost</i>	9.61	19.95	0.06	471.69	Defined as: $0.04 \text{earnings} (\text{days}/\text{trips})^a$
<i>personal income</i> (\$1000)	33.86	28.40	2.30	155.15	Respondent's personal income

<sup>a</sup>To calculate *opportunity cost* of time, we set the total number of potential work days at 250 and used the household income variable to obtain daily earnings. Daily earnings are useful here because the time cost, which is measured by the amount of time spent wildlife watching, is recorded in days. We divided the commonly used wage factor (.333) by 8 to arrive at the earnings factor 0.04.

<sup>b</sup>Personal income is a predicted value obtained using interval regression. Model results are available upon request.

Table 3.1: Summary statistics of variables and descriptions continued

Variable	Mean	Std. Dev.	Min	Max	Description
<i>age</i>	43.34	13.94	16	90	Age of respondent
<i>married</i> (=1)	0.69	0.46	0	1	Respondent is married
<i>male</i> (=1)	0.50	0.50	0	1	Respondent is male
<i>photo</i> (=1)	0.41	0.49	0	1	Respondent photographed wildlife around their home
<i>parks</i> (=1)	0.30	0.46	0	1	Respondent visited a public park within a mile of their home
<i>private</i> (=1)	0.42	0.49	0	1	Respondent took wildlife watching trips to privately owned land
<i>hunted</i> (=1)	0.07	0.25	0	1	Respondent hunted
<i>fished</i> (=1)	0.20	0.40	0	1	Respondent fished
<i>forrest &amp; rangeland proportion</i>	0.43	0.17	0.05	0.84	Proportion of land acres in state designated as forrest or rangeland

Note: This sample only includes respondents who have taken at least one trip for the primary purpose of observing, photographing, or feeding wildlife.

Table 3.2: Individual and joint model, and utility-consistent model results

	INDIVIDUAL		JOINT TCM-CVM		UTILITY- CONSISTENT
	ZTNB	CENSORED	ZTNB	CENSORED	ZTNB
<i>travel cost</i> (10\$)	-0.1538**** (-5.21)	—	-0.1557**** (-5.25)	—	-0.1574**** (-5.24)
<i>personal income</i> (1000\$)	-0.0013 (-0.69)	0.0036**** (5.55)	-0.0015 (-0.79)	0.0035**** (5.52)	-0.0020 (-1.04)
<i>age</i> (10 yrs.)	0.2048**** (6.09)	0.0187* (1.80)	0.2067**** (6.85)	0.0179* (1.76)	0.2058**** (6.17)
<i>married</i> (=1)	-0.2699*** (-3.05)	-0.0415 (-1.32)	-0.2755*** (-2.77)	-0.0402 (-1.31)	-0.2678*** (-2.98)
<i>male</i> (=1)	0.1759* (1.89)	0.0279 (0.92)	0.1664* (1.81)	0.0290 (0.97)	0.1585 (1.69)
<i>photo</i> (=1)	0.3058**** (3.76)	0.1499**** (4.66)	0.3057**** (3.42)	0.1478**** (4.72)	0.2826**** (3.47)
<i>parks</i> (=1)	0.4501**** (4.82)	0.0860*** (2.64)	0.4563**** (4.64)	0.0847*** (2.66)	0.4425**** (4.69)
<i>private</i> (=1)	0.7919**** (9.91)	0.1629**** (4.92)	0.7848**** (9.27)	0.1627**** (5.03)	0.7564**** (9.36)
<i>forrest &amp; rangeland proportion</i>	0.1494 (0.67)	-0.0268 (-0.34)	0.1790 (0.75)	-0.0292 (-0.38)	0.1653 (0.73)
<i>hunted</i> (=1)	0.4504*** (3.13)	0.0080 (0.14)	0.4335*** (3.11)	0.0087 (0.16)	0.4218*** (3.04)
<i>fished</i> (=1)	0.1724* (1.70)	-0.0556 (-1.53)	0.1869* (1.82)	-0.0536 (-1.51)	0.1983* (1.92)
<i>constant</i>	-0.6214* (-1.95)	-0.1754** (-2.13)	-0.6382 (-0.91)	-0.1930** (-2.36)	-0.5320* (-1.67)
$\ln(\alpha)$	2.4022*** (8.54)	—	2.4067**** (3.96)	—	2.3879**** (8.53)
$\alpha$	11.05		11.10		10.89
$\ln(\sigma)$	—	-0.4228**** (-3.78)		-0.4470**** (-4.09)	20.10**** (11.12)
$\sigma$	—	0.655		0.640	5.362e+08
$\operatorname{atanh}(\rho)$	—	—	—	0.0708**** (4.37)	0.0932**** (5.09)
$\rho$	—	—	—	0.071	0.093



Table 3.2: Individual and joint model, and utility-consistent model results continued

	INDIVIDUAL		JOINT TCM-CVM		UTILITY- CONSISTENT
	ZTNB	CENSORED	ZTNB	CENSORED	ZTNB
<i>sample mean</i>			8.98		
E[trips   x'0, trips > 0]	8.79	—	8.74	—	8.95
E[CS]	65.03	—	64.22	—	63.51
E[CV]	—	146.59	—	120.61	—
Per-trip E[CV]		16.67		13.80	63.51
log L	-12592.26	-2867.77	-15450.08		-15679.97
		-15460.03			
aic	25210.52	5759.54	30952.16		31389.95
		30970.06			
bic	25293.82	5836.43	31118.76		31486.06
		31130.25			
n			4482		

t statistics in parentheses using robust standard errors clustered by person designation.

\* p<0.10, \*\* p<0.05, \*\*\* p<0.01, \*\*\*\* p<0.001

Estimates were obtained using the available sampling weights provided in the data.

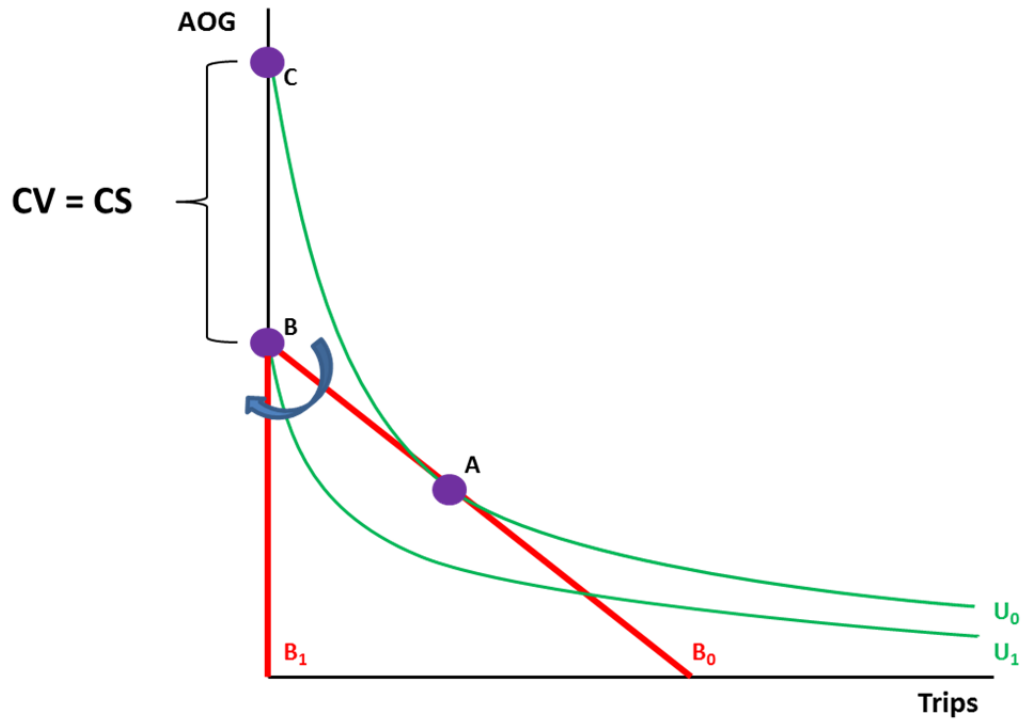
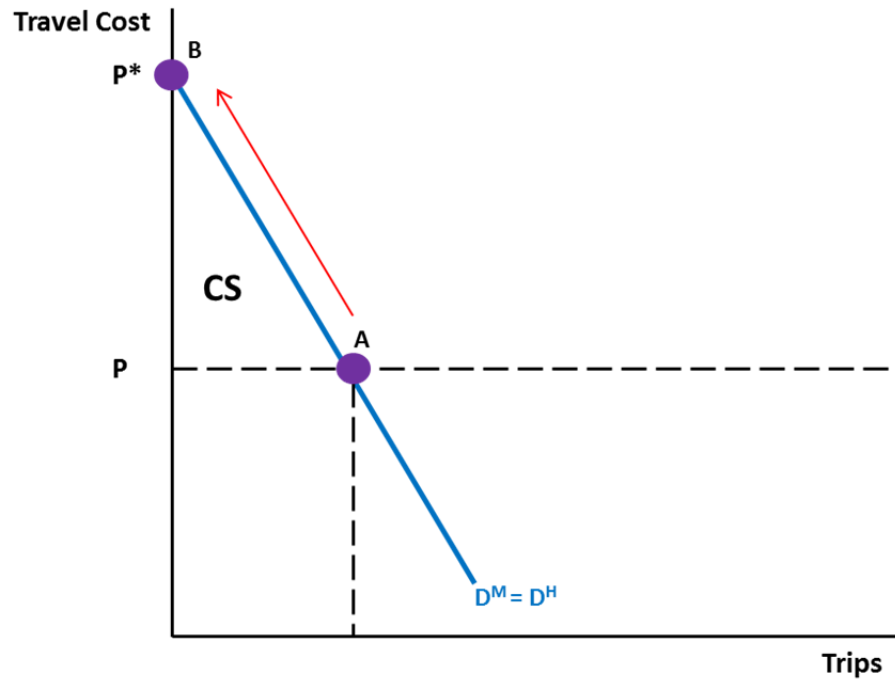


Figure 3.1: Effects of complete substitution from Trips to AOG

## **CHAPTER 4: Recoding Double-bounded Dichotomous Choice Contingent Valuation Method Data with Nonresponse**

### **4.1. Introduction**

The double-bounded dichotomous choice (DB) format in contingent valuation method (CVM) surveys was introduced as a way of obtaining greater statistical efficiency compared to a single-bounded (SB) format (Hanemann, Loomis, and Kanninen, 1991). Similar to a SB format, DB asks survey participants if they would be willing to pay a certain amount of money (BID), contingent upon realizing a described scenario, where BID is the basis for measuring willingness to pay (WTP) (Bishop & Heberlein, 1979; Hanemann, 1984). The difference is that DB includes a *follow-up* question that depends on the response to the *initial* question. For example, suppose the respondent answers *yes* to the initial question. Then a follow-up question asks if they (respondent) would be willing to pay a higher amount, keeping the scenario constant. If the respondent answers *no*, then the follow-up question asks if they would be willing to pay a lower amount, again keeping the scenario constant.

Since its initial introduction, researchers have noticed certain tendencies of respondents in that they may answer the follow-up questions differently than how they answered the initial question (See Haab & McConnell, 2002). For example, respondents initially answering *yes* may answer *no* to the follow-up question because they feel as if they are being exploited. Along the same lines, respondent answering *no* to the initial question may feel the quality of the good

may be lower than expected when faced with a lower bid amount in the follow-up question. The fundamental problem here is that the respondent's expectations concerning the good may have changed from the initial response to the follow-up response. This behavior has been referred to as *strategic behavior* to describe how respondents may strategically alter their responses in order to obfuscate their true preferences. The particular case examined in this paper is where respondents fail or refuse to answer the follow-up question, or nonresponse<sup>xxii</sup>. Failing to answer may be caused by an inability to decide which course of action should be preferred. Refusing to answer may be a sign of strategic behavior in that the respondent may be trying to deter additional questions being asked from the surveyor. Here, there may be a disconnect between the respondent's actual preference structure and the preference structure reflected by their responses (or nonresponses). Whether nonresponses are due to indecisions or strategic refusals, there are losses in statistical efficiency and bias when observations with nonresponses are excluded from the analysis through list-wise deletion (Brox, Kumar & Stollery, 2003). In this paper, a simple recoding procedure for DB data with nonresponse is introduced that allows the analyst to preserve the observation with the missing data and curtail the loss in statistical efficiency and bias from list-wise deletion.

---

<sup>xxii</sup> Issues have been raised as to whether the questions are incentive compatible (Alberini, Kanninen & Carson, 1997; Whitehead, 2002) and also whether the elicited values are consistent across the initial and the follow-up questions (Cameron & Quiggin, 1994). Cameron and Quiggin (1994) find that the value distributions examine in their study are not the same across the initial and follow-up responses. Nevertheless, some researchers have extended the bidding process to a third level or triple-bounded (Langford, Bateman & Langford 1996), or developed other variants such as the one-and-one-half bound (Cooper, Hanemann & Signorello, 2002) and the multiple-bounded formats (Welsh & Poe, 1998). Unfortunately, this scope of this study is limited to addressing the nonresponse itself and therefore does not investigate the underlying causes of strategic nonresponse.

The basic idea is to recode the data so that the remaining (non-missing) information from the initial response can still be incorporated into the analysis. In other words, respondents that do not answer the follow-up question will have their data recoded so that their initial response becomes their final response. The main objective of this study is to compare the statistical efficiencies and biases from recoding and list-wise deletion. The double-bounded interval model introduced by Hanemann, Loomis, and Kanninen (1991) is used to obtain the parameter estimates from Monte Carlo simulations. Two types of nonresponse are examined in this study: random nonresponse and systematic nonresponse. The next section reviews the literature on nonresponse in contingent valuation surveys and its effects, and conventional approaches to address nonresponse in surveys. Section 4.3 presents the recoding procedure applied to nonresponse in the DB follow-up question and the Monte Carlo simulation used to create the nonresponse scenarios. The results of the simulation are reported in section 4.4, and section 4.5 concludes the paper with discussions about the findings.

## **4.2. Literature Review**

### *4.2.1. Nonresponse in Contingent Valuation Surveys*

#### *4.2.1.1. Unit versus Item Nonresponse*

Unit nonresponse (UNR) occurs when recipients of surveys either fail to return the survey or it is returned without any responses. Item nonresponse (INR) occurs when only certain questions are left unanswered. Although contingent valuation method (CVM) surveys are at risk of both types of nonresponses, UNR is considered to be a more serious issue and addressed more often in the

literature. In CVM surveys, the questions having the highest propensity for nonresponse are the bid or WTP questions (Mitchell & Carson, 1989; Dietz, 2001). Because of the importance of WTP questions in CVM, INR in these questions may have the same effect as a UNR by way of list-wise deletion of observations.

#### *4.2.1.2. Sources of Nonresponse*

UNR occurs when the sampled unit is missing because respondents either explicitly refused to participate or there is some failure in establishing contact. INR can occur when individuals are asked to disclose sensitive information such as income and they refuse. Incidentally, studies have shown that the level of income may have a systematic effect on nonresponse (Korinek, Mistiaen & Ravallion, 2007). INR may also arise because individuals are either disinterested in the subject matter of the question or are simply indifferent. Also, if there are too many questions being asked or if the questions themselves are mentally exhausting, individuals may experience survey fatigue which will lead to INR (Riphahn & Serfling, 2005 and references therein).

Another source of INR is the removal of responses by the analyst. Analysts are sometimes faced with deciding the validity of certain responses. For instance, participants may sometimes provide a negative response to express extreme dislike or protest against the question and not necessarily the subject at hand (See Halstead, Luloff, & Stevens, 1992). By the same token, participants may be overtly positive in their response such as expressing that they would contribute more than they are capable in an attempt to influence the final results. Strategic

nonresponse may be another symptom of the issue studied by Cameron and Quiggin (1994) where the preference structure reflected by the initial and follow-up responses are not the same. In such cases, the analyst may resort to discarding the observation. As mentioned by Brox, Kumar and Strollery (2003) however, list-wise deletion of observation may not always be a viable option if the resulting sample size is small as this may bias the estimates (Dalecki, Ilvento, & Moore, 1988).

#### *4.2.1.3. Consequences of Nonresponse*

When the sample available for analysis is not representative of the target population due to nonresponse, parameter estimates may be biased. Nonresponse bias may occur if certain subgroups are under or over represented in the sample thereby skewing the population values. Nonresponse bias can also take the form of sample size bias if the sample available for estimation is too small and lacks any real variation. If the rate of nonresponse is systematically influenced by factors such as income or the WTP amount, a self-selection bias can occur. Self-selection refers to when certain individuals tend to self-select themselves out of the sample (Greene, 2003). Conducting inferences on a sample where self-selection is prevalent can lead to parameter estimates that are biased and inconsistent. Both UNR and INR can lead to nonresponse bias and (self) selection bias, where INR may operate exactly like UNR when certain key variables are missing. Consider also the situation where INR occurs within a sample already hindered by UNR. Here, we may have selection that is driven by one factor causing UNR, while driven by another factor causing INR. Messonnier

et al. (2000) provides an insightful analysis of the two effects and shows how to account for them when data describing the unit nonrespondents are available.

#### *4.2.2. Conventional Methods for Addressing Nonresponse*

##### *4.2.2.1. Weight Adjustments*

Population weights can be applied to make the sample more representative of the target population if certain groups are underrepresented. Typically, weights are calculated based on past demographic (income, gender, race, education), and geographical distributions which are obtained from previous surveys. Most national surveys can provide reliable weights because the sample size is typically large and the survey has been conducted for several decades. On the other hand, CVM samples tend to be smaller and more variable in their implementation and subject matter, so reliability of weights may be difficult to ascertain. For instance, if WTP is not temporally reliable (Whitehead & Aiken, 2007) the use of weights from previous studies could lead to maladjustments.

##### *4.2.2.2. Imputation and Nonparametric Methods*

Imputation methods can be useful when INR is random. The idea is to use information on the respondents to predict values for nonrespondents. For example, Whitehead (1994) used imputation for missing income responses, while Bhat (1994) showed how this could be accomplished when the variables are categorical. A more comprehensive approach to imputing missing data points was introduced by Rubin (1987) where estimates from multiple imputed data sets are averaged together. However, imputation should not be used when the



subsample of respondents do not accurately represent the nonrespondents. This may happen if some individuals are selecting themselves out of the sample. Imputation in the presence of self-selection can lead to estimates that are biased and inconsistent. Horowitz and Manski (1998) corrected for nonresponse bias by estimating asymptotic bounds on the bias due to nonresponse then applying them to the population parameters. This method however, has not been commonly adopted as many CVM analysts prefer parametric or semi-parametric methods.

#### *4.2.2.3. Heckman's (1979) Two-Step Procedure*

The Heckman (1979) procedure is the most popular method for correcting nonresponse bias. In particular, it is designed to detect and correct for self-selection bias. It does so by augmenting the primary equation with a selection variable. Whitehead, Groothuis and Blomquist (1993) applied the Heckman procedure to SB data with UNR. Messonnier et al. (2000) used the procedure to correct for both UNR and INR. Brox, Kumar and Strollery (2003) added the Heckman procedure to interval regression to analyze payment card data with INR. Harpman, Welsh and Sparling (2004) used a bivariate probit model with selection to correct for UNR in multiple-bounded data. Similar to weight adjustments, correction for selection depends on data describing nonrespondents which may not be available with UNR. Correction for selection in INR may be limited by the size of the usable sample and is highly sensitive to misspecification (Puhani, 2000).

The recoding procedure presented in this paper serves as a simple alternative to the more sophisticated approaches mentioned above. In the next section, the intuition behind the recoding procedure and its simplicity of application are demonstrated.

### 4.3. Methodology

#### 4.3.1. Econometric Model for Double-bounded Dichotomous Choice Data

The willingness to pay function for the  $i^{\text{th}}$  individual can be expressed as

$$(1) \quad WTP_i = x_i' \beta + u_i$$

where  $x_i$  is a vector of independent factors,  $\beta$  is a vector of associated coefficients, and  $u_i$  is the stochastic error term with  $u_i \sim N(0, \sigma^2)$ . Let  $y_i^1$  and  $y_i^2$  be binary or dichotomous variables that reflect the individual's response to the initial and the follow-up questions, and  $b^1 < b^2$  be positive constants representing the respective bid values. The probabilities of the response scenarios are

$$(2) \quad y_i^1 = 1, y_i^2 = 0$$

$$\begin{aligned} \Pr(\text{yes}, \text{no}) &= \Pr(b^1 \leq WTP < b^2) \\ &= \Pr(b^1 \leq x_i' \beta + u_i < b^2) \\ &= \Pr\left(\frac{b^1 - x_i' \beta}{\sigma} \leq \frac{u_i}{\sigma} < \frac{b^2 - x_i' \beta}{\sigma}\right) \\ &= \Phi\left(\frac{b^2 - x_i' \beta}{\sigma}\right) - \Phi\left(\frac{b^1 - x_i' \beta}{\sigma}\right) \end{aligned}$$

$$(3) \quad y_i^1 = 1, y_i^2 = 1$$

$$\begin{aligned} \Pr(\text{yes}, \text{yes}) &= \Pr(WTP > b^1, WTP > b^2) \\ &= \Pr(x_i' \beta + u_i > b^1, x_i' \beta + u_i > b^2) \\ &= \Pr(u_i > b^2 - x_i' \beta) \\ &= 1 - \Phi\left(\frac{b^2 - x_i' \beta}{\sigma}\right) \end{aligned}$$

$$(4) \quad y_i^1 = 0, y_i^2 = 1$$

$$\begin{aligned} \Pr(\text{no}, \text{yes}) &= \Pr(b^2 \leq WTP < b^1) \\ &= \Pr(b^2 \leq x_i' \beta + u_i < b^1) \\ &= \Pr\left(\frac{b^2 - x_i' \beta}{\sigma} \leq \frac{u_i}{\sigma} < \frac{b^1 - x_i' \beta}{\sigma}\right) \\ &= \Phi\left(\frac{b^1 - x_i' \beta}{\sigma}\right) - \Phi\left(\frac{b^2 - x_i' \beta}{\sigma}\right) \end{aligned}$$

$$(5) \quad y_i^1 = 0, y_i^2 = 0$$

$$\begin{aligned} \Pr(\text{no}, \text{no}) &= \Pr(WTP < b^1, WTP < b^2) \\ &= \Pr(x_i' \beta + u_i < b^1, x_i' \beta + u_i < b^2) \\ &= \Pr(u_i < b^2 - x_i' \beta) \end{aligned}$$

$$= \Phi\left(\frac{b^2 - x'_i\beta}{\sigma}\right)$$

As shown in Cameron and James (1992), a probit model provides estimates that are a composite of  $\beta$  and  $\sigma$  in the form of  $\hat{\alpha} = \hat{\beta}/\hat{\sigma}$  and  $\hat{\delta} = -1/\hat{\sigma}$ , where  $\hat{\alpha}$  and  $\hat{\delta}$  are the estimated coefficients associated with the independent factors and the bid variable respectively. The expected willingness to pay can therefore be calculated as

$$(6) \quad E(WTP|x) = x' \left(-\frac{\hat{\alpha}}{\hat{\delta}}\right) = x' \hat{\beta}$$

Following Hanemann, Loomis, and Kanninen (1991),  $\beta$  and  $\sigma$  are estimated directly where the log-likelihood function to be maximized is

$$\begin{aligned} \mathcal{L} = & \sum_{i=1}^n \left\{ \pi_i^{yn} \ln \left( \Phi \left( \frac{b^2 - x'_i\beta}{\sigma} \right) - \Phi \left( \frac{b^1 - x'_i\beta}{\sigma} \right) \right) + \pi_i^{yy} \ln \left( 1 - \Phi \left( \frac{b^2 - x'_i\beta}{\sigma} \right) \right) \right. \\ & \left. + \pi_i^{ny} \ln \left( \Phi \left( \frac{b^1 - x'_i\beta}{\sigma} \right) - \Phi \left( \frac{b^2 - x'_i\beta}{\sigma} \right) \right) + \pi_i^{nn} \ln \left( \Phi \left( \frac{b^2 - x'_i\beta}{\sigma} \right) \right) \right\} \end{aligned}$$

where  $\pi_i^{yn}$ ,  $\pi_i^{yy}$ ,  $\pi_i^{ny}$ , and  $\pi_i^{nn}$  are binary variables to indicate each of the cases, and  $\Phi(\cdot)$  is the standard normal cumulative distribution function.

The main hypothesis tested in this study is that recoding leads to better statistical performance. In particular, both statistical efficiency and bias are

examined with data exhibiting random and systematic nonresponse. For efficiency, the root mean squared error (RMSE) is used as the unit of comparison. For statistical bias, we examine the estimated parameters and *WTP* for trends and deviations from true values.

#### 4.3.2. Monte Carlo Simulations

In order to simulate the scenario where the respondent fails or refuses to answer the follow-up question in a DB CVM survey, a Monte Carlo simulation was used. Monte Carlo simulations are useful for comparing competing statistical methodologies with artificially generated data. For example, Bohara, Kerkvliet, and Berrens (2001) used a Monte Carlo simulation to examine the influence of negative willingness to pay on the performance of alternative distributional assumptions. In this study however, the data generating process began with the DB choice format responses in (2)–(5) given by

$$(7) \quad y_i^1 = 1, y_i^2 = 1$$

$$\text{Response}(yes, yes)_i = 1 \quad \text{if} \quad (5 - 0.5b_i^{2b} + 0.75X_i - 1.25Z_i + e_i) \geq 0$$

$$(8) \quad y_i^1 = 0, y_i^2 = 0$$

$$\text{Response}(no, no)_i = 1 \quad \text{if} \quad (5 - 0.5b_i^{2a} + 0.75X_i - 1.25Z_i + e_i) < 0$$

$$(9) \quad y_i^1 = 1, y_i^2 = 0$$

$$\text{Response}(yes, no)_i = 1 \quad \text{if} \quad \{(5 - 0.5b_i^1 + 0.75X_i - 1.25Z_i + e_i) \geq 0 \text{ and}$$

$$(5 - 0.5b_i^{2b} + 0.75X_i - 1.25Z_i + e_i) < 0\}$$

$$(10) \quad y_i^1 = 0, y_i^2 = 1$$

$$\text{Response}(no, yes)_i = 1 \quad \text{if} \quad \{(5 - 0.5b_i^1 + 0.75X_i - 1.25Z_i + e_i) < 0 \text{ and}$$

$$(5 - 0.5b_i^{2a} + 0.75X_i - 1.25Z_i + e_i) \geq 0\}$$

where the initial bid value  $b_i^1$  is continuous and uniformly distributed from 0 to 5, the follow-up bid value to an initial *no* response  $b_i^{2a}$  is uniformly distributed from 0 to 2.5 (half the initial bid), the follow-up bid value to an initial *yes* response  $b_i^{2b}$  is uniformly distributed from 0 to 10 (twice the initial bid), independent factors  $X_i$  and  $Z_i$  are uniformly distributed from 0 to 5, and finally the stochastic error term  $e_i$  follows a standard normal distribution. In essence, the *WTP* for the  $i$ th individual corresponding to framework of (7)–(10) is

$$(11) \quad WTP_i = 10 + 1.5X_i - 1.25Z_i + u_i$$

where  $u_i$  is the stochastic error following a normal distribution.

Three nonresponse scenarios were explored using this simulation: 1) random nonresponse; 2) systematic nonresponse in observations with upper  $X_i$  values; and 3) systematic nonresponse in observations with upper  $Z_i$  values. Estimation was carried out following Hanemann, Loomis, and Kanninen (1991) with a sample size of 1000 observations. The simulation was repeated 5000 times for each scenario at 10%, 25%, and 50% nonresponse rates.

The basic idea of the recoding procedure is to preserve the observation by shifting the bid interval and substituting a response for the missing value. More specifically, if the respondent answered *yes* to the initial question at the bid value  $b_i^1$  but did not answer the follow-up question, we can deduce the respondent would have also answered *yes* to  $b_i^{2a} \ll b_i^1$ . Therefore, we can replace the missing follow-up response and bid  $b_i^2$  with a *yes* response and bid  $b_i^1$ , while also replacing the initial bid  $b_i^1$  with  $b_i^{2a} \ll b_i^1$  and keeping the initial response *yes*. If the respondent answered *no* to the initial question with bid  $b_i^1$ , then we know they would have also answered *no* to a bid  $b_i^{2b} \gg b_i^1$ . In this case, we can replace the missing follow-up response and bid with a *no* response and bid  $b_i^1$ , while also replacing the initial bid with  $b_i^{2b} \gg b_i^1$  and keeping the initial response *no*. A summary of the recoding procedure is given below.

$$(12) \quad y_i^1 = 1, y_i^2 = \textit{missing}$$

Recode  $y_i^2 = 1$ , and then replace the initial bid value  $b_i^1$  with  $b_i^{2a} \ll b_i^1$  and the follow-up bid value  $b_i^2$  with the initial bid value  $b_i^1$ .

$$(13) \quad y_i^1 = 0, y_i^2 = \textit{missing}$$

Recode  $y_i^2 = 0$ , and then replace the initial bid value  $b_i^1$  with  $b_i^{2b} \gg b_i^1$  and the follow-up bid value  $b_i^2$  with the initial bid value  $b_i^1$ .

The root mean squared error (RMSE) was calculated across the replications and used as a measure of statistical efficiency, and was compared between the list-wise deletion case and for when the recoding procedure was used.

## 4.4. Results

### 4.4.1. *Random Nonresponse*

Table 4.1 shows the Monte Carlo simulation results with random nonresponse in the DB follow-up response. An increase in random nonresponse led to larger RMSE values for both recoding and list-wise deletion, but the log difference values (last column) showed that the RMSE values were consistently larger in the list-wise deletion case. In particular, RMSE values for  $\beta_0$ ,  $\beta_X$ ,  $\beta_Z$ ,  $\sigma$ , and  $WTP$  were roughly 12, 13, 11, 12, and 6 percent larger in the list-wise deletion case compared to recoding when half of the observations (500) were randomly assigned a missing follow-up response. There was little evidence of a trend in the estimates across the different nonresponse levels which can be used to assess bias.

### 4.4.2. *Systematic Nonresponse in Upper X*

In table 4.2, the simulation results with systematic nonresponse in the upper  $X$  values are reported. Here, observations with  $X$  variables in upper 10, 25, and 50 percent ranges were assigned a missing follow-up response. This may be similar to the situation where respondents with higher household income levels fail or refuse to answer the follow-up question. The RMSE values with recoding and list-wise deletion were not noticeably different than in the random nonresponse



scenarios in table 4.1. Some of the RMSE values were smaller and some larger. Similar to the random nonresponse scenario, the log-difference of RMSE values showed recoding led to a lower efficiency loss compared to list-wise deletion. Again, there were no noticeable trends in the estimates to indicate bias although the *WTP* estimates with 50% nonresponse were slightly larger for both recoding and list-wise deletion cases.

#### *4.4.3. Systematic Nonresponse in Upper Z*

The results for the systematic nonresponse in upper *Z* values are reported in table 4.3. In this scenario, observations with *Z* variables in upper 10, 25, and 50 percent ranges were assigned a missing follow-up response. This variable was included to simulate the effects of the availability of substitute goods. For example, the willingness to pay for outdoor recreation may decrease with the availability of indoor recreation opportunities. Compared to the random nonresponse and systematic nonresponse in *X* values scenarios, the RMSE values for the list-wise deletion case were noticeably larger, while in the recoding case, the RMSE values were about the same. Notice also, the larger RMSE values in the list-wise deletion case led to larger values in the log-difference of RMSE. For instance, the RMSE for  $\beta_Z$  was over 100 percent larger in the list-wise deletion case compared to recoding, and the RMSE for *WTP* was 57 percent larger in the list-wise deletion case. There was a slight upward trend in the  $\beta_0$  and  $\beta_Z$  estimates in the list-wise deletion case suggesting some upward bias, while there were no noticeable trends in the recoding case.

#### **4.5. Discussion and Conclusion**

This paper introduced a simple procedure for recoding missing follow-up responses from a double-bounded dichotomous choice contingent valuation method survey. Nonresponse can lead to losses in statistical efficiency as well as biased estimates. In this paper, the effects of randomly arising nonresponse and two different types of systematically arising nonresponse scenarios were investigated. The results of Monte Carlo simulations showed the loss in statistical efficiency was lower when the recoding procedure was used compared to when list-wise deletion was used. However, it was unclear from the results whether recoding will always fare better in addressing statistical bias because the magnitudes of the bias from list-wise deletion were not substantial. One possible reason may be that the simulated willingness to pay model was relatively simple in structure in that both explanatory variables followed a uniform distribution and the sample size was fixed at 1000. Therefore, the benefits from recoding may be revealed using a more complex model with a higher potential for bias from list-wise deletion and smaller sample sizes. While there are many other approaches for addressing nonresponse, the simplicity of the recoding procedure and its application may still be preferred over other more sophisticated techniques. A future avenue of research may be to develop other nonstatistics-based approaches to addressing nonresponse in CVM survey data.

Table 4.1: Double-bounded data regression simulation with random nonresponse

No recoding, listwise deletion scenario												
With recoded follow-up nonresponse												
$\beta_0 = 10$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	10.0132	0.3330	8.8621	11.3168	0.3333	1000	10.0132	0.3330	8.8621	11.3168	0.3333	0.0
1000 (10%)	10.0114	0.3445	8.8764	11.5013	0.3447	900	10.0105	0.3477	8.8310	11.4708	0.3478	0.0090
1000 (25%)	10.0112	0.3618	8.7545	11.3531	0.3620	750	10.0105	0.3751	8.6408	11.5203	0.3752	0.0360
1000 (50%)	10.0170	0.4305	8.7200	11.5749	0.4308	500	10.0172	0.4847	8.4765	11.7950	0.4849	0.1183
$\beta_X = 1.5$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	1.5027	0.0823	1.2405	1.8212	0.0824	1000	1.5027	0.0823	1.2405	1.8212	0.0824	0.0
1000 (10%)	1.5030	0.0844	1.2234	1.8412	0.0844	900	1.5031	0.0863	1.1973	1.8074	0.0863	0.0226
1000 (25%)	1.5033	0.0879	1.2103	1.8472	0.0879	750	1.5036	0.0936	1.1796	1.8569	0.0936	0.0630
1000 (50%)	1.5036	0.0997	1.1983	1.8440	0.0998	500	1.5038	0.1138	1.1556	1.9617	0.1138	0.1314
$\beta_Z = -2.5$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	-2.5057	0.0996	-2.8746	-2.1857	0.0997	1000	-2.5057	0.0996	-2.8746	-2.1857	0.0997	0.0
1000 (10%)	-2.5053	0.1027	-2.8713	-2.1752	0.1028	900	-2.5051	0.1037	-2.8788	-2.1820	0.1038	0.0102
1000 (25%)	-2.5053	0.1081	-2.8987	-2.1530	0.1083	750	-2.5055	0.1124	-2.9942	-2.1422	0.1125	0.0385
1000 (50%)	-2.5070	0.1297	-3.0223	-2.1219	0.1299	500	-2.5077	0.1446	-3.0710	-2.0339	0.1448	0.1086

Table 4.1: Double-bounded data regression simulation with random nonresponse continued

No recoding, listwise deletion scenario												
With recoded follow-up nonresponse												
$\sigma = 2$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	1.9933	0.0971	1.6773	2.3817	0.0973	1000	1.9933	0.0971	1.6773	2.3817	0.0973	0.0
1000 (10%)	1.9935	0.1006	1.6850	2.4010	0.1008	900	1.9931	0.1020	1.6642	2.4222	0.1022	0.0137
1000 (25%)	1.9932	0.1059	1.6796	2.4264	0.1062	750	1.9913	0.1107	1.6431	2.4099	0.1111	0.0451
1000 (50%)	1.9914	0.1215	1.5732	2.4428	0.1218	500	1.9863	0.1366	1.5555	2.5276	0.1372	0.1193

$WTP = 7.5$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	7.5058	0.1518	7.0341	8.2165	0.1519	1000	7.5058	0.1518	7.0341	8.2165	0.1519	0.0
1000 (10%)	7.5057	0.1577	6.9578	8.1384	0.1578	900	7.5054	0.1587	6.9637	8.1302	0.1588	0.0063
1000 (25%)	7.5061	0.1669	6.9643	8.2292	0.1670	750	7.5058	0.1699	6.9788	8.2163	0.1700	0.0178
1000 (50%)	7.5083	0.2082	6.8490	8.4244	0.2084	500	7.5073	0.2204	6.7687	8.4698	0.2205	0.0567

Note: RMSE<sub>R</sub> and RMSE<sub>NR</sub> are the estimated root mean squared errors using the recoded and not recoded data.

$\Pr(b^1 < WTP < b^2 \mid X, Z, Bid) = \alpha_0 + \alpha_X X + \alpha_Z Z + \delta Bid + e = 5 + .75X - 1.25Z - .5Bid + e$ , where  $e \sim N(0,1)$

Given the relations and true values:

$\beta_0 = -\alpha_0/\delta = -1/-5 = 2$ ;  $\beta_X = -\alpha_X/\delta = -75/-5 = 15$ ;  $\beta_Z = -\alpha_Z/\delta = 1.25/-5 = -2.5$ ;  $\sigma = -1/\delta = 2$ ;  $X, Z \sim U(2.5, 2.08)$ , we have

$WTP = \beta_0 + \beta_X X + \beta_Z Z + u = 10 + 1.5^*(2.5) - 2.5^*(2.5) + u = 7.5 + u$ , where  $u \sim N(0, \sigma^2)$

RMSE =  $[(1/r)\Sigma(V_r - V)^2]^{0.5}$ , where  $r$  = number of replications,  $V$  = random variable,  $V$  = constant representing the true value of  $V$ . The number of replications  $r = 5000$  in the simulations reported above.

Table 4-2: Double-bounded data regression simulation with systematic nonresponse in upper  $X$

No recoding, listwise deletion scenario												
With recoded follow-up nonresponse												
$\beta_0 = 10$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	$\ln(\text{RMSE}_{NR}/\text{RMSE}_R)$
1000 (0%)	10.0132	0.3330	8.8621	11.3168	0.3333	1000	10.0132	0.3330	8.8621	11.3168	0.3333	0.0
1000 (10%)	10.0140	0.3362	8.8322	11.3570	0.3364	900	10.0143	0.341	8.8122	11.353	0.3412	0.0141
1000 (25%)	10.0146	0.3467	8.8442	11.413	0.3470	750	10.0150	0.3637	8.6836	11.454	0.3640	0.0478
1000 (50%)	10.0141	0.3674	8.8251	11.4503	0.3676	500	10.0194	0.4179	8.5898	11.5898	0.4183	0.1292
$\beta_x = 1.5$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	$\ln(\text{RMSE}_{NR}/\text{RMSE}_R)$
1000 (0%)	1.5027	0.0823	1.2405	1.8212	0.0824	1000	1.5027	0.0823	1.2405	1.8212	0.0824	0.0
1000 (10%)	1.5024	0.0895	1.2192	1.8303	0.0895	900	1.5023	0.0934	1.1698	1.8655	0.0934	0.0428
1000 (25%)	1.5042	0.1024	1.1657	1.8836	0.1025	750	1.5042	0.1174	1.0645	1.9084	0.1175	0.1368
1000 (50%)	1.5055	0.1199	1.1390	2.0354	0.1200	500	1.5020	0.1945	0.8152	2.3126	0.1945	0.4824
$\beta_z = -2.5$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	$\ln(\text{RMSE}_{NR}/\text{RMSE}_R)$
1000 (0%)	-2.5057	0.0996	-2.8746	-2.1857	0.0997	1000	-2.5057	0.0996	-2.8746	-2.1857	0.0997	0.0
1000 (10%)	-2.5058	0.1008	-2.8974	-2.1998	0.1010	900	-2.5059	0.1011	-2.8971	-2.1957	0.1013	0.0030
1000 (25%)	-2.5066	0.1062	-2.9026	-2.1718	0.1064	750	-2.5067	0.1075	-2.9155	-2.1595	0.1077	0.0122
1000 (50%)	-2.5069	0.1183	-3.0238	-2.1017	0.1185	500	-2.5074	0.1226	-2.967	-2.12	0.1228	0.0354

Table 4-2: Double-bounded data regression simulation with systematic nonresponse in upper  $X$  continued

With recoded follow-up nonresponse												
No recoding, listwise deletion scenario												
$\sigma = 2$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	1.9933	0.0971	1.6773	2.3817	0.0973	1000	1.9933	0.0971	1.6773	2.3817	0.0973	0.0
1000 (10%)	1.9926	0.0997	1.6828	2.3700	0.0999	900	1.9925	0.1002	1.6791	2.3752	0.1005	0.0053
1000 (25%)	1.9926	0.1045	1.6728	2.3899	0.1047	750	1.9920	0.1063	1.6514	2.4118	0.1066	0.0181
1000 (50%)	1.9917	0.1192	1.5900	2.4314	0.1195	500	1.9895	0.1262	1.6184	2.4810	0.1267	0.0584

$WTP = 7.5$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	7.5058	0.1518	7.0341	8.2165	0.1519	1000	7.5058	0.1518	7.0341	8.2165	0.1519	0.0
1000 (10%)	7.5055	0.1603	7.0188	8.1654	0.1603	900	7.5056	0.1624	7.0086	8.1703	0.1625	0.0134
1000 (25%)	7.5084	0.1813	6.9251	8.2174	0.1815	750	7.5086	0.1939	6.8777	8.2370	0.1941	0.0668
1000 (50%)	7.5107	0.2224	6.8184	8.4171	0.2227	500	7.5058	0.3188	6.5365	8.7883	0.3188	0.3589

Note: RMSE<sub>R</sub> and RMSE<sub>NR</sub> are the estimated root mean squared errors using the recoded and not recoded data.

$\Pr(b^1 < WTP < b^2 \mid X, Z, Bid) = \alpha_0 + \alpha_x X + \alpha_z Z + \delta Bid + e = 5 + .75X - 1.25Z - .5Bid + e$ , where  $e \sim N(0,1)$

Given the relations and true values:

$\beta_0 = -\alpha_0/\delta = -1/- .5 = 2$ ;  $\beta_x = -\alpha_x/\delta = -.75/- .5 = 1.5$ ;  $\beta_z = -\alpha_z/\delta = 1.25/- .5 = -2.5$ ;  $\sigma = -1/\delta = 2$ ;  $X, Z \sim U(2.5, 2.08)$ , we have

$WTP = \beta_0 + \beta_x X + \beta_z Z + u = 10 + 1.5*(2.5) - 2.5*(2.5) + u = 7.5 + u$ , where  $u \sim N(0, \sigma^2)$

RMSE =  $[(1/r)\Sigma(V_r - V)^2]^{0.5}$ , where  $r$  = number of replications,  $V$  = random variable,  $V$  = constant representing the true value of  $V$ . The number of replications  $r = 5000$  in the simulations reported above.

Table 4.3: Double-bounded data regression simulation with systematic nonresponse in upper Z

No recoding, listwise deletion scenario												
With recoded follow-up nonresponse												
$\beta_o = 10$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	10.0132	0.3330	8.8621	11.3168	0.3333	1000	10.0132	0.3330	8.8621	11.3168	0.3333	0.0
1000 (10%)	10.0144	0.3448	8.8559	11.3368	0.3450	900	10.0144	0.3688	8.7149	11.3563	0.3690	0.0673
1000 (25%)	10.0141	0.3508	8.8513	11.3085	0.3510	750	10.0195	0.4235	8.5821	11.6642	0.4239	0.1886
1000 (50%)	10.0176	0.3652	8.7743	11.7173	0.3656	500	10.0331	0.6203	8.0159	12.7175	0.6211	0.5299
$\beta_x = 1.5$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	1.5027	0.0823	1.2405	1.8212	0.0824	1000	1.5027	0.0823	1.2405	1.8212	0.0824	0.0
1000 (10%)	1.5027	0.0853	1.2341	1.8187	0.0853	900	1.5034	0.0894	1.2108	1.8107	0.0894	0.0471
1000 (25%)	1.5032	0.0936	1.2261	1.8500	0.0937	750	1.5032	0.1090	1.1378	1.9206	0.1090	0.1518
1000 (50%)	1.5057	0.1148	1.1069	1.9994	0.1149	500	1.5085	0.1813	0.9437	2.3010	0.1815	0.4572
$\beta_z = -2.5$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	-2.5057	0.0996	-2.8746	-2.1857	0.0997	1000	-2.5057	0.0996	-2.8746	-2.1857	0.0997	0.0
1000 (10%)	-2.5062	0.1050	-2.9115	-2.1865	0.1052	900	-2.5066	0.1160	-2.9206	-2.1097	0.1161	0.0991
1000 (25%)	-2.5068	0.1097	-2.9934	-2.1339	0.1099	750	-2.5089	0.1520	-3.1653	-2.0132	0.1522	0.3255
1000 (50%)	-2.5084	0.1178	-3.1301	-2.0998	0.1181	500	-2.5197	0.3220	-4.0006	-1.5017	0.3226	1.0052

Table 4.3: Double-bounded data regression simulation with systematic nonresponse in upper  $Z$  continued

With recoded follow-up nonresponse												
No recoding, listwise deletion scenario												
$\sigma = 2$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	1.9933	0.0971	1.6773	2.3817	0.0973	1000	1.9933	0.0971	1.6773	2.3817	0.0973	0.0
1000 (10%)	1.9925	0.1017	1.6687	2.3626	0.1020	900	1.9914	0.1048	1.6305	2.4027	0.1051	0.0301
1000 (25%)	1.9906	0.1096	1.6197	2.3805	0.1100	750	1.9867	0.1208	1.5477	2.4289	0.1215	0.0991
1000 (50%)	1.9906	0.1431	1.4443	2.6397	0.1434	500	1.9699	0.1948	1.2373	2.7909	0.1970	0.3181

$WTP = 7.5$												
Obs (% recoded)	Mean	Std. Dev.	Min	Max	RMSE <sub>R</sub>	Obs	Mean	Std. Dev.	Min	Max	RMSE <sub>NR</sub>	ln(RMSE <sub>NR</sub> /RMSE <sub>R</sub> )
1000 (0%)	7.5058	0.1518	7.0341	8.2165	0.1519	1000	7.5058	0.1518	7.0341	8.2165	0.1519	0.0
1000 (10%)	7.5057	0.1541	7.0162	8.2371	0.1542	900	7.5063	0.1562	7.0078	8.2722	0.1563	0.0135
1000 (25%)	7.5052	0.1606	6.9923	8.2602	0.1607	750	7.5053	0.1700	6.9765	8.1931	0.1701	0.0566
1000 (50%)	7.5109	0.2034	6.8658	8.6867	0.2036	500	7.5050	0.3601	6.3782	9.1896	0.3601	0.5700

Note: RMSE<sub>R</sub> and RMSE<sub>NR</sub> are the estimated root mean squared errors using the recoded and not recoded data.

$\Pr(b^1 < WTP < b^2 \mid X, Z, Bid) = \alpha_0 + \alpha_X X + \alpha_Z Z + \delta Bid + e = 5 + .75X - 1.25Z - .5Bid + e$ , where  $e \sim N(0,1)$

Given the relations and true values:

$\beta_0 = -\alpha_0/\delta = -1/-5 = 2$ ;  $\beta_X = -\alpha_X/\delta = -.75/-5 = 1.5$ ;  $\beta_Z = -\alpha_Z/\delta = 1.25/-5 = -2.5$ ;  $\sigma = -1/\delta = 2$ ;  $X, Z \sim U(2.5, 2.08)$ , we have

$WTP = \beta_0 + \beta_X X + \beta_Z Z + u = 10 + 1.5*(2.5) - 2.5*(2.5) + u = 7.5 + u$ , where  $u \sim N(0, \sigma^2)$

RMSE =  $[(1/r)\Sigma(V_r - V)^2]^{0.5}$ , where  $r$  = number of replications,  $V$  = random variable,  $V$  = constant representing the true value of  $V$ . The number of replications  $r = 5000$  in the simulations reported above.



## **CHAPTER 5: Conclusion**

The aim of this dissertation was to provide a better understanding of the different modeling approaches available for combining travel cost method (TCM) and contingent valuation method (CVM) data within the context of wildlife watching recreation. Emphasis was placed on the methodological aspects as well as in obtaining theoretically consistent measures of the welfare effects from recreation. In addition, this dissertation sought to address the issue of nonresponse which nearly every researcher employing survey data has faced at one time or another. These approaches were demonstrated using data from a national survey as well as data that were computer simulated.

In chapter 2, a finite mixture (FM) model was applied to wildlife watching recreation data from the 2006 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (FHWAR). The results provided strong evidence in support of heterogeneous preferences in the wildlife watching population, which has important implications for sustainable tourism, and wildlife and land management. The analysis revealed there were three subgroups or user classes of wildlife watchers that can be characterized as: local enthusiasts, traveling enthusiasts, and average participants. These user classes exhibited very different visitation and spending patterns, and derived different levels of benefits from their activities. However, post estimation analysis revealed further segmentation was possible that differentiated within the user classes by skill and experience levels. There were also indications that wildlife watchers may consciously reside

closer to their frequented sites in order to reduce their travel costs, and their economic activities may differ when participating locally compared to when they travel. Another unique feature of this model was that CVM choke price data were used to fill-in missing travel cost values for respondents who had not taken any trips. In particular, the predicted mean matching (PMM) multiple imputation (MI) method was used to carry out the procedure of imputing these values which is entirely unique to this study. The advantages of using MI in recreation demand analysis was demonstrated by comparing its performance to that of a zero-truncated FM model. Goodness of fit measures showed the MI FM model performed much better than the zero-truncated FM model.

The use of choke price data in the calculation of consumer surplus revealed some interesting discrepancies that may be an avenue for further study. However, it was mentioned that an important consideration in regards to the results is the difficulty in obtaining credible open-ended CVM responses as previous research has shown. Therefore, a logical extension to this work might investigate the mechanisms people use in deriving their choke prices to develop questions that are less susceptible to these issues. Another extension that was suggested was to explore the performance of models that use MI in other types of CVM data.

Chapter 3 estimated a utility-consistent joint model of wildlife watching recreation using TCM count data and CVM dichotomous choice data from the 1996 FHWAR survey. The chapter advocated for better cohesion between the two types of information by carefully adjusting the language of the CVM questions to match the TCM framework. The chapter pointed out certain types of CVM

scenarios such as the “all or nothing” scenario, can lead to the TCM consumer surplus being equal to the CVM compensating variation measure. However, it was shown that this equivalence needed to be explicitly incorporated into the modeling framework to hold. Thus, the novelty of the utility-consistent model presented in this chapter was that the estimates were consistent across the welfare measures. Although, it is still unclear how the statistical performance of the utility-consistent joint model fared against other joint modeling frameworks, its adoption may be supported purely from a conceptual standpoint. Overall, the argument for combining TCM and CVM data was largely supported. One avenue for future research may be to investigate the policy implications of using a welfare measure that can be interpreted in terms of net benefits as well as willingness to pay, as the only obvious advantage seems to be a more simplified welfare measure for decision making. Finally, while the focus of this chapter was on combining TCM and CVM data, expanding the integration exercise to other data types such as the hedonic price model and choice experiment survey data, could prove fruitful.

The final chapter of this dissertation focused on the issue of nonresponse in double-bounded dichotomous choice (DB) CVM data. The chapter introduced a simple procedure for recoding missing follow-up responses in DB data that can attenuate the loss in statistical efficiency and bias. In particular, the analysis examined scenarios depicting a randomly arising nonresponse and two types of systematically arising nonresponse. Using Monte Carlo simulations, the root mean squared error (RMSE) estimates from recoding was compared to that of when list-wise deletion was used. Comparisons of the RMSE in the three

nonresponse scenarios showed the recoding procedure was favored over using list-wise deletion in all cases. It was unclear whether recoding will always fare better in addressing statistical bias because the magnitudes of the biases from list-wise deletion were not substantial. One possible reason that was mentioned was that the simulated willingness to pay (WTP) model was relatively simple in that the two explanatory variables followed a uniform distribution and the sample size was fixed at 1000. Therefore, the benefits from recoding may be revealed using a more complex model with a higher potential for bias from list-wise deletion and smaller sample sizes. When applicable however, the simplicity of the recoding procedure and its application may make it more attractive than other more sophisticated approaches such as the MI method. The chapter concluded by recommending further research on developing other nonstatistics-based techniques to addressing nonresponse in CVM survey data.

It is my greatest wish that this dissertation can serve as a useful reference for those interested in combining disparate types of nonmarket valuation data as well as providing options for how to remedy issues of nonresponse. I also hope that this dissertation can incite continued research on the value of wildlife watching recreation. Wildlife watching recreation is increasingly becoming the most popular form of outdoor recreation in the U.S. but current valuation models still have not connected the demand for wildlife watching recreation to the benefits of conservation. Future research may try to couple these two topics into a more comprehensive model that can shed light on how humans can coexist more peacefully with the natural environment.

## **APPENDICES**

Appendix A: Contingency Table of Average Posterior Probabilities.....	109
Appendix B: Table of Average Information Criterion and Entropy Values for Truncated FM Models.....	110
Appendix C: Zero-truncated Finite Mixture Count Model.....	111
Appendix D: Derivation of Expected Consumer Surplus with Finite Upper Limit.....	113
Appendix E: Variance Function of Expected Consumer Surplus with Finite Upper Limit.....	114
Appendix F: Goodness of Fit Measures for Count Data Models.....	116
Appendix G: Joint Estimation of TCM and CVM Model Parameters.....	117
Appendix H: Stata Codes.....	119

## Appendix A: Contingency table of average posterior probabilities

---

MI 3-Component FM Model			
	Component 1	Component 2	Component 3
prob[ $i \in 1$ ]	0.850	0.018	0.132
prob[ $i \in 2$ ]	0.029	0.828	0.143
prob[ $i \in 3$ ]	0.110	0.128	0.761

---

**Appendix B: Average information criterion and entropy values for truncated FM models**

Components	$n$	$\ln L$	$p$	AIC	BIC
1	12224	-7590.72	16	15247.44	15453.19
2	12224	-7405.42	33	14876.85	15082.61
3	12224	-7333.15	50	14765.29	14938.05

Note: Values are averages taken over MI estimates.  $p$  is the number of parameters.

## Appendix C: Zero-truncated Finite Mixture Count Model

Following Lourenço, Ferreira, and Barros (2006), a  $G$ -component finite mixture model with vector of covariates  $x_i$  and respective parameters  $\varphi_g$ , can be written in the form

$$(C1) \quad f(y_i; \Phi) = \sum_{g=1}^G \pi_g f_g(y_i; \varphi_g)$$

, where  $f_g(\cdot)$  is the density function for component  $g$ , and  $\pi_g$  is a mixing probability with  $\sum \pi_g = 1$ , and  $0 \leq \pi_g \leq 1$ . The probability that the outcome variable  $y_i$  is greater than zero is

$$(C2) \quad P(y_i > 0; \Phi) = \sum_{g=1}^G \pi_g P_g(y_i > 0; \varphi_g) = 1 - \sum_{g=1}^G \pi_g P_g(y_i = 0; \varphi_g)$$

Therefore, the density of the truncated sample can be expressed as

$$(C3) \quad f(y_i > 0; \Phi) = \frac{\sum_{g=1}^G \pi_g f_g(y_i; \varphi_g)}{1 - \sum_{g=1}^G \pi_g P_g(y_i = 0; \varphi_g)}$$

In Lourenço, Ferreira, and Barros (2006), the authors reexpress (C3) as



$$\begin{aligned}
f(y_i > 0; \Phi) &= \sum_{g=1}^G \left\{ \frac{f_g(y_i; \varphi_g)(1 - P_g(y_i = 0; \varphi_g))}{(1 - \sum_{g=1}^G \pi_g P_g(y_i = 0; \varphi_g))(1 - P_g(y_i = 0; \varphi_g))} \right\} \pi_g \\
&= \sum_{g=1}^G \left\{ \frac{f_g(y_i; \varphi_g)}{1 - P_g(y_i = 0; \varphi_g)} \right\} \tilde{\pi}_g
\end{aligned}$$

, where  $\tilde{\pi}_g$  is a mixing probability in the truncated population and is given by

$$\tilde{\pi}_g = \frac{1 - P_g(y_i = 0; \varphi_g)}{1 - \sum_{g=1}^G \pi_g P_g(y_i = 0; \varphi_g)} \pi_g$$

## Appendix D: Derivation of Expected Consumer Surplus with Finite Upper Limit

Let  $\hat{\mu} = n^{-1} \sum \exp(x'_i \hat{\beta}) = n^{-1} \sum \exp(x'_i \hat{\beta} | c^0)$ , where  $c^0$  and  $c^1 > c^0$  are nonnegative constants. The expected consumer surplus is defined as

$$\begin{aligned} E[CS] &= \int_{c^0}^{c^1} n^{-1} \sum \exp(x'_i \hat{\beta}) dc \\ &= \hat{\beta}_c^{-1} \{ n^{-1} \sum (\exp(x'_i \hat{\beta} | c^1) - \exp(x'_i \hat{\beta} | c^0)) \} \\ &= \hat{\beta}_c^{-1} \{ n^{-1} \sum \exp(x'_i \hat{\beta} | c^1) - n^{-1} \sum \exp(x'_i \hat{\beta} | c^0) \} \end{aligned}$$

Rearrange the above expression after multiplying by  $\hat{\mu}/\hat{\mu}$  which gives

$$= \hat{\beta}_c^{-1} \{ \exp(\hat{\beta}_c (c^1 - c^0)) - 1 \} \cdot \hat{\mu}$$

## Appendix E: Variance Function of Expected Consumer Surplus with Finite Upper Limit

The variance function corresponding to equation (3) in chapter 2 with  $\hat{\mu}$  set to a constant was obtained using the Delta method.

$$\text{var}[CS] \cong (\hat{\mu})^2 K^2 \cdot \text{var}(\hat{\beta}_c) / (\beta_c)^4$$

, where  $K = \exp(\beta_c(c^1 - c^0)) (\beta_c(c^1 - c^0) - 1) - 1$ , and  $\text{var}(\hat{\beta}_c)$  is the estimated variance of  $\hat{\beta}_c$ . Notice  $K = -1$  when  $c^1 = \infty$ . Yen and Adamowicz (1993) argued that mathematical approximations can be problematic for inference especially when the distribution is asymmetric. The authors suggest using the simulation procedure by Krinsky and Robb (1986, 1990) to find the moment estimates. However, it is unclear as to how this can be applied in a multiple imputation framework. Therefore, for the purpose of this study, the approximations are used and the aforementioned issue to be resolved in future works. A variance function approximation to equation (5) in chapter 2 is given by

$$\begin{aligned} \text{var}[\overline{CS}] &\cong \sum_{g=1}^G \left\{ \left( \frac{\partial \overline{CS}}{\partial \pi_g} \right)^2 \cdot \text{var}(\hat{\pi}_g) + \left[ (\hat{\pi}_g) \left( \frac{\partial E[CS]_g}{\partial \beta_g} \right) \right]^2 \cdot \text{var}(\hat{\beta}_g) \right. \\ &\quad \left. + 2\hat{\pi}_g \left( \frac{\partial \overline{CS}}{\partial \pi_g} \right) \left( \frac{\partial E[CS]_g}{\partial \beta_g} \right) \cdot \text{cov}(\hat{\pi}_g, \hat{\beta}_g) \right\} \\ &\cong \sum_{g=1}^G \left\{ (E[CS]_g)^2 \cdot \text{var}(\hat{\pi}_g) + [(\hat{\pi}_g)(K_g / (\beta_g)^2)]^2 \cdot \text{var}(\hat{\beta}_g) \right. \\ &\quad \left. + 2(\hat{\pi}_g E[CS]_g K_g / (\beta_g)^2) \cdot \text{cov}(\hat{\pi}_g, \hat{\beta}_g) \right\} \end{aligned}$$

, where  $\text{cov}(\hat{\pi}_g, \hat{\beta}_g)$  is the estimated covariance between  $\hat{\pi}_g$  and  $\hat{\beta}_g$ . For simplicity, we assumed  $\text{var}(\hat{\pi}_g) = 0$  and  $\text{cov}(\hat{\pi}_g, \hat{\beta}_g) = 0$  which gives

$$\text{var}[\overline{CS}] \cong \sum_{g=1}^G \hat{\pi}_g \text{var}[CS]_g$$

## Appendix F: Goodness of Fit Measures for Count Data Models

Several  $R^2$  measures discussed in Cameron and Windmeijer (1996) were used to assess the models' goodness of fit. The following are the formulas for these measures.

$$\text{Residual Sum of Squares} = 1 - \frac{\sum_i (y_i - \hat{\xi}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

$$\text{Explained Sum of Squares} = 1 - \frac{\sum_i (y_i - \hat{\xi}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

$$\text{Squared Sample Correlation Coefficient} = \frac{[\sum_i (y_i - \bar{y})(\hat{\xi}_i - \bar{\xi})]^2}{\sum_i (y_i - \bar{y})^2 \sum_i (\hat{\xi}_i - \bar{\xi})^2}$$

$$\text{Pearson Residuals} = 1 - \frac{\sum_i [(y_i - \hat{\xi}_i)^2 / (\hat{\xi}_i + \hat{\alpha} \hat{\xi}_i^2)]}{\sum_i [(y_i - \bar{y})^2 / (\bar{y} + \hat{\alpha} \bar{y}^2)]}$$

$$\text{Deviance Residual (DR)} = 1 - \frac{\sum_i [y_i \log(y_i / \hat{\xi}_i) - (y_i + \hat{\alpha}^{-1}) \log((y_i + \hat{\alpha}^{-1}) / (\hat{\xi}_i + \hat{\alpha}^{-1}))]}{\sum_i [y_i \log(y_i / \bar{y}) - (y_i + \hat{\alpha}^{-1}) \log((y_i + \hat{\alpha}^{-1}) / (\bar{y} + \hat{\alpha}^{-1}))]}$$

$$\text{DR Value} = \sum_i 2[y_i \log(y_i / \hat{\xi}_i) - (y_i + \hat{\alpha}^{-1}) \log((y_i + \hat{\alpha}^{-1}) / (\hat{\xi}_i + \hat{\alpha}^{-1}))]$$

, where  $\hat{\xi}_i = \sum_g \hat{\pi}_{gi} \hat{\mu}_{gi}$ , and  $\hat{\mu}_{gi}$  is the component-level predicted value of the count variable  $y_i$ .

## Appendix G: Joint Estimation of TCM and CVM Model Parameters

Parameter estimates of the joint TCM-CVM model is found by maximizing the log-likelihood function given by

$$\begin{aligned} \mathcal{L} = & \sum_{i=1}^n \{y_{cvm,i} \ln[\Phi(W_i)] + (1 - y_{cvm,i}) \ln[1 - \Phi(W_i)]\} \\ & + \sum_{i=1}^n \left\{ y_{tcm,i} \ln \left( \frac{\alpha \exp(\mathbf{x}'_i \theta)}{1 + \alpha \exp(\mathbf{x}'_i \theta)} \right) - \left( \frac{1}{\alpha} \right) \ln(1 + \alpha \exp(\mathbf{x}'_i \theta)) + \ln \Gamma \left( y_{tcm,i} + \frac{1}{\alpha} \right) \right. \\ & \left. - \ln \Gamma(y_{tcm,i} + 1) - \ln \Gamma \left( \frac{1}{\alpha} \right) - \ln[1 - (1 + \alpha \exp(\mathbf{x}'_i \theta))^{-1/\alpha}] \right\} \end{aligned}$$

, where  $y_{cvm,i}$  is the CVM dichotomous response variable,  $\Phi(\cdot)$  is the standard normal cumulative distribution function,  $y_{tcm,i}$  is the TCM count variable that follows a zero-truncated negative binomial distribution with mean function  $\exp(\mathbf{x}'_i \theta) = \exp(\alpha + \beta p_i + \delta y_i + \mathbf{z}'_i \xi)$ , where  $\mathbf{x}_i$  is a vector of covariates and  $\theta$  is a vector of model parameters.

$$W_i = \left( \left( \frac{\mathbf{w}'_i \gamma - B_i}{\sigma} \right) + \rho Z_{tcm,i} \right) / \sqrt{(1 - \rho^2)}$$

, where  $\mathbf{w}'_i \gamma$  is the set of CVM regressors,  $B_i$  is the bid value,  $\sigma > 0$  is the standard normal standard deviation, and  $\rho$  is the correlation coefficient. Also,  $Z_{tcm,i}$  is a standardized fitted error in the demand function (Cameron, 1992) defined as

$Z_{tcm,i} = (y_{tcm,i} - E(y_{tcm,i}))/\sqrt{\text{var}(y_{tcm,i})}$ , with conditional expected value function

$$E(y_{tcm,i}) = \frac{\exp(\mathbf{x}'_i\theta)}{1 - (1 + \alpha \exp(\mathbf{x}'_i\theta))^{-1/\alpha}}$$

, and the conditional variance function as shown in Grogger and Carson (1991) is

$$\text{var}(y_{tcm,i}) = \frac{E(y_{tcm,i})}{(1 + \alpha \exp(\mathbf{x}'_i\theta))^{-1}} \{1 - [(1 + \alpha \exp(\mathbf{x}'_i\theta))^{-(1+\alpha)/\alpha}]E(y_{tcm,i})\}$$

## Appendix H: Stata Codes

### Chapter 1 Stata Codes

```
//=====FILE 1=====//  
  
//This file prepares the data from the screening survey (FH2) for merging with  
the data from the detailed wildlife watching survey(FH4)./  
  
//Screening file//  
use "I:\fh2.dta"  
  
//Prepare data for merging with FH4//  
replace everhunt=0 if everhunt==.  
replace everfish=0 if everfish==.  
replace observe05=0 if observe05==.  
replace feed05=0 if feed05==.  
replace photo05=0 if photo05==.  
replace plant05=0 if plant05==.  
replace interest06=0 if interest==.  
replace trip05=0 if trip05==.  
replace tripday05=0 if tripday05==.  
replace tripspend05=0 if tripspend05==.  
replace trip06=0 if trip06==.  
replace hunt05=0 if (hunt05==. & everhunt==0)  
replace huntday05=0 if hunt05==0  
replace huntspend05=0 if hunt05==0  
replace hunt06=0 if everhunt==0  
replace fish05=0 if everfish==0  
replace fishday05=0 if fish05==0  
replace fishspend05=0 if fish05==0  
replace fish06=0 if everfish==0  
gen retired=retire==1  
replace retire=5 if (job==1 & retire==.)  
ren hincome hincome_fh2 //renamed to compare with fh4 and use as substitute  
in case of nonresponse//  
ren i_resident i_resident_fh2  
  
sum i_resident_fh2 relation job retire hunt05 huntfirst05 huntday05 ///  
huntspend05 hunt06 huntyear huntlike06 fish05 fishfirst05 fishday05 ///  
fishspend05 fish06 fishyear fishlike06 intlike06 tripday05 tripspend05 ///  
triplike06 hincome_fh2 everhunt everfish observe05 feed05 photo05 plant05 ///  
interest06 trip05 trip06 perwgt personid id retired  
  
//Reducing data set to utilized variables//
```



```

keep i_resident_fh2 relation job retire hunt05 huntfirst05 huntday05 ///
huntspend05 hunt06 huntyear huntlike06 fish05 fishfirst05 fishday05 ///
fishspend05 fish06 fishyear fishlike06 intlike06 tripday05 tripspend05 ///
triplike06 hincome_fh2 everhunt everfish observe05 feed05 photo05 plant05 ///
interest06 trip05 trip06 perwgt personid id retired

```

```

//Saving a thinned version of the data set to be used for merging with detailed
wildlife watching survey data FH4//
save "I:\fh2_thin.dta", replace

```

```

//=====FILE 2=====//

```

```

//This file was used primarily to merge the detailed wildlife watching sample file
FH4 with the screening file FH2 which were prepared in FILE 1.//
//Other data that includes state level variables and GIS shape file data are
included in the merge also.//

```

```

clear

```

```

use "I:\fh4.dta" //Detailed Wildlife Watching sample file//

```

```

merge m:1 i_resident using "I:\medinc.dta" //state median income, wildlife
watching sites, number of species//

```

```

drop _merge

```

```

merge m:1 i_resident using "I:\usdb.dta" //us states and territories GIS shape
file//

```

```

drop if personid==.

```

```

drop _merge

```

```

merge m:1 personid using "I:\fh2_thin.dta" //thinned screening file//

```

```

drop if _merge==2

```

```

drop _merge

```

```

count if i_resident!=resstate //70 obs where the individual moved to a different
state between wave1 and wave3 interviews//

```

```

//Saving merged data//

```

```

save "I:\fh2_4merge2.dta", replace

```

```

//=====FILE 3=====//

```

```

//This file was used to create a data set that includes only the contingent
valuation responses from participants with positive trips.//

```

```

//The file includes additional preparations to data such as renaming, recoding,
transformations, and in some cases, simple data imputing. The user should be
aware that some//

```

```
//variables were not used in the main analysis, but rather served as instruments
for conducting diagnostics. The final data set was formatted//
//to long form.//
```

```
clear
```

```
use "I:\fh2_4merge2.dta"
```

```
//Prepare data for Analysis//
replace gemsasz=. if gemsasz==9
replace gemsast=. if gemsast==9
replace geur=. if geur==9
gen urban=(geur==1)
```

```
gen cendiv_ne=cendiv==1
gen cendiv_ma=cendiv==2
gen cendiv_enc=cendiv==3
gen cendiv_wnc=cendiv==4
gen cendiv_sa=cendiv==5
gen cendiv_esc=cendiv==6
gen cendiv_wsc=cendiv==7
gen cendiv_mt=cendiv==8
gen cendiv_pac=cendiv==9
```

```
gen agesq=age^2
gen male=1 if sex==1
replace male=0 if sex==2
gen white=(srace==1)
gen black=(srace==2)
gen natam=(srace==3)
gen asian=(srace==4)
gen pacif=(srace==5)
gen hisp=(hispanic==1)
gen marr=(marital==1)
gen divor=(marital==3)
recode school (0/12=1 "High School or less")(13/15=2 "1 to 3 yrs. College") ///
(16=3 "4 or more/Bachelor's")(17/18=4 "Some grad/prof school or degree"),
gen(educ)
```

```
//Simple imputation of missing values//
replace retire=5 if (hincome!=. & retire==.) //have a job/business//
replace retire=1 if (age>=65 & retire==.) //retired//
replace retire=4 if retire==. //doing something else//
```

```
foreach X of varlist observe photograph feed private public observe_bird
photo_bird ///
```

```

feed_bird bird_prey bird_fowl bird_water bird_song bird_other feed_fish
feed_large ///
feed_small feed_marine feed_other ///
{
replace `X`=0 if `X`=.,
}
ren moreqp5 wildbooks
ren moreqp6 wildcontr

//Resident state//
ren econadd wildtrip1
ren econcstz wildexp1
ren econncuz wildchoke1

//Non-resident state//
ren econadd2 wildtrip2
ren econcstz2 wildexp2
ren econncuz2 wildchoke2

//Environmental Indices//
egen birdspec_min=min(birdspec)
egen birdspec_max=max(birdspec)
egen wildsites_min=min(wildsites)
egen wildsites_max=max(wildsites)
gen birdspec_std=(birdspec-birdspec_min)/(birdspec_max-birdspec_min)
gen wildsites_std=(wildsites-wildsites_min)/(wildsites_max-wildsites_min)
gen env_index=birdspec_std+wildsites_std

//Household income intervals//
gen hhinc_max=9999 if hincome==1
replace hhinc_max=19999 if hincome==2
replace hhinc_max=24999 if hincome==3
replace hhinc_max=29999 if hincome==4
replace hhinc_max=34999 if hincome==5
replace hhinc_max=39999 if hincome==6
replace hhinc_max=49999 if hincome==7
replace hhinc_max=74999 if hincome==8
replace hhinc_max=99999 if hincome==9
replace hhinc_max=200000 if hincome==10
replace hhinc_max=200000 if hincome==.

gen hhinc_min=1 if hincome==1
replace hhinc_min=10000 if hincome==2
replace hhinc_min=20000 if hincome==3
replace hhinc_min=25000 if hincome==4
replace hhinc_min=30000 if hincome==5
replace hhinc_min=35000 if hincome==6

```

```

replace hhinc_min=40000 if hincome==7
replace hhinc_min=50000 if hincome==8
replace hhinc_min=75000 if hincome==9
replace hhinc_min=100000 if hincome==10
replace hhinc_min=1 if hincome==.

gen lhinc_min=ln(hhinc_min)
gen lhinc_max=ln(hhinc_max)

//Household income estimates obtained using Interval Regression//
quietly intreg lhinc_min lhinc_max age agesq male black natam asian ///
pacif hisp i.retire i.marital school urban i.cendiv, vce(robust) ///
predict lhincome_hat, e(lhinc_min, lhinc_max)

//Preparing cost variables for creating final trip cost variable(nctc):nctc= ///
airplane cost+train cost+bus cost+private vehicle+public land access fee ///
+private land access fee//

foreach X of varlist ncushr1d1 ncushr1d2 ncushr1d3 ncushr1d4 ncushr1d5 ///
ncushr1d6 ncushr1d7 ncushr1d8 ncushr1d9 ncushr1d10 ///
ncushr2d1 ncushr2d2 ncushr2d3 ncushr2d4 ncushr2d5 ncushr2d6 ///
ncushr2d7 ncushr2d8 ncushr2d9 ncushr2d10 ///
ncushr3d1 ncushr3d2 ncushr3d3 ncushr3d4 ncushr3d5 ncushr3d6 ///
ncushr3d7 ncushr3d8 ncushr3d9 ncushr3d10 ///
ncushr4d1 ncushr4d2 ncushr4d3 ncushr4d4 ncushr4d5 ncushr4d6 ///
ncushr4d7 ncushr4d8 ncushr4d9 ncushr4d10 ///
ncushr5d1 ncushr5d2 ncushr5d3 ncushr5d4 ncushr5d5 ncushr5d6 ///
ncushr5d7 ncushr5d8 ncushr5d9 ncushr5d10 ///
ncushr6d1 ncushr6d2 ncushr6d3 ncushr6d4 ncushr6d5 ncushr6d6 ///
ncushr6d7 ncushr6d8 ncushr6d9 ncushr6d10 ///
ncushr7d1 ncushr7d2 ncushr7d3 ncushr7d4 ncushr7d5 ncushr7d6 ///
ncushr7d7 ncushr7d8 ncushr7d9 ncushr7d10 ncushr8d1 ncushr8d2 ///
ncushr8d3 ncushr8d4 ncushr8d5 ncushr8d6 ncushr8d7 ncushr8d8 ///
ncushr8d9 ncushr8d10 ncushr9d1 ncushr9d2 ncushr9d3 ncushr9d4 ///
ncushr9d5 ncushr9d6 ncushr9d7 ncushr9d8 ncushr9d9 ncushr9d10 ///
ncushr10d1 ncushr10d2 ncushr10d3 ncushr10d4 ncushr10d5 ///
ncushr10d6 ncushr10d7 ncushr10d8 ncushr10d9 ncushr10d10 ///
ncushr11d1 ncushr11d2 ncushr11d3 ncushr11d4 ncushr11d5 ///
ncushr11d6 ncushr11d7 ncushr11d8 ncushr11d9 ncushr11d10 ///
ncushr12d1 ncushr12d2 ncushr12d3 ncushr12d4 ncushr12d5 ///
ncushr12d6 ncushr12d7 ncushr12d8 ncushr12d9 ncushr12d10 ///
{
replace `X'=0 if `X'==.
}

//=====//

```

```

//Format data into long form//

//These variables were generated to make the procedure feasible//
gen wildchoke3=.
gen wildchoke4=.
gen wildchoke5=.
gen wildchoke6=.
gen wildchoke7=.
gen wildchoke8=.
gen wildchoke9=.
gen wildchoke10=.
gen wildtrip3=.
gen wildtrip4=.
gen wildtrip5=.
gen wildtrip6=.
gen wildtrip7=.
gen wildtrip8=.
gen wildtrip9=.
gen wildtrip10=.
gen wildexp3=.
gen wildexp4=.
gen wildexp5=.
gen wildexp6=.
gen wildexp7=.
gen wildexp8=.
gen wildexp9=.
gen wildexp10=.

//Reshape data to long format//
reshape long wildchoke wildtrip wildexp ncu_std ncutotd ncudaysd ///
trp1_q1d trp1_q2d trp1_q3d trpday1d trpday2d trpday3d ncuprivd ///
ncupubd birds1d birds1ad birds2d birds3d typbrd1d typbrd2d ///
typbrd3d typbrd4d typbrd5d animls1d animls2d animls3d ///
animls4d animls5d ncushr1d ncushr2d ncushr3d ncushr4d ///
ncushr5d ncushr6d ncushr7d ncushr8d ncushr9d ncushr10d ///
ncushr11d ncushr12d ncutotamd, i(personid id) j(choke)

//Renaming variables//
ren ncu_std ncstate
ren ncutotd nctrips
ren ncudaysd ncdays
ren trp1_q1d ncobsrv
ren trp1_q2d ncphoto
ren trp1_q3d ncfeed
ren trpday1d obday
ren trpday2d phday
ren trpday3d feday

```

```

ren ncuprivd pvtlnd
ren ncupubd publnd
ren birds1d obbird
ren birds1ad birdday
ren birds2d phbird
ren birds3d febird
ren typbrd1d brdprey
ren typbrd2d brdwater
ren typbrd3d brdshore
ren typbrd4d brdsong
ren typbrd5d brdothor
ren animls1d opffish
ren animls2d opfbig
ren animls3d opfsml
ren animls4d opfmarine
ren animls5d opfturtle
ren ncushr1d foodcost
ren ncushr2d lodgcost
ren ncushr3d aircost
ren ncushr4d traincost
ren ncushr5d carcost
ren ncushr6d guidecost
ren ncushr7d pubfee
ren ncushr8d pvtfee
ren ncushr9d equipcost
ren ncushr10d fuelcost
ren ncushr11d boatcost
ren ncushr12d cookcost
ren ncutotamtd totrescost

```

```

//Collapsing data to only those having contingent valuation responses(Choke
price)//
drop if wildchoke==.

```

```

//Generating zeros for associated trips//
replace nctrips=0

```

```

//Creating another unique identifier to differentiate these responses from
others//
gen personid2=(((personid+id+choke)/age)+retire)/relation*choke*age

```

```

//Checking uniqueness of identifier//
duplicates list personid2

```

```

//Additional data cleaning and augmentation//
ren lhincome_hat lhinc
gen outstate=0 if (ncstate==resstate)

```

```

replace outstate=1 if (outstate==.)
gen partic=1 if nctrips>0
replace partic=0 if partic==.
gen lwildsites=ln(wildsites)
gen lbirdspec=ln(birdspec)
gen nchunt=ncu_hnt
replace nchunt=1 if (hunto6==1 & nchunt==.)
gen nchunto6=hunto6
replace nchunto6=1 if (ncu_hnt==1 & hunto6==0)
gen ncfisho6=fisho6
replace ncfisho6=1 if (ncu_fish==1 & fisho6==0)

//Dropping miscellaneous variables//
drop perstat1 linenum i_wave1 i_wave2 mode_a letter_a perstat2 letter_b ///
outcome1 outcome2 mode_b i_wave3 perstat3 letter_c outcome3 mode_c ///
Name i_resident_fh2

replace ncstate=ncustate if ncstate==""

save "I:\fh24mergelongchoke.dta", replace

```

```
//=====FILE 4=====//
```

This file begins with the same data in FILE 3, but is used primarily for generating the travel cost variable for participants with positive trips. The file includes additional preparations to data such as renaming, recoding, transformations, and in some cases, simple data imputing. The user should be aware that some variables were not used in the main analysis, but rather served as instruments for conducting diagnostics. The final data set was formatted to long form and includes all respondents including nonparticipants. The procedure however is identical to that in FILE 3 until the reshaping procedures. This file also merges the data obtained from FILE 3 with that obtained here.

```
clear
```

```
use "I:\fh2_4merge2.dta"
```

```
//Prepare data for Analysis//
replace gemsasz=. if gemsasz==9
replace gemsast=. if gemsast==9
replace geur=. if geur==9
gen urban=(geur==1)

```

```
gen cendiv_ne=cendiv==1
gen cendiv_ma=cendiv==2
gen cendiv_enc=cendiv==3

```

```

gen cendiv_wnc=cendiv==4
gen cendiv_sa=cendiv==5
gen cendiv_esc=cendiv==6
gen cendiv_wsc=cendiv==7
gen cendiv_mt=cendiv==8
gen cendiv_pac=cendiv==9

gen agesq=age^2
gen male=1 if sex==1
replace male=0 if sex==2
gen white=(srace==1)
gen black=(srace==2)
gen natam=(srace==3)
gen asian=(srace==4)
gen pacif=(srace==5)
gen hisp=(hispanic==1)
gen marr=(marital==1)
gen divor=(marital==3)
recode school (0/12=1 "High School or less")(13/15=2 "1 to 3 yrs. College") ///
(16=3 "4 or more/Bachelor's")(17/18=4 "Some grad/prof school or ///
degree"), gen(educ)

//Simple imputation of missing values//
replace retire=5 if (hincome!=. & retire==.) //have a job/business//
replace retire=1 if (age>=65 & retire==.) //retired//
replace retire=4 if retire==. //doing something else//

foreach X of varlist observe photograph feed private public observe_bird ///
photo_bird feed_bird bird_prey bird_fowl bird_water bird_song ///
bird_other feed_fish feed_large feed_small feed_marine feed_other ///
{
replace `X'=0 if `X'==.
}
ren moreqp5 wildbooks
ren moreqp6 wildcontr

//Resident state//
ren econadd wildtrip1
ren econcstz wildexp1
ren econncuz wildchoke1

//Non-resident state//
ren econadd2 wildtrip2
ren econcst2z wildexp2
ren econncu2z wildchoke2

//Environmental Indices//

```



```

egen birdspec_min=min(birdspec)
egen birdspec_max=max(birdspec)
egen wildsites_min=min(wildsites)
egen wildsites_max=max(wildsites)
gen birdspec_std=(birdspeak-birdspec_min)/(birdspeak_max-birdspec_min)
gen wildsites_std=(wildsites-wildsites_min)/(wildsites_max-wildsites_min)
gen env_index=birdspec_std+wildsites_std

```

```

/*Household income intervals*/
gen hhinc_max=9999 if hincome==1
replace hhinc_max=19999 if hincome==2
replace hhinc_max=24999 if hincome==3
replace hhinc_max=29999 if hincome==4
replace hhinc_max=34999 if hincome==5
replace hhinc_max=39999 if hincome==6
replace hhinc_max=49999 if hincome==7
replace hhinc_max=74999 if hincome==8
replace hhinc_max=99999 if hincome==9
replace hhinc_max=2000000 if hincome==10
replace hhinc_max=2000000 if hincome==.

```

```

gen hhinc_min=1 if hincome==1
replace hhinc_min=10000 if hincome==2
replace hhinc_min=20000 if hincome==3
replace hhinc_min=25000 if hincome==4
replace hhinc_min=30000 if hincome==5
replace hhinc_min=35000 if hincome==6
replace hhinc_min=40000 if hincome==7
replace hhinc_min=50000 if hincome==8
replace hhinc_min=75000 if hincome==9
replace hhinc_min=100000 if hincome==10
replace hhinc_min=1 if hincome==.

```

```

gen lhhinc_min=ln(hhinc_min)
gen lhhinc_max=ln(hhinc_max)

```

```

//Household income estimates obtained using Interval Regression//
quietly intreg lhhinc_min lhhinc_max age agesq male black natam asian ///
pacif hisp i.retire i.marital school urban i.cendiv, vce(robust) predict ///
lhincome_hat, e(lhhinc_min, lhhinc_max)

```

```

//Preparing cost variables for creating final trip cost variable(nctc):nctc=
airplane cost+train cost+bus cost+private vehicle+public land access fee+private
land access fee//

```

```

foreach X of varlist ncushr1d1 ncushr1d2 ncushr1d3 ncushr1d4 ///
ncushr1d5 ncushr1d6 ncushr1d7 ncushr1d8 ncushr1d9 ncushr1d10 ///
ncushr2d1 ncushr2d2 ncushr2d3 ncushr2d4 ncushr2d5 ncushr2d6 ///
ncushr2d7 ncushr2d8 ncushr2d9 ncushr2d10 ncushr3d1 ncushr3d2 ///
ncushr3d3 ncushr3d4 ncushr3d5 ncushr3d6 ncushr3d7 ncushr3d8 ///
ncushr3d9 ncushr3d10 ncushr4d1 ncushr4d2 ncushr4d3 ncushr4d4 ///
ncushr4d5 ncushr4d6 ncushr4d7 ncushr4d8 ncushr4d9 ncushr4d10 ///
ncushr5d1 ncushr5d2 ncushr5d3 ncushr5d4 ncushr5d5 ncushr5d6 ///
ncushr5d7 ncushr5d8 ncushr5d9 ncushr5d10 ncushr6d1 ncushr6d2 ///
ncushr6d3 ncushr6d4 ncushr6d5 ncushr6d6 ncushr6d7 ncushr6d8 ///
ncushr6d9 ncushr6d10 ncushr7d1 ncushr7d2 ncushr7d3 ncushr7d4 ///
ncushr7d5 ncushr7d6 ncushr7d7 ncushr7d8 ncushr7d9 ncushr7d10 ///
ncushr8d1 ncushr8d2 ncushr8d3 ncushr8d4 ncushr8d5 ncushr8d6 ///
ncushr8d7 ncushr8d8 ncushr8d9 ncushr8d10 ncushr9d1 ncushr9d2 ///
ncushr9d3 ncushr9d4 ncushr9d5 ncushr9d6 ncushr9d7 ncushr9d8 ///
ncushr9d9 ncushr9d10 ncushr10d1 ncushr10d2 ncushr10d3 ///
ncushr10d4 ncushr10d5 ncushr10d6 ncushr10d7 ncushr10d8 ///
ncushr10d9 ncushr10d10 ncushr11d1 ncushr11d2 ncushr11d3 ///
ncushr11d4 ncushr11d5 ncushr11d6 ncushr11d7 ncushr11d8 ///
ncushr11d9 ncushr11d10 ncushr12d1 ncushr12d2 ncushr12d3 ///
ncushr12d4 ncushr12d5 ncushr12d6 ncushr12d7 ncushr12d8 ///
ncushr12d9 ncushr12d10 ///
{
replace `X'=0 if `X'==.
}

//=====//
//Format data into long form//

reshape long neu_std neutotd ncudaysd trp1_q1d trp1_q2d trp1_q3d ///
trpday1d trpday2d trpday3d ncuprivd ncupubd birds1d birds1ad birds2d ///
birds3d typbrd1d typbrd2d typbrd3d typbrd4d typbrd5d animls1d ///
animls2d animls3d animls4d animls5d ncushr1d ncushr2d ncushr3d ///
ncushr4d ncushr5d ncushr6d ncushr7d ncushr8d ncushr9d ncushr10d ///
ncushr11d ncushr12d neutotamt, i(personid id) j(site)

ren neu_std ncstate
ren neutotd nctrips
ren ncudaysd ncdays
ren trp1_q1d ncobsrv
ren trp1_q2d ncphoto
ren trp1_q3d ncfeed
ren trpday1d obday
ren trpday2d phday
ren trpday3d feday
ren ncuprivd pvtlnd
ren ncupubd publnd

```

```

ren birds1d obbird
ren birds1ad birdday
ren birds2d phbird
ren birds3d febird
ren typbrd1d brdprey
ren typbrd2d brdwater
ren typbrd3d brdshore
ren typbrd4d brdsong
ren typbrd5d brdother
ren animls1d opffish
ren animls2d opfbig
ren animls3d opfsm1
ren animls4d opfmarine
ren animls5d opfturtle
ren ncushr1d foodcost
ren ncushr2d lodgcost
ren ncushr3d aircost
ren ncushr4d traincost
ren ncushr5d carcost
ren ncushr6d guidecost
ren ncushr7d pubfee
ren ncushr8d pvtfee
ren ncushr9d equipcost
ren ncushr10d fuelcost
ren ncushr11d boatcost
ren ncushr12d cookcost
ren neutotamtd totrescost

```

```
gen personid2=.
```

```
//Travel Cost//
```

```
gen tc=(aircost+traincost+carcost+pubfee+pvtfee)/nctrips
```

```
//Calculating opportunity costs: total potential work days=250 days(weekdays
less 10 federal holidays//
```

```
//Since we are using daily earnings instead of hourly, the wage factor should be
divided by eight, i.e. .333/8=.04//
```

```
gen ncoppcost=.04*earn_hat*(ncdays/nctrips)
```

```
gen nctc=tc+ncoppcost
```

```
replace nctrips=0 if (nctrips==. & wildtrip1==0 & wildtrip2==0 & site==1)
```

```
replace ncstate=resstate if nctrips==0
```

```
replace ncdays=0 if nctrips==0
```

```
replace nctc=medtc1a if (nctrips==0)
```

```
drop if nctrips==.
```

```
//Additional data cleaning and augmentation//
```

```

ren lhincome_hat lhinc
gen outstate=0 if (ncstate==resstate)
replace outstate=1 if (outstate==.)
gen partic=1 if nctrips>0
replace partic=0 if partic==.
gen lwildsites=ln(wildsites)
gen lbirdspec=ln(birdspec)
gen nchunt=ncu_hnt
replace nchunt=1 if (hunto6==1 & nchunt==.)
gen nchunto6=hunto6
replace nchunto6=1 if (ncu_hnt==1 & hunto6==0)
gen ncfisho6=fisho6
replace ncfisho6=1 if (ncu_fish==1 & fisho6==0)

//Dropping miscellaneous variables//
drop perstat1 linenum i_wave1 i_wave2 mode_a letter_a perstat2 letter_b ///
outcome1 outcome2 mode_b i_wave3 perstat3 letter_c outcome3 mode_c Name
i_resident_fh2

//Dropping entry errors/uncertain data values//
drop if (nctrips==0 & opffish!=.) //these were contradictions in the data: 134
obs//
drop if (nctrips==0 & ncobsrv!=.) //1 obs//

save "I:\fh24mergelong.dta", replace

//Merging with choke price data and state level land data//
clear

clear matrix

use "I:\fh24mergelong.dta"

merge m:1 personid2 using "I:\fh24mergelongchoke.dta"
replace nctc=. if (personid2==. & nctrips==0)
replace nctc=wildchoke if personid2!=.
drop _merge
merge m:1 ncstate using "I:\rangeforest.dta" //2007 state level rangeland and
forest (1000) acres from Natural Resources Inventory
(NRI)(http://www.nrcs.usda.gov)//
drop _merge

//Correcting Data entries//
replace range=0 if ncstate=="DC"
replace forest=0 if ncstate=="DC"
replace forest=101918.5 if ncstate=="AK"

```

```

replace range=201528.5 if ncstate=="AK"
gen range1=range/1000
gen forest1=forest/1000
gen land=(range+forest)/(pop06/1000)
gen land2=(range+forest)/(acres*1000)
replace land=0 if land==.
replace range1=range1+1
replace forest1=forest1+1
gen land3=forest/(acres*1000)
gen land4=range/(acres*1000)

//Data used for FM Model analysis//
save "I:\fh24mergelongimpute.dta", replace

//=====FILE 5=====//

//This file was used to obtain the PMM imputations, and the FM model
estimates//

use "I:\fh24mergelongimpute.dta"

//Data contains additional rows of choke prices in the travel cost column. 3128
additional rows were added totaling 6906 and grand total 15400//

drop if (nctc>=10000 & nctc!=.) //33 obs(1.1% of choke price obs) were deleted
because they were extreme values max=.999 million.//
//Including them led to unreasonably high imputed values and problems in
achieving convergence//

replace job=0 if job==. //all 50 observations were either retired, going to
school,...//

replace env_ind=env_index if wildchoke!=. //in order to remain consistent w/
current zero trippers, env_ind is replace w/ home state values//

//Opportunity cost of a single trip to zero trippers & contingent zero trippers to
be added after imputation//

//Imputation model based on 'tc' which is nctc less opportunity cost//
replace ncoppcost=.04*earn_hat if nctrips==0 //
gen nctc_choke=wildchoke+ncoppcost
replace nctc=nctc_choke if wildchoke!=.
replace tc=wildchoke if wildchoke!=.
gen lcp=ln(wildchoke)
gen ltc=ln(tc) //missing values generated from tc=0//
gen lnctc=ln(nctc)

```

```

//Strata indicators following fhwar06 report; p.152//
gen wstrata=1 if (trip06==1 | interest06==1) //active//
replace wstrata=2 if ((trip06==0 & trip05==1) & (tripday05==5 |
tripspend05==5)) //avid//
replace wstrata=3 if (((trip06==0 & trip05==1) & (tripday05<5 & tripspend<5)) |
intlke06==1 | triplike06==1 | observe05==1 | feed05==1 | photo05==1 |
planto5==1) //average//
replace wstrata=4 if ((trip06==0 & trip05==0) & (intlke06==2 | intlke06==3 |
triplike06==2 | triplike06==3)) //infrequent//
replace wstrata=5 if ((trip06==0 & trip05==0) & (intlke06==4 | triplike06==4))
//nonparticipant//
tab wstrata if personid2==. & site==1, plot

gen nonpartic= wstrata==5
sum nonpartic if personid2==. & nctrips==0
sum nonpartic if personid2==. & nctrips>0

gen npartic=1 if nctrips==0
replace npartic=0 if nctrips>0

//Create participation potential indicators//
gen partind=1 if nctrips==0 & wstrata==5 //trip05==0 & (intlke06==4 &
triplike06==4) & (observe05==0 | feed05==0 | photo05==0 | planto5==0)//
//market nonparticipant//
replace partind=2 if nctrips==0 & (trip06==1 | trip05==1) & (intlke06!=4 |
triplike06!=4 | intlke06==. | triplike06==.) & (tripday05>0 | tripspend05>0) &
(observe05==1 | feed05==1 | photo05==1 | planto5==1) //market participant//
*tabstat env_ind if personid2==., by(ncstate) stat(n mean p50 sd min max) //141
obs for AK & 137 obs for HI & 9 for DC//
replace partind=3 if nctrips>0
tab partind if personid2==. & site==1, plot

sum nctrips age male marr lhinc wildlife wildcontr trip05 nchunto6 ncfish06
env_ind parks private natrarea retire urban photo feedbrd feedfish aown alease
mainplnt ///
wildbooks school cendiv tripspend05 white observe05 feed05 photo05 planto5
outstate if personid2==. /*& nctrips>0 & site==1*/
//initially nctrips(12272 obs), then 12224 obs =>dropped 48//

foreach X of varlist nctrips age agesq male marr lhinc wildlife wildcontr trip05
nchunto6 ncfish06 env_ind parks private /*natrarea*/ retire urban photo
feedbrd feedfish /*aown alease mainplnt*/ ///
wildbooks school cendiv tripspend05 white observe05 feed05 photo05 planto5
///
{

```

```

drop if (`X'==.) //52 obs//
}

egen totncwgt=total(ncwgt) if personid2==.
gen totncwgt1000=totncwgt/1000
dis totncwgt1000
gen ncwgt2=ncwgt*12224/115957630 //adjusts weights so it sums to sample size,
n=12224//
egen totncwgt2=total(ncwgt2) if personid2==.
dis totncwgt2

//Multiple Imputation of Travel Cost for nonparticipants//
mi set wide
mi register imputed lnctc
set seed 12380

mi svyset [pweight=ncwgt2]

//The natural log of nctc was used because nctc failed the normality test. Also,
the predicted mean matching (PMM) used 3 closest neighbors//

mi impute pmm lnctc nctrips age agesq male marr lhinc wildlife wildcontr ///
trip05 nchunto6 ncfish06 parks private i.retire urban ///
photo feedbrd feedfsh wildbooks school i.cendiv i.tripsendo5 white ///
observe05 feed05 photo05 plant05 [pweight=ncwgt2], add(10) knn(3)

mi passive: gen nctc_mi=exp(lnctc)
mi passive: replace nctc_mi=nctc if nctc!=. & personid2==.

//Table 2.2: PMM Results for travel cost//
mi estimate: svy: mean nctc_mi if personid2==.
mi estimate: svy, subpop(partic): mean nctc_mi if personid2==.
mi estimate: svy: mean nctc_mi if personid2==. & nctrips==0
mi estimate: svy: mean nctc_mi if personid2==. & nctrips==1
mi estimate: svy: mean nctc_mi if personid2==. & nctrips==2
mi estimate: svy: mean nctc_mi if personid2==. & nctrips==3
mi estimate: svy: mean nctc_mi if personid2==. & nctrips>3

//End of PMM procedure. Begin FM Model Estimation//

//Table 2.3: Average information criterion and entropy values; Testing for
components//

/*1-Component*/
forvalues i=1/10 {

```

```

nbreg nctrrips age agesq male marr lhinc school _`i'_ncte_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
fitstat
}

```

```

/*2-Component*/
forvalues i=1/10 {
fmm nctrrips age agesq male marr lhinc school _`i'_ncte_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) ///
components(2)
fmmlc
}

```

```

/*3-Component*/
forvalues i=1/10 {
fmm nctrrips age agesq male marr lhinc school _`i'_ncte_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) ///
components(3) from(.0780568 -.000842 -.0091116 -.1632147 ///
.1960878 .0695426 -.0005897 .8433348 .4699351 1.20976 ///
.703454 1.678239 .5607098 .6972714 -7.532573 -.028683 ///
.0003464 .7186719 -.0508844 .0368727 .0157604 -.0098542 ///
.5299923 .6883075 .4587726 .474313 .3594805 -.0816081 ///
1.026804 1.725451 .1077329 -.0011352 .3739535 -.5148367 ///
.6594796 .049071 -.0346685 .7949065 .6319239 ///
.949179 .4923119 .804073 1.032272 .5959123 -8.755179 ///
1.418229 -1.761949 .338601 .6615518 .229392)
*fmmlc
}

```

```

/*4-Component*/
forvalues i=1/10 {
fmm nctrrips age agesq male marr lhinc school _`i'_ncte_mi photo feedbrd parks
wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) ///
components(4) from(.0932981 -.0009609 .3735991 -.2547126 ///
.2930519 -.0202652 -.0037552 .4848303 .2242133 .4896062 ///
.5758752 .4690702 .9045633 .1436699 -3.728027 -.0176296 ///
.0001295 .6371824 -.2780966 .7016975 .1065705 -.0783202 ///
.9455834 .7727241 .6346586 .3497921 .5117884 .6174938 ///
1.179355 -5.958894 .0912965 -.0012018 -.01206 -.1599273 ///
.4287372 .0195941 -.0085934 1.139986 .8376455 1.084964 ///
.9203508 1.671505 .6398412 1.045063 -8.015345 .0924681 ///
-.0008889 .0198871 -.2316165 .1060584 .0892847 -.0002252 ///

```



```
.8650737 .130724 1.050429 .7178906 1.842565 .7531802 ///
.3606856 -7.788876 -2.2534 -1.764246 -1.016855 -.5642304 ///
.6693763 -.5458257 -.4208959)
fmmlc
}
```

```
//Estimating Single-component models for the 10 imputed data sets//
```

```
nbreg ntrips age agesq male marr lhinc school _1_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
predict c1 if e(sample)
nbreg ntrips age agesq male marr lhinc school _2_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
predict c2 if e(sample)
nbreg ntrips age agesq male marr lhinc school _3_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
predict c3 if e(sample)
nbreg ntrips age agesq male marr lhinc school _4_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
predict c4 if e(sample)
nbreg ntrips age agesq male marr lhinc school _5_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
predict c5 if e(sample)
nbreg ntrips age agesq male marr lhinc school _6_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
predict c6 if e(sample)
nbreg ntrips age agesq male marr lhinc school _7_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
predict c7 if e(sample)
nbreg ntrips age agesq male marr lhinc school _8_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
predict c8 if e(sample)
nbreg ntrips age agesq male marr lhinc school _9_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
predict c9 if e(sample)
nbreg ntrips age agesq male marr lhinc school _10_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)
```

```

predict c101 if e(sample)

gen tp_nb=(c1+c2+c3+c4+c5+c6+c7+c8+c9+c10)/10

//Obtain MI estimates for Single-Component model//
mi estimate: nbreg nctrips age agesq male marr lhinc school nctc_mi photo
feedbrd parks wildcontr trip05 nchunto6 ncfish06 if personid2==.
[pweight=ncwgt2], ///
difficult vce(cluster personid) technique(nr)

//Using estimates from mi estimate to create CS lower limit in Single-component
model//
gen tcon=exp(-3.046768+.0303139*age-.0003226*agesq+.4101752*male- ///
.2577584*marr+.0642338*lhinc+.0548194*school+.8909739*photo+ ///
.8141454*feedbrd+1.073308*parks+.4086888*wildcontr+1.023345*trip05 ///
+.9778609*nchunto6+.8616468*ncfish06) if personid2==.

gen co_nb=(1/-.0029207)*ln(3.276973/4.26014) //co_nb=89.835274//
dis co_nb

//Estimating 3-component models for the 10 imputed data sets//

fmm nctrips age agesq male marr lhinc school _1_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfish06 if personid2==. [pweight=ncwgt2],
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) ///
components(3) /*trace*/ ///
from(.0780568 -.000842 -.0091116 -.1632147 .1960878 .0695426 ///
-.0005897 .8433348 .4699351 1.20976 .703454 1.678239 ///
.5607098 .6972714 -7.532573 -.028683 .0003464 .7186719 /// -
.0508844 .0368727 .0157604 -.0098542 .5299923 .6883075 ///
.4587726 .474313 .3594805 -.0816081 1.026804 1.725451 ///
.1077329 -.0011352 .3739535 -.5148367 .6594796 .049071 /// -
.0346685 .7949065 .6319239 .949179 .4923119 .804073 ///
1.032272 .5959123 -8.755179 1.418229 -1.761949 .338601 ///
.6615518 .229392)

predict tp11 if e(sample), equation(component1)
predict tp21 if e(sample), equation(component2)
predict tp31 if e(sample), equation(component3)
predict tp1231 if e(sample)
predict pst11 if e(sample), posterior equation(component1)
predict pst21 if e(sample), posterior equation(component2)
predict pst31 if e(sample), posterior equation(component3)
gen g1=1 if (pst11>pst21 & pst11>pst31)
replace g1=2 if (pst21>pst11 & pst21>pst31)
replace g1=3 if (pst31>pst11 & pst31>pst21)

```

```

univar nctrips _1_nctc_mi outstate age male marr lhinc school retired wildlife
wildcontr trip05 nchunto6 ncfish06 nonpartic ///
ncdays ncobsrv ncphoto ncfed pvtlnd publnd land if nctrips>0, by(g1)

```

```

gen tcon11=exp(-7.532573+0.0780568*age-0.000842*agesq-0.0091116*///
male0.1632147*marr+0.1960878*lhinc+0.0695426*school+0.8433348*///
photo+0.4699351*feedbrd+1.20976*parks+0.703454*wildcontr+1.678239///
*trip05+0.5607098*nchunto6+0.6972714*ncfish06) if e(sample)
gen tcon21=exp(1.725451-0.028683*age+0.0003464*agesq+0.7186719*///
male-0.0508844*marr+0.0368727*lhinc+0.0157604*school+///
0.5299923*photo+0.6883075*feedbrd+0.4587726*parks+0.474313*///
wildcontr+0.3594805*trip05-0.0816081*nchunto6+1.026804*ncfish06) if
e(sample)
gen tcon31=exp(-8.755179+0.1077329*age-0.0011352*agesq+0.3739535*///
male-0.5148367*marr+0.6594796*lhinc+0.049071*school+0.7949065*///
photo+0.6319239*feedbrd+0.949179*parks+0.4923119*wildcontr+///
0.804073*trip05+1.032272*nchunto6+0.5959123*ncfish06) if e(sample)

```

```

/*****/
frm nctrips age agesq male marr lhinc school _2_nctc_mi photo feedbrd parks
wildcontr trip05 nchunto6 ncfish06 if personid2==. [pweight=ncwgt2], ///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) ///
components(3) /*trace*/ ///
from(0.0708973 -0.0007806 -0.0313866 -0.174372 0.2072993 ///
0.0692952 -0.0005079 0.847606 0.5003767 1.204931 ///
0.6979081 1.664115 0.5913673 0.7001766 -7.454391 ///
-0.0280852 0.0003415 0.7161696 -0.0396754 0.0152953 ///
0.012303 -0.0098941 0.5316211 0.6895315 0.5118858 ///
0.4737978 0.333105 -0.0043157 0.9879732 1.923338 ///
0.110506 -0.0011647 0.3615294 -0.5173733 0.7071932 ///
0.0569523 -0.036461 0.768718 0.670128 0.9261724 0.4756491 ///
0.8289184 1.02183 0.6217959 -9.423879 ///
1.464485 -1.634971 0.4065391 0.6916477 0.2148769)
predict tp12 if e(sample), equation(component1)
predict tp22 if e(sample), equation(component2)
predict tp32 if e(sample), equation(component3)
predict tp1232 if e(sample)
predict pst12 if e(sample), posterior equation(component1)
predict pst22 if e(sample), posterior equation(component2)
predict pst32 if e(sample), posterior equation(component3)
gen g2=1 if (pst12>pst22 & pst12>pst32)
replace g2=2 if (pst22>pst12 & pst22>pst32)
replace g2=3 if (pst32>pst12 & pst32>pst22)
univar _2_nctc_mi outstate age male marr lhinc school retired wildlife ///
wildcontr trip05 nchunto6 ncfish06 nonpartic ///
ncdays ncobsrv ncphoto ncfed pvtlnd publnd land if nctrips>0, by(g2)

```

```

gen tcon12=exp(-7.454391+0.0708973*age-0.0007806*agesq- ///
0.0313866*male-0.174372*marr+0.2072993*lhinc+0.0692952*
school+0.847606*/// photo+0.5003767*feedbrd ///
+1.204931*parks+0.6979081*wildcontr+1.664115*trip05+0.5913673*///
nchunto6+0.7001766*ncfisho6) if e(sample)
gen tcon22=exp(1.923338-0.0280852*age+0.0003415*agesq+///
0.7161696*male- ///
0.0396754*marr+0.0152953*lhinc+0.012303*school+0.5316211*photo ///
+0.6895315*feedbrd ///
+0.5118858*parks+0.4737978*wildcontr+0.333105*trip05- ///
0.0043157*nchunto6+0.9879732*ncfisho6) if e(sample)
gen tcon32=exp(-9.423879+0.110506*age-0.0011647*agesq+///
0.3615294*male-
///0.5173733*marr+0.7071932*lhinc+0.0569523*school+0.768718 ///
*photo+0.670128*feedbrd ///
+0.9261724*parks+0.4756491*wildcontr+0.8289184*trip05+1.02183 ///
*nchunto6+0.6217959*ncfisho6) if e(sample)

/*****/
fmm nctrips age agesq male marr lhinc school _3_nctc_mi photo feedbrd ///
parks wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2],
///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) components(3)
/*trace*/ ///
from(.074843 -.0008105 -.0134107 -1.847175 .2188725 .0731565 ///
-.0005993 .8459815 .4977331 1.195974 .6993784 1.682093 ///
.5164949 ///
.7014508 -7.742872 -.0259345 .0003208 .7189236 -.047614 ///
.0153803 .0138082 -.0097269 .5144137 .6916606 .4819593 ///
.4457557 .3284057 .0573018 .9979791 1.904004 .1011659 ///
-.0010588 .3827304 -.5212476 .6829229 .0615322 -.035335 ///
.8245169 .6783392 .950509 .5117617 .8648746 .9921521 ///
.6303132 -9.134355 1.432894 -1.721493 .3579461 .6532512 ///
.2631812)
predict tp13 if e(sample), equation(component1)
predict tp23 if e(sample), equation(component2)
predict tp33 if e(sample), equation(component3)
predict tp1233 if e(sample)
predict pst13 if e(sample), posterior equation(component1)
predict pst23 if e(sample), posterior equation(component2)
predict pst33 if e(sample), posterior equation(component3)
gen g3=1 if (pst13>pst23 & pst13>pst33)
replace g3=2 if (pst23>pst13 & pst23>pst33)
replace g3=3 if (pst33>pst13 & pst33>pst23)

```

```

univar _3_nctc_mi outstate age male marr lhinc school retired wildlife ///
wildcontr trip05 nchunto6 ncfish06 nonpartic ncdays ncobsvr nphoto ///
ncfeed pvtlnd publnd, by(g3)

```

```

gen tcon13=exp(-7.742872+0.074843*age-0.0008105*agesq- ///
0.0134107*male-0.1847175*marr+0.2188725*lhinc+0.0731565 ///
*school+0.8459815*photo+0.4977331*feedbrd ///
+1.195974*parks+0.6993784*wildcontr+1.682093*trip05+0.5164949*///
nchunto6+0.7014508*ncfish06) if e(sample)
gen tcon23=exp(1.904003-0.0259344*age+0.0003208*agesq+ ///
0.7189234*male-0.0476142*marr+0.0153803*lhinc+0.0138082 ///
*school+0.5144138*photo+0.6916607*feedbrd ///
+0.4819598*parks+0.4457556*wildcontr+0.3284058*trip05+ ///
0.057302*nchunto6+0.9979787*ncfish06) if e(sample)
gen tcon33=exp(-9.134362+0.1011659*age-0.0010588*agesq+ ///
0.3827307*male- ///
0.5212474*marr+0.6829236*lhinc+0.0615322*school+0.8245164 ///
*photo+0.6783394*feedbrd ///
+0.9505087*parks+0.5117615*wildcontr+0.8648747*trip05+ ///
0.9921523*nchunto6+0.6303128*ncfish06) if e(sample)

```

```

/*****/
fmm nctrips age agesq male marr lhinc school _4_nctc_mi photo ///
feedbrd parks wildcontr trip05 nchunto6 ncfish06 if personid2==. ///
[pweight=newgt2], mixtureof(negbin2) vce(cluster personid) difficult ///
technique(nr) components(3) /*trace*/ ///
from(.0757717 -.0008249 -.0114037 -.1868786 .2100781 ///
.0695924 -.0006193 .8263674 .5017736 1.212104 .6853117 ///
1.685827 .6142868 .6796289 -7.597151 -.0297359 .0003538 ///
.724043 -.0483157 .0282348 .0147377 -.0096973 .5496611 ///
.6791634 .4820176 .4685902 .3557187 -.1146064 1.047677 ///
1.871717 .1049171 -.0011132 .3814262 -.5119896 .6445525 ///
.0555395 -.0331326 .7546436 .6866012 .9440965 .4912773 ///
.8166562 1.046998 .6141657 -8.654384 1.481821 -1.736899 ///
.383567 .6393538 .2106884)
predict tp14 if e(sample), equation(component1)
predict tp24 if e(sample), equation(component2)
predict tp34 if e(sample), equation(component3)
predict tp1234 if e(sample)
predict pst14 if e(sample), posterior equation(component1)
predict pst24 if e(sample), posterior equation(component2)
predict pst34 if e(sample), posterior equation(component3)
gen g4=1 if (pst14>pst24 & pst14>pst34)
replace g4=2 if (pst24>pst14 & pst24>pst34)
replace g4=3 if (pst34>pst14 & pst34>pst24)

```

```

univar _4_nctc_mi outstate age male marr lhinc school retired wildlife ///
wildcontr trip05 nchunto6 ncfisho6 nonpartic ncdays ncobsvr ncphoto ///
ncfeed pvtlnd publnd, by(g4)

```

```

gen tcon14=exp(-7.597148+0.0757708*age-0.0008249*agesq-///
0.0114025*male-///
0.1868764*marr+0.210081*lhinc+0.0695917*school+0.8263661///
*photo+0.5017751*feedbrd ///
+1.212102*parks+0.6853109*wildcontr+1.685828*trip05+ ///
0.614284*nchunto6+0.6796322*ncfisho6) if e(sample)
gen tcon24=exp(1.871658-0.0297346*age+0.0003538*agesq+///
0.7240283*male-///
0.0483324*marr+0.0282313*lhinc+0.0147382*school+0.5496614 ///
*photo+0.6791655*feedbrd ///
+0.4820541*parks+0.4685838*wildcontr+0.3557261*trip05- ///
0.1145798*nchunto6+1.047629*ncfisho6) if e(sample)
gen tcon34=exp(-8.65496+0.1049177*age-0.0011132*agesq+ ///
0.3814535*male-///
0.5119717*marr+0.6445525*lhinc+0.05554*school+0.7546101* ///
photo+0.6866155*feedbrd ///
+0.9440807*parks+0.4912602*wildcontr+0.8166628*trip05+ ///
1.047009*nchunto6+0.6141359*ncfisho6) if e(sample)

```

```

/*****/
fmn nctrips age agesq male marr lhinc school _5_nctc_mi photo feedbrd parks
wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) components(3)
/*trace*/ ///
from(.0779535 -.0008352 -.0098337 -.1703383 .1922358 ///
.0778584 -.0006187 .8254089 .4784475 1.188945 ///
.6962863 1.688665 .5625217 ///
.6869882 -7.602526 -.0290204 .0003487 .7145895 ///
-.0486902 .0232256 .0192482 -.0097451 .5461551 ///
.6788694 .5297874 .4548459 ///
.3570483 -.1014167 1.016827 1.765604 .1069923 -.0011328 ///
.4119678 -.5000939 .7035616 .0613461 -.0371987 .7686411 .6869706
.8875225 .5116731 .8265945 1.047386 .6005649 -9.42004 ///
1.380241 -1.696266 .3105988 .6914445 .2702445)
predict tp15 if e(sample), equation(component1)
predict tp25 if e(sample), equation(component2)
predict tp35 if e(sample), equation(component3)
predict tp1235 if e(sample)
predict pst15 if e(sample), posterior equation(component1)
predict pst25 if e(sample), posterior equation(component2)
predict pst35 if e(sample), posterior equation(component3)
gen g5=1 if (pst15>pst25 & pst15>pst35)
replace g5=2 if (pst25>pst15 & pst25>pst35)

```

```

replace g5=3 if (pst35>pst15 & pst35>pst25)
univar _5_nctc_mi outstate age age male marr lhinc school retired wildlife ///
wildcontr trip05 nchunto6 ncfish06 nonpartic ///
ncdays ncobsrv ncphoto ncfed pvtlnd publnd, by(g5)

gen tcon15=exp(-7.602526+0.0779535*age-0.0008352*agesq-///
0.0098337*male-///
0.1703383*marr+0.1922358*lhinc+0.0778584*school+0.8254089 ///
*photo+0.4784475*feedbrd ///
+1.188945*parks+0.6962863*wildcontr+1.688665*trip05+0.5625217 ///
*nchunto6+0.6869882*ncfish06) if e(sample)
gen tcon25=exp(1.765602-0.0290204*age+0.0003487*agesq+ ///
0.7145896*male-///
0.0486906*marr+0.0232258*lhinc+0.0192482*school+0.546155* ///
photo+0.6788698*feedbrd ///
+0.5297875*parks+0.4548459*wildcontr+0.3570483*trip05- ///
0.1014149*nchunto6+1.016827*ncfish06) if e(sample)
gen tcon35=exp(-9.42004+0.1069923*age-0.0011328*agesq+ ///
0.4119678*male-///
0.5000939*marr+0.7035616*lhinc+0.0613461*school+0.7686411* ///
photo+0.6869706*feedbrd ///
+0.8875225*parks+0.5116731*wildcontr+0.8265945*trip05+ ///
1.047386*nchunto6+0.6005649*ncfish06) if e(sample)

/*****
fmm nctrips age agesq male marr lhinc school _6_nctc_mi photo feedbrd parks
wildcontr trip05 nchunto6 ncfish06 if personid2= . [pweight=ncwgt2], ///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) components(3)
/*trace*/ ///
from(.0708887 -.0007873 -.0279639 -.163999 .2066887 ///
.0691788 -.0005342 .8360727 .5053213 1.193486 .7042503 ///
1.690595 .6464416 ///
.6991508 -7.436987 -.0302653 .000359 .7281187 -.0480176 ///
.0133438 .0180519 -.0096869 .5453554 .6813864 .5544011 ///
.4445855 ///
.3588027 -.1132916 1.030052 1.937496 .1090012 -.0011589 ///
.3665494 -.5041877 .7202556 .059273 -.0350962 .7439782 ///
.713064 .8826755 .4986136 .8269235 1.088872 .6070921 ///
-9.57355 1.521671 -1.641607 .4184927 .6644148 .1684165)
predict tp16 if e(sample), equation(component1)
predict tp26 if e(sample), equation(component2)
predict tp36 if e(sample), equation(component3)
predict tp1236 if e(sample)
predict pst16 if e(sample), posterior equation(component1)
predict pst26 if e(sample), posterior equation(component2)
predict pst36 if e(sample), posterior equation(component3)
gen g6=1 if (pst16>pst26 & pst16>pst36)

```

```

replace g6=2 if (pst26>pst16 & pst26>pst36)
replace g6=3 if (pst36>pst16 & pst36>pst26)
univar _6_nctc_mi outstate age male marr lhinc school retired wildlife w ///
ildcontr trip05 nchunto6 ncfish06 nonpartic ///
ncdays ncobsrv ncphoto ncfed pvtlnd publnd, by(g6)

```

```

gen tcon16=exp(-7.436987+0.0708887*age-0.0007873*agesq- ///
0.0279639*male- ///
0.163999*marr+0.2066887*lhinc+0.0691788*school+0.8360727 ///
*photo+0.5053213*feedbrd ///
+1.193486*parks+0.7042503*wildcontr+1.690595*trip05+0.6464416 ///
*nchunto6+0.6991508*ncfish06) if e(sample)
gen tcon26=exp(1.937496-0.0302653*age+0.000359*agesq+ ///
0.7281187*male-///
0.0480176*marr+0.0133438*lhinc+0.0180519*school+0.5453554* ///
photo+0.6813864*feedbrd ///
+0.5544011*parks+0.4445855*wildcontr+0.3588027*trip05- ///
0.1132916*nchunto6+1.030052*ncfish06) if e(sample)
gen tcon36=exp(-9.5735+0.1090012*age-0.0011589*agesq+ ///
0.3665494*male-///
0.5041877*marr+0.7202556*lhinc+0.059273*school+0.7439782 ///
*photo+0.713064*feedbrd ///
+0.8826755*parks+0.4986136*wildcontr+0.8269235*trip05+ ///
1.088872*nchunto6+0.6070921*ncfish06) if e(sample)

```

```

/*****/
fmm nctrips age agesq male marr lhinc school _7_nctc_mi photo feedbrd parks
wildcontr trip05 nchunto6 ncfish06 if personid2==. [pweight=ncwgt2], ///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) components(3)
/*trace*/ ///
from(.0758311 -.0008291 -.0229071 -.1482405 .2058946 ///
.074374 -.0005461 .8249749 .5260225 1.22422 .7123037 ///
1.681839 .5493531 ///
.7210285 -7.681908 -.0290843 .0003459 .7235422 -.0448934 ///
.0110161 .015717 -.0093675 .5497895 .6726555 .5097662 ///
.4431818 ///
.3622747 -.0973945 1.064539 2.001135 .1077662 -.001148 ///
.3623654 -.4931171 .6309466 .0575541 -.0318545 .7577236 ///
.715547 .9516074 .5238737 .8339863 1.041989 .6107923 - ///
8.653926 1.515831 -1.698703 .38185 .6403727 .2307349)
predict tp17 if e(sample), equation(component1)
predict tp27 if e(sample), equation(component2)
predict tp37 if e(sample), equation(component3)
predict tp1237 if e(sample)
predict pst17 if e(sample), posterior equation(component1)
predict pst27 if e(sample), posterior equation(component2)
predict pst37 if e(sample), posterior equation(component3)

```



```

gen g7=1 if (pst17>pst27 & pst17>pst37)
replace g7=2 if (pst27>pst17 & pst27>pst37)
replace g7=3 if (pst37>pst17 & pst37>pst27)
univar _7_ncte _mi outstate age male marr lhinc school retired wildlife wildcontr
trip05 nchunto6 ncfisho6 nonpartic ///
ncdays ncobsrv ncphoto ncfed pvtlnd publnd, by(g7)

```

```

gen tcon17=exp(-7.681908+0.0758311*age-0.0008291*agesq-0.0229071*male-
0.1482405*marr+0.2058946*lhinc+0.074374*school+0.8249749*photo+0.526
0225*feedbrd ///
+1.22422*parks+0.7123037*wildcontr+1.681839*trip05+0.5493531*nchunto6+
0.7210285*ncfisho6) if e(sample)
gen tcon27=exp(2.001135-0.0290843*age+0.0003459*agesq+0.7235422*male-
0.0448934*marr+0.0110161*lhinc+0.015717*school+0.5497895*photo+0.67265
55*feedbrd ///
+0.5097662*parks+0.4431818*wildcontr+0.3622747*trip05-
0.0973945*nchunto6+1.064539*ncfisho6) if e(sample)
gen tcon37=exp(-8.653926+0.1077662*age-0.001148*agesq+0.3623654*male-
0.4931171*marr+0.6309466*lhinc+0.0575541*school+0.7577236*photo+0.7155
47*feedbrd ///
+0.9516074*parks+0.5238737*wildcontr+0.8339863*trip05+1.041989*nchunto
6+0.6107923*ncfisho6) if e(sample)

```

```

/*****/
fmm nctrips age agesq male marr lhinc school _8_ncte _mi photo feedbrd parks
wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) components(3)
/*trace*/ ///
from(.0724633 -.0007957 -.0076931 -.1688684 .2042587 .0702085 -
.0004974 .8384893 .497014 1.202331 .693626 1.696706 .547249
///
.700792 -7.492898 -.032672 .000378 .7114763 -.0357256 .0041805
.0193429 -.00959 .5596117 .6771816 .5811016 .4577401 ///
.3661606 -.0892211 1.025671 1.99336 .1068909 -.0011353 .4077529
-.511473 .7497179 .0602064 -.0384632 .7039106 .6972749 ///
.869247 .4898828 .8220303 1.044915 .5283158 -9.78146 1.50528
-1.551913 .4038016 .7246557 .1696543)
predict tp18 if e(sample), equation(component1)
predict tp28 if e(sample), equation(component2)
predict tp38 if e(sample), equation(component3)
predict tp1238 if e(sample)
predict pst18 if e(sample), posterior equation(component1)
predict pst28 if e(sample), posterior equation(component2)
predict pst38 if e(sample), posterior equation(component3)
gen g8=1 if (pst18>pst28 & pst18>pst38)
replace g8=2 if (pst28>pst18 & pst28>pst38)
replace g8=3 if (pst38>pst18 & pst38>pst28)

```

```

univar _8_nctc_mi outstate age male marr lhinc school retired wildlife wildcontr
trip05 nchunto6 ncfish06 nonpartic ///
ncdays ncobsrv ncphoto ncfed pvtlnd publnd, by(g8)

```

```

gen tcon18=exp(-7.492898+0.0724633*age-0.0007957*agesq-0.0076931*male-
0.1688684*marr+0.2042587*lhinc+0.0702085*school+0.8384893*photo+0.49
7014*feedbrd ///
+1.202331*parks+0.693626*wildcontr+1.696706*trip05+0.547249*nchunto6+0
.700792*ncfish06) if e(sample)

```

```

gen tcon28=exp(1.99336-0.032672*age+0.000378*agesq+0.7114763*male-
0.0357256*marr+0.0041805*lhinc+0.0193429*school+0.5596117*photo+0.6771
816*feedbrd ///
+0.5811016*parks+0.4577401*wildcontr+0.3661606*trip05-
0.0892211*nchunto6+1.025671*ncfish06) if e(sample)

```

```

gen tcon38=exp(-9.78146+0.1068909*age-0.0011353*agesq+0.4077529*male-
0.511473*marr+0.7497179*lhinc+0.0602064*school+0.7039106*photo+0.69727
49*feedbrd ///
+0.869247*parks+0.4898828*wildcontr+0.8220303*trip05+1.044915*nchunto
6+0.5283158*ncfish06) if e(sample)

```

```

/***** /
fmm ntrips age agesq male marr lhinc school _9_nctc_mi photo feedbrd parks
wildcontr trip05 nchunto6 ncfish06 if personid2==. [pweight=ncwgt2], ///
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) components(3)
/*trace*/ ///
from(.072783 -.0007934 -.0140748 -.1590192 .2155736 .0687449 -
.00057 .8282696 .4830566 1.187148 .6828708 1.706293 .5410034
///
.7062497 -7.604495 -.029041 .0003476 .6982083 -.0451293
.0094589 .0170762 -.0097117 .5380089 .6656172 .5418408 .458521
///
.3620847 -.0901166 .9998603 1.908933 .1085862 -.00115 .4141809
-.5114906 .7799658 .0589894 -.0385898 .7493254 .7184858 ///
.8941871 .4759828 .8632638 1.075834 .6149584 -10.25801 1.414485
-1.598671 .3457343 .7087563 .2501986)
predict tp19 if e(sample), equation(component1)
predict tp29 if e(sample), equation(component2)
predict tp39 if e(sample), equation(component3)
predict tp1239 if e(sample)
predict pst19 if e(sample), posterior equation(component1)
predict pst29 if e(sample), posterior equation(component2)
predict pst39 if e(sample), posterior equation(component3)
gen g9=1 if (pst19>pst29 & pst19>pst39)
replace g9=2 if (pst29>pst19 & pst29>pst39)
replace g9=3 if (pst39>pst19 & pst39>pst29)
univar _9_nctc_mi outstate age male marr lhinc school retired wildlife wildcontr
trip05 nchunto6 ncfish06 nonpartic ///

```

ncdays ncobsrv ncphoto ncfed pvtlnd publnd, by(g9)

```
gen tcon19=exp(-7.604495+0.072783*age-0.0007934*agesq-0.0140748*male-
0.1590192*marr+0.2155736*lhinc+0.0687449*school+0.8282696*photo+0.483
0566*feedbrd ///
+1.187148*parcs+0.6828708*wildcontr+1.706293*trip05+0.5410034*nchunto6
+0.7062497*ncfisho6) if e(sample)
gen tcon29=exp(1.908933-0.029041*age+0.0003476*agesq+0.6982083*male-
0.0451293*marr+0.0094589*lhinc+0.0170762*school+0.5380089*photo+0.66
56172*feedbrd ///
+0.5418408*parcs+0.458521*wildcontr+0.3620847*trip05-
0.0901166*nchunto6+0.9998603*ncfisho6) if e(sample)
gen tcon39=exp(-10.25801+0.1085862*age-0.00115*agesq+0.4141809*male-
0.5114906*marr+0.7799658*lhinc+0.0589894*school+0.7493254*photo+0.718
4858*feedbrd ///
+0.8941871*parcs+0.4759828*wildcontr+0.8632638*trip05+1.075834*nchunto
6+0.6149584*ncfisho6) if e(sample)
```

```
/******  
fmm nctrips age agesq male marr lhinc school _10_nctc_mi photo feedbrd parcs  
wildcontr trip05 nchunto6 ncfisho6 if personid2==. [pweight=ncwgt2], ///  
mixtureof(negbin2) vce(cluster personid) difficult technique(nr) components(3)  
/*trace*/ ///  
from(.0759582 -.0008286 -.0156957 -.1963789 .20828 .0717438 -  
.0006094 .8321341 .5184441 1.213525 .6954017 1.712399 .6199228  
///  
.6586268 -7.636743 -.0280322 .0003371 .7478974 -.0857765  
.0487991 .0145827 -.0096031 .537938 .7347365 .5530767 .4357401  
///  
.3572624 .1394161 1.027387 1.631027 .099214 -.0010617 .3546976  
-.5146526 .5977464 .058045 -.0300469 .7761371 .6585633 ///  
.9368653 .4759611 .8577358 .9623489 .5964677 -8.060461 1.557239  
-1.729004 .3854412 .6127237 .1746849)  
predict tp110 if e(sample), equation(component1)  
predict tp210 if e(sample), equation(component2)  
predict tp310 if e(sample), equation(component3)  
predict tp12310 if e(sample)  
predict pst110 if e(sample), posterior equation(component1)  
predict pst210 if e(sample), posterior equation(component2)  
predict pst310 if e(sample), posterior equation(component3)  
gen g10=1 if (pst110>pst210 & pst110>pst310)  
replace g10=2 if (pst210>pst110 & pst210>pst310)  
replace g10=3 if (pst310>pst110 & pst310>pst210)  
univar _10_nctc_mi outstate age male marr lhinc school retired wildlife  
wildcontr trip05 nchunto6 ncfisho6 nonpartic ///  
ncdays ncobsrv ncphoto ncfed pvtlnd publnd, by(g10)
```

```

gen tcon110=exp(-7.636728+0.0759579*age-0.0008286*agesq-
0.0156968*male-
0.1963782*marr+0.2082801*lhinc+0.0717435*school+0.8321341*photo+0.5184
439*feedbrd ///
+1.213524*parcs+0.6954028*wildcontr+1.7124*trip05+0.6199287*nchunto6+0.
6586281*ncfish06) if e(sample)
gen tcon210=exp(1.631205-0.0280357*age+0.0003371*agesq+0.7478951*male-
0.0857385*marr+0.0487865*lhinc+0.0145838*school+0.5379462*photo+0.734
706*feedbrd ///
+0.5530573*parcs+0.435745*wildcontr+0.3572623*trip05+0.1392546*nchunto
6+1.027429*ncfish06) if e(sample)
gen tcon310=exp(-8.060484+0.0992154*age-
0.0010617*agesq+0.3546983*male-
0.5146653*marr+0.5977471*lhinc+0.0580448*school+0.7761326*photo+0.658
5734*feedbrd ///
+0.9368644*parcs+0.4759624*wildcontr+0.8577316*trip05+0.9624*nchunto6
+0.5964613*ncfish06) if e(sample)

```

//The MI coefficient estimates for the 3-component FM model reported in Table 2.4 were calculated in MS Excel//

//The MI estimate of the relative entropy value was obtained using the calculations reported below//

//Calculating MI mixing probabilities//

```

gen pst1=(pst11+pst12+pst13+pst14+pst15+pst16+pst17+pst18+pst19+pst110)/10
gen
pst2=(pst21+pst22+pst23+pst24+pst25+pst26+pst27+pst28+pst29+pst210)/10
gen
pst3=(pst31+pst32+pst33+pst34+pst35+pst36+pst37+pst38+pst39+pst310)/10
gen g=1 if (pst1>pst2 & pst1>pst3)
replace g=2 if (pst2>pst1 & pst2>pst3)
replace g=3 if (pst3>pst1 & pst3>pst2)

```

//=====Relative Entropy=====//

```

gen lpst1=ln(pst1)
gen lpst2=ln(pst2)
gen lpst3=ln(pst3)
replace lpst1=-81 if pst1==0
replace lpst2=-81 if pst2==0
replace lpst3=-81 if pst3==0

```

```

gen pst1lpst1=pst1*lpst1
gen pst2lpst2=pst2*lpst2
gen pst3lpst3=pst3*lpst3
gen sum__pstlpst=pst1lpst1+pst2lpst2+pst3lpst3
egen tot__pstlpst=total(sum__pstlpst)

```

```

gen entropy=1+(tot__pstlpst/(12224*ln(3)))

dis entropy //e=0.57688624//

//Table 2.5: Component-level mean values//
univar nctc_miavg nctrips outstate age male marr lhinc school retired wildlife
wildcontr trip05 nchunto6 ncfisho6 nonpartic ///
ncdays ncobsrv ncphto ncfed pvtlnd publnd land land2 if personid2==., by(g)
univar nctc_miavg nctrips outstate age male marr lhinc school retired wildlife
wildcontr trip05 nchunto6 ncfisho6 nonpartic ///
ncdays ncobsrv ncphto ncfed pvtlnd publnd land land2 if personid2==. &
nctrips>0, by(g)
univar nctc_miavg nctrips outstate age male marr lhinc school retired wildlife
wildcontr trip05 nchunto6 ncfisho6 nonpartic ///
ncdays ncobsrv ncphto ncfed pvtlnd publnd land land2 if personid2==. &
nctrips==0, by(g)

//Calculate component-level lower bounds co and expected counts using MI
estimates//

gen
tcon1=(tcon11+tcon12+tcon13+tcon14+tcon15+tcon16+tcon17+tcon18+tcon19+t
con110)/10
gen
tcon2=(tcon21+tcon22+tcon23+tcon24+tcon25+tcon26+tcon27+tcon28+tcon29
+tcon210)/10
gen
tcon3=(tcon31+tcon32+tcon33+tcon34+tcon35+tcon36+tcon37+tcon38+tcon39
+tcon310)/10

gen tp1=(tp11+tp12+tp13+tp14+tp15+tp16+tp17+tp18+tp19+tp110)/10
gen tp2=(tp21+tp22+tp23+tp24+tp25+tp26+tp27+tp28+tp29+tp210)/10
gen tp3=(tp31+tp32+tp33+tp34+tp35+tp36+tp37+tp38+tp39+tp310)/10
gen
tp123=(tp1231+tp1232+tp1233+tp1234+tp1235+tp1236+tp1237+tp1238+tp1239
+tp12310)/10

sum tcon1 tcon2 tcon3 tp1 tp2 tp3 tp123 pst1 pst2 pst3 nctrips if e(sample)

gen co1=(1/-0.0005692)*ln(.7500889/.8142377)
gen co2=(1/-0.00968767)*ln(18.48249/32.42402)
gen co3=(1/-0.03508482)*ln(7.15158/30.42196)

sum co1 co2 co3 tp1 tp2 tp3 tp123 pst1 pst2 pst3 if e(sample)

//Data used for obtaining CS estimates//
save "I:\fmm_post.dta", replace

```

```

//=====FILE 6=====//

//This file is used to conduct post-estimation analyses//

use "I:\fmm_post.dta"

gen ncdaystrip=ncdays/nctrips
gen ncdaystrip2=ncdaystrip
replace ncdaystrip2=0 if ncdays==0 & personid2==.

//Table 2.5: Component-level mean values//
mi estimate: mean nctc_mi if personid2==. & g==1
mi estimate: mean nctc_mi if personid2==. & g==1 & nctrips==0
mi estimate: mean nctc_mi if personid2==. & g==1 & nctrips>0
mi estimate: mean nctc_mi if personid2==. & g==1 & nctrips>0 & outstate==0
mi estimate: mean nctc_mi if personid2==. & g==1 & nctrips>0 & outstate==1

mi estimate: mean nctc_mi if personid2==. & g==2
mi estimate: mean nctc_mi if personid2==. & g==3

mi estimate: mean nctc_mi if personid2==.
mi estimate: mean nctc_mi if personid2==. & nctrips==0
mi estimate: mean nctc_mi if personid2==. & nctrips>0

univar tp1 tp2 tp3 nctrips ncdaystrip outstate age male marr lhinc school retired
wildlife wildcontr trip05 nchunto6 ncfish06 nonpartic ///
ncdays ncobsrv ncphoto ncfed pvtlnd publnd obbird phbird febird birdday
brdprey brdwater brdshore brdsong brdother opffish opfbig opfsml opfmarine
opfturtle ///
land land2 urban birdspec wildsites if personid2==. & nctrips>0, by(g)
univar tp1 tp2 tp3 nctrips ncdaystrip outstate age male marr lhinc school retired
wildlife wildcontr trip05 nchunto6 ncfish06 nonpartic ///
ncdays ncobsrv ncphoto ncfed pvtlnd publnd obbird phbird febird birdday
brdprey brdwater brdshore brdsong brdother opffish opfbig opfsml opfmarine
opfturtle ///
land land2 urban birdspec wildsites if personid2==. & nctrips>0 & outstate==0,
by(g)
univar tp1 tp2 tp3 nctrips ncdaystrip outstate age male marr lhinc school retired
wildlife wildcontr trip05 nchunto6 ncfish06 nonpartic ///
ncdays ncobsrv ncphoto ncfed pvtlnd publnd obbird phbird febird birdday
brdprey brdwater brdshore brdsong brdother opffish opfbig opfsml opfmarine
opfturtle ///
land land2 urban birdspec wildsites if personid2==. & nctrips>0 & outstate==1,
by(g)

```

```

univar nctrips ncdaystrip outstate age male marr lhinc school retired wildlife
wildcontr trip05 nchunto6 ncfish06 nonpartic ///
ncdays ncobsrv ncphto ncfed pvtlnd publnd obbird phbird febird birdday
brdprey brdwater brdshore brdsong brdother opffish opfbig opfsm1 opfmarine
opfturtle ///
land land2 urban birdspec wildsites if personid2==. & nctrips==0, by(g)

```

```

//Calculating survey weighted expected trip count and lower limit values needed
to calculate consumer surplus(CS) estimates//
//Final CS values reported in Table 2.6 were obtained using MS Excel//

```

```

mi unset
svyset [pweight=ncwgt2]

```

```

svy: mean ncte_choke //408.9574//
svy: mean partic nctrips tp_nb tp1 tp2 tp3 tp123 co_nb co1 co2 co3 tcon tcon1
tcon2 tcon3 if personid2==.

```

```

gen svyco_nb=(1/-.0029207)*ln(2.68695/3.503886) //90.89//
gen svyco1=(1/-0.0005692)*ln(.585412/.6314361) //132.96//
gen svyco2=(1/-0.00968767)*ln(16.58438/28.33156) //55.28//
gen svyco3=(1/-0.03508482)*ln(5.967343/25.4918) //41.39//

```

```

svy: mean partic nctrips tp_nb tp1 tp2 tp3 tp123 co_nb svyco_nb svyco1 svyco2
svyco3 svyc1_nb svyc11 svyc12 svyc13 tcon1 tcon2 tcon3 svyco_nb nchunto6
ncfish06 if personid2==.

```

## Chapter 2 Stata Codes

```

//This file merges the screening survey data with the wildlife watching sample

```

```

data(merge fh24_96.do)//

```

```

use "G: \fh4_96"

```

```

merge m:1 personid using "G:\fhwar_96 dta\fh2_96" //screening file//
drop if _merge==2 //dropping observations not included in fh4_96 file giving
n=11,759//
drop _merge

```

```

save "G: \fh24_96.dta", replace

```

```

//This file recodes and cleans the data for analysis//

```

```

clear

use "G:\ \fh24_96.dta"

//Investigating participants only//

//Renaming variables//
ren econww cvm1
ren econww2 cvm2
ren wwpick1 bid1
ren wwpick2 bid2
ren econadd wildtrip1
replace wildtrip1=0 if wildtrip1==.
ren econadd2 wildtrip2
replace wildtrip2=0 if wildtrip2==.
ren econest totcost1
ren econest2 totcost2
ren travel_a dist_a
ren travel_b dist_b
ren travel_c dist_c
ren rectrp_a sitetype_a
ren rectrp_b sitetype_b
ren rectrp_c sitetype_c

recode direct_a (1=1 "North")(2=2 "Northeast")(3=3 "East")(4=4 "Southeast")
///
(5=5 "South")(6=6 "Southwest")(7=7 "West")(8=8 "Northwest"),
gen(sitedirect_a)
recode direct_b (1=1 "North")(2=2 "Northeast")(3=3 "East")(4=4 "Southeast")
///
(5=5 "South")(6=6 "Southwest")(7=7 "West")(8=8 "Northwest"),
gen(sitedirect_b)
recode direct_c (1=1 "North")(2=2 "Northeast")(3=3 "East")(4=4 "Southeast")
///
(5=5 "South")(6=6 "Southwest")(7=7 "West")(8=8 "Northwest"),
gen(sitedirect_c)

recode sitetype_a (1=1 "Ocean side")(2=2 "Lake or stream side")(3=3 "Marsh,
wetland, or swamp")(4=4 "Woodland") ///
(5=5 "Brush covered area")(6=6 "Open field")(7=7 "Man made area")(8=8
"Other"), gen(type_a)
recode sitetype_b (1=1 "Ocean side")(2=2 "Lake or stream side")(3=3 "Marsh,
wetland, or swamp")(4=4 "Woodland") ///
(5=5 "Brush covered area")(6=6 "Open field")(7=7 "Man made area")(8=8
"Other"), gen(type_b)
recode sitetype_c (1=1 "Ocean side")(2=2 "Lake or stream side")(3=3 "Marsh,
wetland, or swamp")(4=4 "Woodland") ///

```



```
(5=5 "Brush covered area")(6=6 "Open field")(7=7 "Man made area")(8=8  
"Other"), gen(type_c)
```

```
//Per-trip cost//  
gen cost1=totcost1/wildtrip1  
gen cost2=totcost2/wildtrip2
```

```
gen cendiv_ne=cendiv==1  
gen cendiv_ma=cendiv==2  
gen cendiv_enc=cendiv==3  
gen cendiv_wnc=cendiv==4  
gen cendiv_sa=cendiv==5  
gen cendiv_esc=cendiv==6  
gen cendiv_wsc=cendiv==7  
gen cendiv_mt=cendiv==8  
gen cendiv_pac=cendiv==9
```

```
gen age_10=age/10  
gen agesq=age^2  
gen agesq_10=(age_10)^2  
gen male=1 if sex==1  
replace male=0 if sex==2  
gen white=(race==1)  
gen black=(race==2)  
gen natam=(race==3)  
gen asianpac=(race==4)  
gen othe=(race==5)  
gen hisp=(hispanic==1)  
gen marr=(marital==1)  
gen divor=(marital==3)  
recode school (0/12=1 "High School or less")(13/15=2 "1 to 3 yrs. College")(16=3  
"4 or more/Bachelor's")(17/28=4 "Some grad/prof school or degree"), gen(educ)  
gen retired=1 if retire==1  
replace retired=0 if retire!=1  
replace retire=5 if (job==1 & retire==.)
```

```
//Imputing missing values*/  
replace retire=5 if (income!=. & retire==.) /*have a job/business*/  
replace retire=1 if (age>=65 & retire==.) /*retired*/  
replace retire=4 if retire==. /*doing something else*/  
gen urban=1 if place=="1"  
replace urban=0 if place!="1"
```

```
ren ntrip95 trip95  
replace trip95=0 if trip95==.
```

```

//Creating hunting and fishing variables//
gen hunt_96=1 if (hunt96==1 | ww_hunt==1)
replace hunt_96=0 if (hunt96==0 | ww_hunt==0)
replace hunt_96=0 if hunt_96==. //imputing hunt_96//
gen fish_96=1 if (fish96==1 | ww_fish==1)
replace fish_96=0 if (fish96==0 | ww_fish==0)
replace fish_96=0 if fish_96==. //imputing fish_96//

//Personal income intervals//
gen pinc_max=9999 if income==1
replace pinc_max=19999 if income==2
replace pinc_max=24999 if income==3
replace pinc_max=29999 if income==4
replace pinc_max=34999 if income==5
replace pinc_max=39999 if income==6
replace pinc_max=49999 if income==7
replace pinc_max=74999 if income==8
replace pinc_max=99999 if income==9
replace pinc_max=2000000 if income==10
replace pinc_max=2000000 if income==.

gen pinc_min=1 if income==1
replace pinc_min=10000 if income==2
replace pinc_min=20000 if income==3
replace pinc_min=25000 if income==4
replace pinc_min=30000 if income==5
replace pinc_min=35000 if income==6
replace pinc_min=40000 if income==7
replace pinc_min=50000 if income==8
replace pinc_min=75000 if income==9
replace pinc_min=100000 if income==10
replace pinc_min=1 if income==.

gen lpinc_min=ln(pinc_min)
gen lpinc_max=ln(pinc_max)

//Total number of days spent wildlife watching in resident state//
gen wildday1=wwday1 if resstate==wwstate1
replace wildday1=wwday2 if resstate==wwstate2
replace wildday1=wwday3 if resstate==wwstate3
replace wildday1=wwday4 if resstate==wwstate4
replace wildday1=wwday5 if resstate==wwstate5
replace wildday1=wwday6 if resstate==wwstate6
replace wildday1=0 if wildday==.
replace wildday1=wildtrip1 if (wildday1==0 & wildtrip1>0)
replace wildtrip1=0 if wildtrip1==.

```

```

//Total number of days spent wildlife watching in nonresident state//
gen wildday2=wwday1 if wwst==wwstate1
replace wildday2=wwday2 if wwst==wwstate2
replace wildday2=wwday3 if wwst==wwstate3
replace wildday2=wwday4 if wwst==wwstate4
replace wildday2=wwday5 if wwst==wwstate5
replace wildday2=wwday6 if wwst==wwstate6
replace wildday2=wildtrip2 if (wildday2==0 & wildtrip2>0)
replace wildday2=wildtrip2 if (wildday2==0 & wildtrip2>0)
replace wildtrip2=0 if wildtrip2==.

//Visited private land in resident state//
gen private1=wwpriv1 if resstate==wwstate1
replace private1=wwpriv2 if resstate==wwstate2
replace private1=wwpriv3 if resstate==wwstate3
replace private1=wwpriv4 if resstate==wwstate4
replace private1=wwpriv5 if resstate==wwstate5
replace private1=wwpriv6 if resstate==wwstate6

//Visited private land in nonresident state//
gen private2=wwpriv1 if wwst==wwstate1
replace private2=wwpriv2 if wwst==wwstate2
replace private2=wwpriv3 if wwst==wwstate3
replace private2=wwpriv4 if wwst==wwstate4
replace private2=wwpriv5 if wwst==wwstate5
replace private2=wwpriv6 if wwst==wwstate6

//Visited public land in resident state//
gen public1=wwpub1 if resstate==wwstate1
replace public1=wwpub2 if resstate==wwstate2
replace public1=wwpub3 if resstate==wwstate3
replace public1=wwpub4 if resstate==wwstate4
replace public1=wwpub5 if resstate==wwstate5
replace public1=wwpub6 if resstate==wwstate6

//Visited public land in nonresident state//
gen public2=wwpub1 if wwst==wwstate1
replace public2=wwpub2 if wwst==wwstate2
replace public2=wwpub3 if wwst==wwstate3
replace public2=wwpub4 if wwst==wwstate4
replace public2=wwpub5 if wwst==wwstate5
replace public2=wwpub6 if wwst==wwstate6

//Creating predicted annual personal income//
intreg lpinc_min lpinc_max age agesq male black natam asianpac hisp other
i.retire i.marital school urban i.cendiv, vce(robust)
predict lpinc_hat, e(lpinc_min, lpinc_max)

```

```

gen pinc_hat=exp(lpinc_hat)
sum lpinc_hat pinc_hat, detail

//Creating daily earnings: 250 days//
gen earn_hat=pinc_hat/250
gen oppcost_quart1=.04*earn_hat*(wildday1/wildtrip1) if wildtrip1>0
gen oppcost_quart2=.04*earn_hat*(wildday2/wildtrip2) if wildtrip2>0
gen tottc1=totcost1+oppcost_quart1*wildtrip1
gen tottc2=totcost2+oppcost_quart2*wildtrip2
gen tc1=cost1+oppcost_quart1
gen tc2=cost2+oppcost_quart2
gen pinc_1000=pinc_hat/1000
gen bid1_1000=bid1/1000
gen bid2_1000=bid2/1000
gen bid1_10=bid1/10
gen bid2_10=bid2/10
gen lbid1=ln(bid1)
gen lbid2=ln(bid2)
gen tc1_10=tc1/10
gen tc2_10=tc2/10
gen hypcost1=tottc1+bid1
gen hypcost2=tottc2+bid2
gen hypcost1_1000=hypcost1/1000
gen hypcost2_1000=hypcost2/1000
gen hyptc1=hypcost1/wildtrip1
gen hyptc2=hypcost2/wildtrip2
gen hyptc1_10=hyptc1/10
gen hyptc2_10=hyptc2/10
gen lhypcost1=ln(hypcost1)
gen lhypcost2=ln(hypcost2)
gen tc_pchng1=100*(hypcost1-tottc1)/tottc1
gen tc_pchng2=100*(hypcost2-tottc2)/tottc2
gen dp1=bid1/wildtrip1
gen dp1_10=dp1/10
gen dp2=bid2/wildtrip2
gen dp2_10=dp2/10

// Format data into long form //
reshape long ncstate wildtrip cvm bid totcost cost wildday private public
oppcost_quart tottc tc hypcost hyptc lhypcost tc_pchng dp a b la lb, i(personid
id) j(site)

sum wildtrip if wildtrip>0
gen tc_10=tc/10
gen dp_10=dp/10

```

```
save "G:\fh24_96long.dta"
```

```
//=====//
```

```
//This file estimates the models from Table 3.3//
```

```
use "I:\fh24_96long.dta"
```

```
//Merging with 1996 state level rangeland and forest (1000) acres from NRI//  
merge m:1 ncstate using "I: rangeforest96.dta"
```

```
drop _merge  
gen bid_10=bid/10  
gen bid_1000=bid/1000  
gen land=(range+forest)/(pop96/1000)  
gen land2=(range+forest)/(totarea)  
sum land land2
```

```
gen outstate=1 if site==1  
replace outstate=0 if site==2  
foreach X of varlist wildtrip tc tc_10 pinc_1000 marr age_10 male bid cvm  
private public ///  
{  
drop if (`X'==.)  
}  
drop if wildtrip==0  
drop if ncstate=="DC"
```

```
//Table 3.2//  
sum wildtrip tc tc_10 pinc_1000 pinc_hat marr age age_10 male photo parks  
bid cvm private public land land2 oppcost_quart hunt_96 fish_96 ///  
wildtrip tc_10 pinc_1000 age_10 marr male photo parks private land2 hunt_96  
fish_96  
drop if tc>500  
drop if wildtrip>250
```

```
//Canned procedure for ZTNB//
```

```
tnbreg wildtrip tc_10 pinc_1000 age_10 marr male photo parks private land2 ///  
hunt_96 fish_96, cluster(personid) tech(bfgs nr)
```

```
est sto ZTNB  
predict cmu_hat, cm  
sum wildtrip cmu_hat  
di "E[CS] = " -10/_b[tc_10]
```

```

//Program for Censored(threshold) normal model (Cameron(1991))//
capture program drop cnormal
program define cnormal
args ln1 b0 b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 s
tempvar xb pi sigma

//WTP model//
qui gen double
`xb'=`b0'+`b1'*pinc_1000+`b2'*age_10+`b3'*marr+`b4'*male+`b5'*photo ///
+`b6'*parks+`b7'*private+`b8'*land2+`b9'*hunt_96+`b10'*fish_96

//Defining Sigma//
qui gen double `sigma'=exp(`s')

//Defining CVM probit//
qui gen double `pi'=normprob((`xb'-bid_1000)/`sigma')

//Log-likelihood Function//
qui replace `lnl' = cvm*ln(`pi')+(1-cvm)*ln(1-`pi')

end

ml model lf cnormal (wtp_cons:) (wtp_pinc_1000:) (wtp_age_10:)
(wtp_marr:) (wtp_male:) (wtp_photo:) ///
(wtp_parks:) (wtp_private:) (wtp_land2:) (wtp_hunt96:) (wtp_fish96:)
(s:), technique(nr bfgs) cluster(personid)
*ml check
ml init -.1754269 .0036068 .0187455 -.0414852 .027948 .1498656
.0860281 .162887 -.0267908 .0080004 -.0555778 -.4227889, copy
ml search
ml maximize, difficult /*trace*/

di "sigma = " exp([s]_cons)
gen
wtp_cn=[wtp_cons]_cons+[wtp_pinc_1000]_cons*pinc_1000+[wtp_age_10]_
cons*age_10+[wtp_male]_cons*male+[wtp_marr]_cons*marr ///
+[wtp_photo]_cons*photo+[wtp_parks]_cons*parks+[wtp_private]_con ///
s*private+[wtp_land2]_cons*land2+[wtp_hunt96]_cons*hunt_96+ ///
[wtp_fish96]_cons*fish_96
gen wtp=wtp_cn*1000
sum wtp_cn wtp
est sto THRESHOLD

//Program for Joint TCM-CVM model//
capture program drop utility1
program define utility1

```

```

args ln1 a0 a1 a2 a3 a4 a5 a6 a7 a8 a9 a10 a11 b0 b1 b2 b3 b4 b5 b6 ///
b7 b8 b9 b10 /*b11*/ lalpha s kap
tempvar xb exb zb po cmean cvar z sigma rho pi f

//Count(TCM) model//
qui gen double
`xb'=`a0'+`a1'*tc_10+`a2'*pinc_1000+`a3'*age_10+`a4'*marr+`a5'*male+`a6'
*photo+`a7'*parks+`a8'*private+`a9'*land2+`a10'*hunt_96+`a11'*fish_96
qui gen double `exb' = exp(`xb')

//Prob(y>0)//
qui gen double `po'=1/((1+exp(`lalpha')*`exb')^(1/exp(`lalpha'))))

//E(y|y>0)//
qui gen double `cmean'=`exb'/(1-`po')

//Var(y|y>0)//
qui gen double `cvar'=(`cmean'/((`po')^exp(`lalpha')))*(1-
`cmean'*(`po')^(1+exp(`lalpha'))))

*qui gen double `z'=(wildtrip-`cmean')/(sqrt(`cvar'))
qui gen double `z'=sqrt(((wildtrip-`cmean')^2)/(`cvar'))

//Probit regression//
qui gen double
`zb'=`b0'+`b1'*pinc_1000+`b2'*age_10+`b3'*marr+`b4'*male+`b5'*photo+`b
6'*parks+`b7'*private+`b8'*land2+`b9'*hunt_96+`b10'*fish_96

//Defining Sigma//
qui gen double `sigma'=exp(`s')

//Defining rho//
qui gen double `rho'=tanh(`kap') //hyperbolic tangent function//

//Defining CVM probit//
*qui gen double `pi'=normprob((`zb'+`sigma'*`rho'*`z')/(`sigma'*sqrt(1-
(`rho')^2)))
qui gen double `pi'=normprob((bid_1000-`zb'-
`sigma'*`rho'*`z')/(`sigma'*sqrt(1-(`rho')^2)))

//Defining TCM Zero-truncated NB2//
qui gen double `f'=exp(lngamma(wildtrip+1/exp(`lalpha')))-
lnfactorial(wildtrip)-lngamma(1/exp(`lalpha')) ///
+wildtrip*ln(exp(`lalpha')*`exb')-
(wildtrip+1/exp(`lalpha'))*ln(1+exp(`lalpha')*`exb')-ln(1-`po'))

//Log-likelihood Function//

```

```

qui replace `lnl' = (1-cvm)*ln(`pi')+cvm*ln(1-`pi')+ln(`f')
*qui replace `lnl' = cvm*ln(`pi')+(1-cvm)*ln(1-`pi')+ln(`f')

end

ml model lf utility1 (ztnb_cons:) (ztnb_tc_10:) (ztnb_pinc_1000:)
(ztnb_age_10:) (ztnb_marr:) (ztnb_male:) ///
(ztnb_photo:) (ztnb_parks:) (ztnb_private:) (ztnb_land2:)
(ztnb_hunt96:) (ztnb_fish96:) ///
(prob_cons:)/*(prob_bid_10:)/(prob_pinc_1000:) (prob_age_10:)
(prob_marr:) (prob_male:) ///
(prob_photo:) (prob_parks:) (prob_private:) (prob_land2:)
(prob_hunt96:) (prob_fish96:) ///
(alpha:) (s:) (kap:), technique(nr) cluster(personid)

ml init -.6214306 -.1537789 -.0013296 .2047646 -.2699383
.1759307 .3058146 .4501337 .791881 .1494254 .4504297
.1723992 -.1754328 .0036069 .0187457 -.041486 .0279484
.1498675 .0860289 .1628891 -.0267907 .0080003 -.0555786
2.402233 -.4227794 .01, copy

ml maximize, trace

est sto JOINT_CORR

di "alpha = " exp([alpha]_cons)
di "sigma = " exp([s]_cons)
di "rho = " tanh([kap]_cons)
di "Joint ZTNB2/Utility tc beta = " [ztnb_tc_10]_cons
di "Joint ZTNB2/Utility E[CS] = " -10/[ztnb_tc_10]_cons

gen
etrip_util=exp([ztnb_cons]_cons+[ztnb_tc_10]_cons*tc_10+[ztnb_pinc_1000]
_cons*pinc_1000+[ztnb_age_10]_cons*age_10+[ztnb_male]_cons*male+[ztnb
_marr]_cons*marr ///
+[ztnb_photo]_cons*photo+[ztnb_parks]_cons*parks+[ztnb_private]_cons*pri
vate+[ztnb_land2]_cons*land2+[ztnb_hunt96]_cons*hunt_96+[ztnb_fish96]_
cons*fish_96)

gen pro=1/(1+exp([alpha]_cons)*etrip_util)^(1/exp([alpha]_cons))

gen etrip_tr=etrip_util/(1-pro)

sum wildtrip cmu_hat etrip_tr, detail

```



```

gen
wtp_joint=1000*([prob_cons]_cons+[prob_pinc_1000]_cons*pinc_1000+[pro
b_age_10]_cons*age_10+[prob_male]_cons*male+[prob_marr]_cons*marr ///
+[prob_photo]_cons*photo+[prob_parks]_cons*parks+[prob_private]_cons*pr
ivate+[prob_land2]_cons*land2+[prob_hunt96]_cons*hunt_96+[prob_fish96]
_cons*fish_96)

```

```

sum wtp_joint

```

```

//Program for Utility-consistent model//

```

```

capture program drop utility1
program define utility1
args lnl a0 a1 a2 a3 a4 a5 a6 a7 a8 a9 a10 a11 lalpha s kap
tempvar xb exb uo u1 du po cmean cvar z sigma rho pi f

//Count(TCM) model//
qui gen double
`xb'=`a0'+`a1'*tc_10+`a2'*pinc_1000+`a3'*age_10+`a4'*marr+`a5'*male+`a6'
*photo+`a7'*parks+`a8'*private+`a9'*land2+`a10'*hunt_96+`a11'*fish_96
qui gen double `exb' = exp(`xb')

//Prob(y>0)//
qui gen double `po'=1/((1+exp(`lalpha')*`exb')^(1/exp(`lalpha'))))

//E(y|y>0)//
qui gen double `cmean'=`exb'/(1-`po')

//Var(y|y>0)//
qui gen double `cvar'=(`cmean'/((`po')^exp(`lalpha')))*(1-
`cmean'*(`po')^(1+exp(`lalpha'))))

qui gen double `z'=(wildtrip-`cmean')/(sqrt(`cvar'))

//Utility Difference//
qui gen double `uo'=(-1/(`a2'*exp(`a2'*pinc_1000)))
qui gen double `u1'=(-`a1'-
`a2'*wildtrip*(1+`a1'*(dp_10)))/(`a1'*`a2'*(exp(`a2'*pinc_1000))) //dp_10 =
hyptc_10-tc_10//
qui gen double `du'=-
wildtrip*(1+`a1'*dp_10)/(`a1'*exp(`a2'*pinc_1000))

//Defining Sigma//
qui gen double `sigma'=exp(`s')

//Defining rho//
qui gen double `rho'=tanh(`kap') //hyperbolic tangent function//

```

```

//Defining CVM probit//
qui gen double `pi'=normprob((`du'+`sigma'*`rho'*`z')/(`sigma'*sqrt(1-
(`rho')^2)))

//Defining TCM Zero-truncated NB2//
qui gen double `f'=exp(lgamma(wildtrip+1/exp(`alpha'))-
lnfactorial(wildtrip)-lgamma(1/exp(`alpha'))) ///
+wildtrip*ln(exp(`alpha')*`exb')-
(wildtrip+1/exp(`alpha'))*ln(1+exp(`alpha')*`exb')-ln(1-`po'))

//Log-likelihood Function//
*qui replace `lnl' = (1-cvm)*ln(`pi')+cvm*ln(1-`pi')+ln(`f)
qui replace `lnl' = cvm*ln(`pi')+(1-cvm)*ln(1-`pi')+ln(`f)

end

ml model lf utility1 (ztnb_cons:) (ztnb_tc_10:) (ztnb_pinc_1000:)
(ztnb_age_10:) (ztnb_marr:) (ztnb_male:) ///
(ztnb_photo:) (ztnb_parks:) (ztnb_private:) (ztnb_land2:)
(ztnb_hunt96:) (ztnb_fish96:) (alpha:) (s:) (kap:), technique(bfgs nr)
cluster(personid)
*ml check
ml init -.3801276 -.1432526 -.0025163 .1926361 -.2558721 .1261794
.2276727 .4064981 .6874352 .1471791 .3692333 .2148833 2.401282 6.970658 -
.200626, copy
*ml search
ml maximize, difficult /*trace*/

est sto ZTNB_RESUTILITY

di "alpha = " exp([alpha]_cons)
di "sigma = " exp([s]_cons)
di "rho = " tanh([kap]_cons)
di "Joint ZTNB2/Utility tc beta = " [ztnb_tc_10]_cons
di "Joint ZTNB2/Utility E[CS] = " -10/[ztnb_tc_10]_cons

gen
etrip_util=exp([ztnb_cons]_cons+[ztnb_tc_10]_cons*tc_10+[ztnb_pinc_1000]
_cons*pinc_1000+[ztnb_age_10]_cons*age_10+[ztnb_male]_cons*male+[ztnb
_marr]_cons*marr ///
+[ztnb_photo]_cons*photo+[ztnb_parks]_cons*parks+[ztnb_private]_cons*pri
vate+[ztnb_land2]_cons*land2+[ztnb_hunt96]_cons*hunt_96+[ztnb_fish96]_
cons*fish_96)
gen pro=1/(1+exp([alpha]_cons)*etrip_util)^(1/exp([alpha]_cons))
gen etrip_tr=etrip_util/(1-pro)

```

```
sum wildtrip cmu_hat etrip_tr, detail
```

```
//Table 3.3//
```

```
esttab ZTNB THRESHOLD JOINT_CORR ZTNB_RESUTILITY, b(%12.4f) t  
stats(N ll aic bic) mtitles parentheses star(* .10 ** .05 *** .01 **** .001)
```

### Chapter 3 Stata Codes

```
//=====Double-Bound Regression Simulation=====//
```

```
clear all
```

```
set obs 1000
```

```
set seed 12345
```

```
gen bido = 5*uniform() //range is [0,5] w/ mean=2.5//
```

```
gen bid1 = 2*bido
```

```
gen bid2 = bido/2
```

```
gen x = 5*uniform()
```

```
gen z = 5*uniform()
```

```
gen wtp_fixed=10+1.5*x-2.5*z //wtp_true=10+1.5*(2.5)-2.5*(2.5)=7.5//
```

```
gen double xb = 5 + .75*x - 1.25*z
```

```
sum bido bid1 bid2 x z wtp_fixed
```

```
//Save data to be called in simulation program//
```

```
save "D:/ doubleb.dta", replace
```

```
//Program to be called for simulation//
```

```
program sim1, rclass
```

```
version 12.0
```

```
use "D:\doubleb.dta", clear
```

```
gen u=rnormal(0,1)
```

```
gen byte YY = (-.5*bid1+xb+u) >= 0
```

```
gen byte NN = (-.5*bid2+xb+u) < 0
```

```
gen byte YN = ((-.5*bido+xb+u) >= 0) & ((-.5*bid1+xb+u) < 0)
```

```
gen byte NY = ((-.5*bido+xb+u) < 0) & ((-.5*bid2+xb+u) >= 0)
```

```
//Creating responses//
```

```
gen r1=1 if (YY ==1 | YN ==1)
```

```
replace r1 =0 if r1 ==.
```

```
gen r2 =1 if (NY ==1 | YY ==1)
```

```
replace r2 =0 if r2 ==.
```

```

//Creating first and second bids for doubleb regression//
gen b1 =bido
gen b2 =bid1 if (YYr==1 | YN ==1)
replace b2 =bid2 if (NY ==1 | NN ==1)

//Creating random missing r2//
*replace r2 =. in 901/1000
*replace r2 =. in 751/1000
*replace r2 =. in 501/1000

//Creating systematic missing r2 at largest/smallest 10,25,50 percent//

//dropping largest x//
*sort x
*replace r2 =. in 901/1000
*replace r2 =. in 751/1000
*replace r2 =. in 501/1000

//dropping largest z//
*sort z
*replace r2 =. in 901/1000
*replace r2 =. in 751/1000
*replace r2 =. in 501/1000

//Creating missing adjustments(recoding)//
replace r2 =1 in 501/1000 if r1 ==1
replace b1 =bid2 in 501/1000 if r1 ==1
replace b2 =bido in 501/1000 if r1 ==1
replace r2 =0 in 501/1000 if r1 ==0
replace b1 =bid1 in 501/1000 if r1 ==0
replace b2 =bido in 501/1000 if r1 ==0

doubleb b1 b2 r1 r2 x z

return scalar b=_b[Beta:_cons]
return scalar bx=_b[Beta:x]
return scalar bz=_b[Beta:z]
return scalar s=_b[Sigma:_cons]
return scalar wtp=( _b[Beta:_cons]+_b[Beta:x]*(2.5)+_b[Beta:z]*(2.5))
return scalar se_b=( _b[Beta:_cons]-10)^2
return scalar se_bx=( _b[Beta:x]-1.5)^2
return scalar se_bz=( _b[Beta:z]+2.5)^2
return scalar se_s=( _b[Sigma:_cons]-2)^2

```

```
return scalar se_wtp=((_b[Beta:_cons]+_b[Beta:x]*(2.5)+_b[Beta:z]*(2.5))-  
7.5)^2
```

```
end
```

```
simulate s=r(s) b=r(b) bx=r(bx) bz=r(bz) wtp=r(wtp) se_b=r(se_b)  
se_bx=r(se_bx) se_bz=r(se_bz) se_s=r(se_s) se_wtp=r(se_wtp), reps(5000):  
sim1
```

```
//Tables 4.1 to 4.3//  
//Root mean squared error//  
sum b bx bz s wtp  
sum se_b  
dis (r(mean)^.5)  
sum se_bx  
dis (r(mean)^.5)  
sum se_bz  
dis (r(mean)^.5)  
sum se_s  
dis (r(mean)^.5)  
sum se_wtp  
*dis r(mean)  
dis (r(mean)^.5)
```

## REFERENCES

- Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management*, 26, 271-292.
- Aiken, R. (July 2009). *Net economic values of wildlife-related recreation in 2006*. Division of Policy and Programs, Wildlife and Sport Fish Restoration Programs, U.S. Fish and Wildlife Service, Washington, D.C.
- Aiken, R., & Rouche, G. P. (September 2003). *Net economic values for wildlife-related recreation in 2001*. Washington D.C.: Division of Federal Aid, U.S. Fish and Wildlife Service.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Alberini, A., Kanninen, b., & Carson, R. (1997). Modeling Response Incentive Effects in Dichotomous Choice Contingent Valuation Data. *Land Economics*, 73, 309-324.
- An, D., & Little, R. J. (2007). Multiple imputation: An alternative to top coding for statistical disclosure control. *Journal of the Royal Statistical Society: Series A*, 170, 923-940.

- Azevedo, C., Herriges, J., & Kling, C. (2003). Combining revealed and stated preferences: Consistency tests and their interpretations. *American Journal of Agriculture Economics*, 85, 525-537.
- Baerenklau, K. (2010). A latent class approach to modeling endogenous spatial sorting in zonal recreation demand models. *Land Economics*, 86(4), 800-816.
- Beatty, T. K., Brozovic, N., & Ward, M. B. (July 24-27, 2005). Consumer surplus estimates and the source of regression error. *Selected paper for presentation at the American Agricultural Economics Association Annual Meeting*. Providence, RI.
- Bhat, C. (1994). Imputing a Continuous Income Variable from Grouped and Missing Income Observations. *Economics Letters*, 46, 311-319.
- Bishop, R. C. (1978). Endangered species and uncertainty: The economics of a safe minimum standard. *American Journal of Agricultural Economics*, 60, 10-18.
- Bishop, R., & Heberlein, T. (1979). Measuring values of extramarket goods: Are indirect measures biased? *American Journal of Agricultural Economics*, 61, 926-930.
- Bishop, R., & Heberlein, T. (1986). Does contingent valuation work? In R. Cummings, D. Brookshire, & e. W. Schulze, *A state of the art assessment of the contingent valuation method*. Totowa, NJ: Rowman and Allenheld.

- Bockstael, N., & Strand Jr., I. E. (1987). The effect of common sources of regression error on benefit estimates. *Land Economics*, 63(1), 11-20.
- Bockstael, N., Strand, I. E., & Hannemann, W. (1987). Time and the recreational demand model. *American Journal of Agricultural Economics*, 69, 298-302.
- Bohara, A. K., Kerkvliet, J., & Berrens, R. P. (2001). Addressing negative willingness to pay in dichotomous choice contingent valuation: A monte carlo simulation. *Environmental and Resource Economics*, 20, 173-195.
- Boxall, P., & Adamowicz, W. (2002). Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and Resource Economics*, 23(4), 421-446.
- Boyle, K. J., Roach, B., & Waddington, D. G. (August 1998). *1996 net economic values for Bass, Trout and Walleye fishing, Deer, Elk and Moose hunting, and wildlife watching*. Washington D.C.: Division of Federal Aid, U.S. Fish and Wildlife Service.
- Bozdogan, H. (1987). Model selection and akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345-370.
- Brefle, W., & Morey, E. (2000). Investigating preference heterogeneity in a repeated discrete choice recreatio demand model of Atlantic Salmon fishing. *Marine Resource Economics*, 15, 1-20.



- Brown, J. G. (2002). Using a multiple imputation technique to merge data sets. *Applied Economics Letters*, 9, 311-314.
- Brox, J. A., Kumar, R. C., & Stollery, K. R. (2003). Estimating willingness to pay for improved water quality in the presence of item nonresponse bias. *American Journal of Agricultural Economics*, 85(2), 414-428.
- Cameron, A. C., & Trivedi, P. K. (1998). *Regression Analysis of Count Data*. New York, NY: Cambridge University Press.
- Cameron, T. A. (1988). A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics and Management*, 15, 355-379.
- Cameron, T. A. (1992). Combining contingent valuation and travel cost data for the valuation of nonmarket goods. *Land Economics*, 68(3), 302-317.
- Cameron, T. A., & Englin, J. (1997). Respondent experience and contingent valuation of environmental goods. *Journal of Environmental Economics and Management*, 33(3), 296-313.
- Cameron, T. A., & Quiggin, J. (1994). Estimation using contingent valuation data from a "dichotomous choice with follow-up" questionnaire. *Journal of Environmental Economics and Management*, 27(3), 218-234.

- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification* , 13, 195-212.
- Cesario, F., & Knetsch, J. (1976). Time bias in recreation benefit estimates. *Land Economics*, 52, 32-41.
- Chen, H., & Cosselett, S. (1998). Environmental quality preference and benefit estimation in multinomial probit models: A simulation approach. *American Journal of Agricultural Economics*, 80, 512-520.
- Cicchetti, C. J. (1973). *Forecasting recreation in the United States*. Lexington, MA: Lexington Books.
- Cooper, J., Hanemann, M., & Signorello, G. (2002). One-and-One-Half Bound Dichotomous Choice Contingent Valuation. *Review of Economics and Statistics*, 84, 742-756.
- Cragg, J. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39, 829-844.
- Creel, M. D., & Loomis, J. B. (1990). Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in california. *American Journal of Agricultural Economics*, 72(2), 434-441.
- Cummings, R. G., Brookshire, D. S., & Schulze, W. D. (1986). *Valuing environmental goods: A state of the arts assessment of the contingent valuation method*. Totwa, NJ: Roweman and Allanheld.

- Dalecki, M., Ilvento, T., & Moore, D. (1988). The effects of multiwave mailings on the external validity of mail surveys. *Journal of the Community Development Society, 19*, 51-70.
- Davis, D., Tisdell, C., & Hardy, M. (2001). *The role of economics in managing wildlife tourism*. Gold Coast: CRC for Sustainable Tourism.
- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1-38.
- Dhazn, G., Woodward, R., Ozuna Jr., T., & Griffin, W. L. (2003). Joint estimation of revealed and stated preference data: An application to recreational red snapper valuation. *Agricultural and Resource Economics Review, 32*(2), 209-221.
- Dias, J. G., & Vermunt, J. K. (2006). Bootstrap methods for measuring classification uncertainty in latent class analysis. In A. Rizzi, & M. Vichi (Ed.), *Comstat 2006- Proceedings in Computational Statistics* (pp. Part 1, 31-41). Rome: Physica-Verlag Heidelberg.
- Dias, J. G., & Vermunt, J. K. (2008). A bootstrap-based aggregate classifier for model-based clustering. *Computational Statistics, 23*(4), 643-659.
- Durrant, G. B., & Skinner, C. (2006). Using data augmentation to correct for non-ignorable non-response when surrogate data are available: An application

to the distribution of hourly pay. *Journal of the Royal Statistical Society: Series A*, 169, 605-623.

English, D. B., & Bowker, J. (1996). Sensitivity of whitewater rafting consumer surplus to pecuniary travel cost specifications. *Journal of Environmental Management*, 47, 79-91.

English, E. (2008). Recreation nonparticipation as choice behavior rather than statistical outcome. *American Journal of Agricultural Economics*, 90(1), 186-196.

Eom, Y., & Larson, D. M. (2006). Improving environmental valuation estimates through consistent use of revealed and stated preference information. *Journal of Environmental Economics and Management*, 52, 501-516.

Glynn, R. J., Laird, N. M., & Rubin, D. B. (1993). Multiple imputation in mixture models for nonignorable nonresponse with follow-ups. *Journal of the American Statistical Association*, 88(423), 984-993.

González-Sepúlveda, J. M. (2008). *Challenges and solutions in combining RP and SP data to value recreation*. Dissertation, Colorado State University, Department of Agricultural and Resource Economics, Fort Collins, CO.

González-Sepúlveda, J. M., Loomis, J. B., & González-Cabán, A. (2008). 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*, 40(2), 681-695.

Greene, W. H. (2003). *Econometric Analysis* (5th Edition ed.). Upper Saddle River, New Jersey: Pearson Education, Inc.

Gurmu, S., & Trivedi, P. K. (1996). Excess zeros in count models for recreation trips. *Journal of Business & Economic Statistics*, 14(4), 469-477.

Haab, T. C., & McConnell, K. E. (1996). Count data models and the problem of zeros in recreation demand analysis. *American Journal of Agricultural Economics*, 78, 89-102.

Habb, T. C., & McConnell, K. E. (1998). Referendum models and economic values: Theoretical, intuitive, and practical bounds on willingness to pay. *Land Economics*, 74(2), 216-229.

Habb, T. C., & McConnell, K. E. (2002). *Valuing environmental and natural resources: The econometrics of non-market valuation*. Northhampton, MA: Edward Elgar.

Halstead, J., Luloff, A., & Stevens, T. H. (1992). Protest bidders in contingent valuation. *Northeastern Journal of Agricultural and Resource Economics*, 21(2), 160-169.

Hanemann, W. M., Loomis, J., & Kanninen, B. (1991). Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics*, 73(4), 1255-1263.

- Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics*, 66, 322-341.
- Hay, M. J., & McConnell, K. E. (1979). An analysis of participation in nonconsumptive wildlife recreation. *Land Economics*, 55(4), 460-471.
- Hay, M. J., & McConnell, K. E. (1984). Harvesting and nonconsumptive wildlife recreation decisions. *Land Economics*, 60(4), 388-396.
- Hay, M. J., & McConnell, K. E. (1981). An analysis of participation in nonconsumptive wildlife recreation: Reply. *Land Economics*, 57(2), 288-292.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161.
- Hellerstein, D. M. (1991). Using count data models in travel cost analysis with aggregate data. *American Journal of Agricultural Economics*, 73, 860-867.
- Hellerstein, D. M. (1992). Correcting for bias when average values are used to compute changes in consumer surplus. *Leisure Sciences*, 14(4), 337-345.
- Hellerstein, D. M. (1998). An analysis of wildlife recreation using the FHWAR. *1998 W-133 Proceedings*. Draft 5/7/98.

- Hellerstein, D. M., & Mendelsohn, R. (1993). A theoretical foundation for count data models. *American Journal of Agricultural Economics*, 75, 604-611.
- Herriges, J. A., & Kling, C. L. (1999). *Valuing recreation and the environmental: Revealed preference methods in theory and practice*. Cheltenham, UK: Edward Elgar.
- Hilbe, J. M. (2011). *Negative Binomial Regression* (2nd ed.). New York, NY: Cambridge University Press.
- Holmes, T. P., & Englin, J. E. (2010). Preference heterogeneity in a count data model of demand for off-highway vehicle recreation. *Agricultural and Resource Economics Review*, 39(1), 75-88.
- Horowitz, J., & Manski, C. F. (1998). Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations. *Journal of Econometrics*, 84(1), 37-58.
- Huang, J., Habb, T. C., & Whitehead, J. C. (1997). Willingness to pay for quality improvements: Should revealed and stated preference data be combined? *Journal of Environmental Economics and Management*, 34(3), 240-255.
- Hwang, J. W. (2012). Temporal reliability test of nonconsumptive wildlife recreation benefits constructed from choke price data. *Economics Bulletin*, 32(1), 788-798.
- Hynes, S., Hanley, N., & Scarpa, R. (2008). Effects on welfare measures of alternative means of accounting for preference heterogeneity in

recreational demand models. *American Journal of Agricultural Economics*, 90(4), 1011-1027.

Johnson, N. L., & Kotz, S. (1969). *Distributions in statistics*. New York, NY: John Wiley & Sons.

Kofman, P., & Sharpe, I. G. (2003). Using multiple imputation in the analysis of incomplete observations in finance. *Journal of Financial Econometrics*, 1(2), 216-249.

Kohnen, C., & Reiter, J. P. (2009). Multiple imputation for combining confidential data owned by two agencies. *Journal of the Royal Statistical Society: Series A*, 172, 511-528.

Korinek, A., Mistiaen, J. A., & Ravallion, M. (2007). An Econometric Method of Correcting for Unit Nonresponse Bias in Surveys. *Journal of Econometrics*, 136, 213-235.

Krinsky, I., & Robb, A. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 68, 715-719.

Krinsky, I., & Robb, A. (1990). On approximating the statistical properties of elasticities: A correction. *The Review of Economics and Statistics*, 72, 189-190.

Kuriyama, K., Hanemann, W. M., & Hilger, J. R. (2010). A latent segmentation approach to a Kuhn-Tucker model: An application to recreation demand. *Journal of Environmental Economics and Management*, 60, 209-220.



- Landry, C. E., & Liu, H. (2009). A semi-parametric estimator for revealed and stated preference data: An application to recreational beach visitation. *Journal of Environmental Economics and Management*, 57, 205-218.
- Langford, I. H., Bateman, I. J., & Langford, H. D. (1996, April). A Multilevel Modelling Approach to Triple-Bounded Dichotomous Choice Contingent Valuation. *Environmental and Resource Economics*, 7, 197-211.
- Leonard, J. (July 2008). *Wildlife Watching in the U.S.: The Economic Impacts on National and State Economies in 2006: Addendum to the 2006 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation*. Report 2006-1, U.S. Fish and Wildlife Service, Wildlife and Sport Fish Restoration Programs, Arlington, VA.
- Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys*. New York, NY: John Wiley & Sons, Inc.
- Li, H., Berrens, R. P., Bohara, A. K., Jenkins-Smith, H. C., Silva, C. L., & Weimer, D. L. (2005). Exploring the beta model using proportional budget information in a contingent valuation study. *Economics Bulletin*, 17(8), 1-9.
- Little, R. J. (1988). Missing-Data Adjustments in Large Surveys. *Journal of Business and Economic Statistics*, 6(3), 287-296.
- Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: John Wiley.

- Loomis, J. C. (1997). Panel estimators to combine revealed and stated preference dichotomous choice data. *Journal of Agricultural and Resource Economics*, 22, 233-245.
- Lourenço, O. D., Ferreira, P. L., & Barros, P. P. (2006, August). Econometric analysis of health care utilization: An alternative hurdle using latent class models. Center of Studies and Research in Health of the University of Coimbra.
- Majumdar, S. (2011). Consumptive nature-based recreation in the United States: Welfare measures from truncated and untruncated count data models. In *Three essays on the valuation of natural resources in recreation and tourism*. (Doctoral Dissertation), Auburn University.
- Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic Theory*. New York, NY: Oxford University Press.
- McConnell, K. (1990). Models for referendum data: The structure of discrete choice models for contingent valuation. *Journal of Environmental Economics and Management*, 18, 19-34.
- McConnell, K., & Strand, I. (1981). Measuring the cost of time in recreation demand analysis: An application to sportsfishing. *American Journal of Agricultural Economics*, 63, 153-156.
- McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. New York, NY: John Wiley & Sons.

- Messonnier, M. L., Bergstrom, J. C., Cornwell, C. M., Teasley, R. J., & Cordell, H. K. (2000, May). Survey Response-Related Biases in Contingent Valuation: Concepts, Remedies, and Empirical Application to Valuing Aquatic Plant Management. *American Journal of Agricultural Economics*, 83, 438-450.
- Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods: The contingent valuation method*. New York, NY: Resource for the Future.
- Moeltner, K. (2003). Addressing aggregation bias in zonal recreation models. *Journal of Environmental Economics and Management*, 45(1), 128-144.
- Moeltner, K., & von Haefen, R. (2011). Microeconomic strategies for dealing with unobservable and endogenous variables in recreation demand models. *Annual Review of Resource Economics*, 3, 375-396.
- Morey, E., Thacher, J., & Breffle, W. (2006). Using Anger Characteristics and Attitudinal Data to Identify Environmental Preference Classes: A Latent-Class Model. *Environmental & Resource Economics*, 34, 91-115.
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33, 341-365.
- Parsons, G. R. (1991). A note on choice of residential location in travel cost demand models. *Land Economics*, 67(3), 360-364.
- Phaneuf, D. J., Kling, C. L., & Herriges, J. A. (2000). Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand. *Review of Economics and Statistics*, 82, 83-92.

- Phaneuf, D. J., Taylor, L. O., & Braden, J. B. (2013). Combining revealed and stated preference data to estimate preferences for residential amenities: A GMM approach. *Land Economics*, 89(1), 30-52.
- Provencher, B., Baerenklau, K. A., & Bishop, R. C. (2002). A finite mixture logit model of recreational angling with serially correlated random utility. *American Journal of Agricultural Economics*, 84(4), 1066-1075.
- Pudney, S. (1989). *Modeling individual choice: The econometrics of corners, kinks and holes*. New York, NY: Basil Blackwell.
- Puhani, P. (2000). The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Surveys*, 14(1), 53-68.
- Riphahn, R. T., & Serfling, O. (2005). Item Non-Response on Income and Wealth Questions. *Empirical Economics*, 30, 521-538.
- Rockel, M. L., & Kealy, M. J. (1991). The value of nonconsumptive wildlife recreation in the United States. *Land Economics*, 67(4), 422-434.
- Rubin, D. B. (1986). Statistical Matching using File Concatenation with Adjusted Weights and Multiple Imputations. *Journal of Business and Economic Statistics*, 4, 87-94.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons.

- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434), 473-789.
- Scarpa, R., & Thiene, M. (2005). Destination choice models for rock climbing in the Northeastern Alps: A latent-class approach based on intensity of preferences. *Land Economics*, 81(3), 426-444.
- Scarpa, R., Thiene, M., & Tempesta, T. (2007). Latent class count models of total visitation demand: Days out hiking in the eastern Alps. *Environmental and Resource Economics*, 38, 447-460.
- Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8, 3-15.
- Schenker, N., & Taylor, J. (1996). Partially parametric techniques for multiple imputation. *Computational Statistics and Data Analysis*, 22, 425-446.
- Schenker, N., Raghunathan, T. E., Chiu, P.-L., Makuc, D. M., Zhang, G., & Cohen, A. J. (2006). Multiple imputation of missing income data in the national health interview survey. *Journal of the American Statistical Association*, 101(475), 924-933.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.
- Shonkwiler, J., & Shaw, W. D. (1996). Hurdle count-data models in recreation demand analysis. *Journal of Agricultural and Resource Economics*, 21(2), 210-219.

- Skinner, C. (1989). Domain means, regression and multivariate analysis. In D. H. C.J. Skinner (Ed.), *Analysis of complex surveys* (pp. 59-87). New York, NY: Wiley.
- Smith, V. K., Desvousges, W. H., & Fisher, A. (1986). A comparison of direct and indirect methods for estimating environmental benefits. *American Journal of Agricultural Economics*, 68(2), 280-290.
- StataCorp LP. (2011). *Stata user's guide*. College Station, TX: Stata Press.
- Tapper, R. (2006). *Wildlife watching and tourism: A study on the benefits and risks of a fast growing tourism activity and its impacts on species*. UNEP/CMS Convention on Migratory Species and TUI. Bonn, Germany: United Nations Environment Programme and the Secretariat of the Convention on the Conservation of Migratory Species of Wild Animals.
- Train, K. E. (1998). Recreation demand models with taste differences over people. *Land Economics*, 74(2), 230-239.
- U.S Department of Agriculture. (2009). *Hawaii Supplement (2012, 7 pages) to Summary Report: 2007 National Resources Inventory*. Natural Resources Conservation Service, Washington DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa. Retrieved from [http://www.nrcs.usda.gov/Internet/FSE\\_DOCUMENTS/stelprdb1041379.pdf](http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1041379.pdf)

- U.S. Department of Agriculture. (2009). *2007 National Resources Inventory*. Natural Resources Conservation Service, Washington DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa. Retrieved from [http://www.nrcs.usda.gov/technical/NRI/2007/2007\\_NRI\\_Summary.pdf](http://www.nrcs.usda.gov/technical/NRI/2007/2007_NRI_Summary.pdf)
- U.S. Department of Agriculture. (2012). *Alaska Summary Report: 2007 National Resources Inventory*. Natural Resources Conservation Service, Washington DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa. Retrieved from [http://www.nrcs.usda.gov/Internet/FSE\\_DOCUMENTS/stelprdb1047451.pdf](http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1047451.pdf)
- U.S. Department of the Interior, Fish & Wildlife Service and U.S Department of Commerce, Bureau of the Census. (December 2012). *2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation*. Washington DC: U.S. Government Printing Office.
- U.S. Department of the Interior, Fish & Wildlife Service and U.S Department of Commerce, Bureau of the Census. (Winter 2012). *Celebrating the wildlife and sport fish restoration program*. Washington DC: U.S. Government Printing Office.
- U.S. Department of the Interior, Fish & Wildlife Service and U.S. Department of Commerce, Bureau of Census. (November 1997). *1996 National Survey of*

*Fishing, Hunting, and Wildlife-Associated Recreation*. Washington D.C.:  
U.S. Government Printing Office.

- Vaughan, W. J., & Russell, C. S. (1981). An analysis of participation in nonconsumptive wildlife recreation: Comment. *Land Economics*, 57(2), 279-287.
- von Haefen, R. H. (2002). A complete characterization of the linear, log-linear, and semi-log incomplete demand system models. *Journal of Agricultural Resource Economics*, 27, 281-319.
- von Haefen, R. H. (2010). Incomplete demand systems, corner solutions, and welfare measurement. *Agricultural and Resource Economics Review*, 39(1), 22-36.
- Wales, T., & Woodland, A. (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics*, 21, 263-285.
- Wedel, M., & Kamakura, W. (2000). *Market Segmentation: Conceptual and Methodological Foundations* (2nd ed.). Boston, MA: Kluwer Academic Publishers.
- Wedel, M., ter Hofstede, F., & Steenkamp, J. E. (1998). Mixture Model Analysis of Complex Samples. *Journal of Classification*, 15, 225-244.
- Welsh, M., & Poe, G. (1998). Elicitation Effects in Contingent Valuation: Comparisons to a Multiple-bounded Discrete Choice Approach. *Journal of Environmental Economics and Management*, 36, 170-185.



- Whitehead, J. C. (2002). Incentive Incompatibility and Starting Point Bias in Iterative Valuation Questions. *Land Economics*, 78, 285-297.
- Whitehead, J. C. (1994). Item nonresponse in contingent valuation: Should CV researchers impute values for missing independent variables. *Journal of Leisure Research*, 26(3), 296-303.
- Whitehead, J. C., & Aiken, R. (2007). Temporal reliability of willingness to pay from the National Survey of Fishing, Hunting and Wildlife-Associated Recreation. *Applied Economics*, 39, 777-786.
- Whitehead, J. C., Dumas, C. F., Hertine, J., Hill, J., & Buerger, B. (2008). Valuing beach access and width with revealed and stated preference data. *Marine Resource Economics*, 23, 119-135.
- Whitehead, J. C., Groothuis, P., & Blomquist, G. (1993). Testing for non-response and sample selection bias in contingent valuation: Analysis of a combination phone/mail survey. *Economics Letters*, 41(2), 215-220.
- Whitehead, J. C., Pattanayak, S. K., Houtven, G. L., & Gelso, B. R. (2008). Combining revealed and stated preference data to estimate the nonmarket value of ecological services: An assessment of the state of the science. *Journal of Economic Surveys*, 22, 872-908.
- Whitehead, J. C., Phaneuf, D. J., Dumas, C. F., Herstine, J., Hill, J., & Buerger, B. (2010). Convergent validity of revealed and stated recreation behavior

- with quality change: A comparison of multiple and single site demands. *Environmental & Resource Economics*, 45, 91-112.
- Willig, R. D. (1976). Consumer's surplus without apology . *The American Economic Review*, 66(4), 589-597.
- Woodward, R. T., Gillig, D., Griffin, W. L., & Ozuna Jr., T. (2001). The welfare impacts of unanticipated trip limitations in travel cost models. *Land Economics*, 77(3), 327-338.
- Yen, S. T., & Adamowicz, W. L. (1993). Statistical properties of welfare measures from count-data models of recreation demand. *Review of Agricultural Economics*, 15(2), 203-215.
- Zarnoch, S. J., Cordell, H. K., Betz, C. J., & Bergstrom, J. C. (May 2010). *Multiple imputation: An application to income nonresponse in the national survey on recreation and the environment*. Southern Research Station. Forest Service, US Department of Agriculture.
- Zawacki, W. T., Marsinko, A., & Bowker, J. (2000). A travel cost analysis of nonconsumptive wildlife-associated recreation in the United States. *Forest Science*, 46(4), 496-506.
- Zhang, P. (2003). Multiple imputation: Theory and method. *International Statistical Review*, 71(3), 581-592.