A SITE-PORTFOLIO MODEL FOR MULTIPLE-DESTINATION RECREATION TRIPS: VALUING TRIPS TO NATIONAL PARKS IN THE SOUTHWESTERN USA

by

Zhe Chen

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

Summer 2016

A SITE-PORTFOLIO MODEL FOR MULTIPLE-DESTINATION

RECREATION TRIPS:

VALUING TRIPS TO NATIONAL PARKS IN THE SOUTHWESTERN USA

by

Zhe Chen

Approved: ____
James L. Butkiewicz, Ph.D.
Chair of the Department of Economics

Approved: ____
Bruce Weber, Ph.D.
Dean of the Lerner College of Business and Economics

Approved: ____
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

George R. Parsons, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Evangelos M. Falaris, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

David E. Black, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Christopher G. Leggett, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to express my sincere thanks to my advisor, George Parsons for his kind support, intelligent guidance, and mentorship during my graduate studies. This dissertation draws heavily from a forthcoming manuscript I jointly wrote with George Parsons, Chris Leggett, Joe Herriges, and Nancy Bockstael, a 2002 memorandum by Nancy Bockstael, Joe Herriges, and Cathy Kling, and a 2003 memorandum by Joe Herriges. I wish to express my gratitude to all of these authors for their help in completing this dissertation.

The data used in this dissertation come from a 2002 national park survey funded by the National Park Service. I would like to thank Chris Leggett for providing this well-documented survey data and other related documents.

Lastly, I wish to thank the other members of my dissertation committee, Evangelos Falaris, David Black and Chris Leggett for their help, encouragement and valuable comments.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The multiple-site visitation problem has afflicted travel cost models since their inception. Over the years, several approaches have been proposed to address the issue of allocating travel costs when multiple-site visitations are involved; however, these approaches have proven to be problematic for generalized application. I propose a new method for analyzing multiple-destination recreation trips and apply it to visitation to national parks in the southwestern United States, including well-known parks such as the Grand Canyon and Zion National Parks. I use conventional random utility theory and treat groups of parks (portfolios) as choice alternatives. I consider one choice occasion per respondent and condition that choice on the person visiting at least one park in the choice set, so the participation decision (go/no-go) is not modeled. Trip cost includes time, travel, lodging, and food cost for visiting all sites in the portfolio. Variation in trip cost is generated by where individuals enter and exit the region and by variation in the specific set of parks in each portfolio. Specialized sampling weights are used in the model to correct for on-site sampling. I estimated three empirical versions of this choice model: Standard Logit with Additive Site Utilities (SL), Mixed Logit with Additive Site Utilities (MXL), and Portfolio Specific Constants as Utilities (PSC). I found that the PSC model performs relatively better than the SL model in terms of accounting for the complementary effects among parks. MXL model with a constrained distribution of the random parameters provides more behaviorally reasonable estimates compared to other traditionally assumed distributions. Finally, I provide estimates of values for closing individual parks or groups of parks. The loss-

to-trip ratios (per trip value) for individual park closures range from \$143 to \$255 for Additive Site Utility Models. The aggregated welfare losses for individual park closures over the season (June 2002) range from \$2.4 million for Canyonlands to \$40.9 million for Grand Canyon.

Chapter 1

INTRODUCTION

National parks in the U.S. are well known for their breathtaking views. Each year, they attract millions of visitors from across the nation and around the world. According to the National Park Service (NPS), in 2014 the overall visitations to national parks, seashores, monuments, or historical sites hit a record-high of 292.8 million. National park visits alone reached 68.9 million. Among the 59 national parks in the U.S., the ones in the Southwestern region are famous for their unique landscapes and their cultural and historic significance. According to the NPS Annual Recreation Visitation Report, in 2014 over 30% of national park visits were to southwestern parks[1]. Each year, millions of people with a variety of tastes and preferences visit southwestern national parks due to these parks' diverse characteristics. Countless fascinating hiking trails, wild backcountry experiences, and breathtaking contrasting colors and landscapes provide visitors numerous choices amongst different sites.

One of the purposes of this study is to obtain estimates of the damages the public would incur in the event of a short-term closure of one or more national parks in the southwestern United States. This study is mainly focusing on seven relatively popular national parks in the "Four Corners" states (Utah, Colorado, Arizona, and New Mexico), specifically Arches National Park, Bryce Canyon National Park,

---

[1] See National Park Service Stats Reports: National Park Service Visitor Use Statistics https://irma.nps.gov/Stats/Reports/National

Canyonlands National Park, Grand Canyon National Park, Mesa Verde National Park, Petrified Forest National Park, and Zion National Park.

Trips to these national parks differ from the trips modeled in most recreational demand analyses in a number of ways. First, for most individuals, trips to these parks are usually taken no more than once a year and for many they may represent a once or twice in a lifetime event. Given that no more than one trip is taken by most people who visit these parks, little about individual preferences can be learned with trip frequency information. Relating the number of trips taken to the trip's cost, as has been done in most single-site demand analyses, is not feasible. Second, a visit to one of these national parks tends to be a non-single-site trip. This is due to the fact that a majority of visitors travel a considerable distance to reach these parks and most national parks in this four state region are located fairly close to one another. People visiting southwestern national parks usually take week- or even month-long trips and they often choose to visit multiple parks. In many cases, their trips also involve visits to other, non-park, destinations in the area. This makes the recreational "commodity" being consumed more complex than a usual recreational trip. Third, because visitors to these national parks come from all over the U.S. (and the world), the possibility of contacting a visitor who has been to any of these seven national parks in a general population survey is rather small. Therefore, collecting enough information about the dichotomous decision to take one or more of these trips by sampling the population randomly from certain off-site regions or even nationwide is very challenging.

The other purpose of this study is to develop new method that can overcome the issue of multiple-site visitations in the traditional travel cost model. To address the multiple-site visitation problem, I frame the site choice problem using a portfolio-based approach. Each visitor is considered to choose one or multiple parks in the four-state region to visit, conditioned on taking at least one trip to one of the seven national parks. In other words, I argue that it can be assumed that individuals face a choice among a set of park portfolios. Each portfolio contains a unique combination of the national parks in the region. For instance, if there were only three national parks (A, B, and C) in the region, then the visitor can choose to visit one park or two parks or three parks in a single trip, and the possible portfolios for the visitor to choose from are {A}, {B}, {C}, {AB}, {AC}, {BC}, and {ABC}. Given that I focus on seven southwestern national parks, individuals will be choosing between 127 different portfolios. Following traditional random utility maximization theory, individuals have utilities for all alternative recreation portfolios and are assumed to choose the portfolio that maximizes their utility. The utility from each portfolio depends on the sites included in the portfolio, trip costs, characteristics of the decision-makers, and random factors that are unobservable to the researchers. A person's trip costs for visiting a portfolio consist of two parts; out-of-pocket costs and the opportunity cost. Out-of-pocket costs include park entrance fees, driving, lodging, and dining costs, while the opportunity cost is mainly the cost of travel time. I estimated three empirical versions of this choice model: Standard Logit with Additive Site Utilities (SL), Mixed Logit with Additive Site Utilities (MXL), and Portfolio Specific Constants as Utilities (PSC). For the MXL model, I also tested different random parameter distributions and compared the results. I found that the PSC model performs relatively better than the SL model in terms of accounting for the substitution/complementary effects among parks. The flexible nature of the MXL model allows for correlation among error terms, thus is considered to be a better fit for the portfolio-based model where portfolios sharing the same park(s) are likely to have correlated error terms. Also, I found that MXL model with a constrained distribution of the random parameters provides more behaviorally reasonable estimates compare to other traditionally assumed distributions.

The data to study national park portfolio choices was collected in a two-step process. First, participants were randomly recruited on-site at each of the seven national parks during a two-week period in June 2002. In the second step, Southwest National Park Visitor Surveys were mailed to all recruits in July 2002 to follow up on their trip detail information. It is worth noting that this analysis is entirely conditional on the individual making a trip to at least one national park in this region. Therefore, the question that can be answered in this dissertation is: What are the losses to individuals who have planned a trip to at least one of the seven major national parks in the southwest, if they learn, after they have committed to the trip, that one of those parks is closed to the public during their trip? These losses can be considered short-term losses, since the study design excludes any cases when individuals find out about the closure in advance and cancel their entire trips. The per party per trip welfare losses for closing individual parks range from $12 for the least popular park – Canyonlands to $161 for the most popular park – Grand Canyon (2002$). The loss-to-trip ratios for individual park closures range from $143 to $255 for Additive Site Utility Models.

This dissertation is organized as follows. Chapter 2 provides a brief literature review on previous studies regarding travel cost model, recreational demand for multiple-destination/multiple-purpose trips, as well as a review of studies on portfolio-based discrete choice model in different contexts.

Chapter 3 presents the survey design and data collection process. Summary statistics for the data are presented in this chapter.

Chapter 4 describes the theoretical models. It first provides an overview of Random Utility Theory and then explains in depth how the site-portfolio models are formed. In addition, it lays out the different types of Random Utility Models (RUMs) used for recreational demand estimation.

Chapter 5 presents how the on-site sample is adjusted using exogenous population choice weights. Details on weights computation and examples are presented in this chapter.

Chapter 6 presents the estimation results from the random utility models presented in Chapter 4. Then, it offers estimated welfare losses to the public due to hypothetical single or multiple park closures in the U.S. southwest.

Finally, Chapter 7 provides conclusions and potential questions for future research.

Chapter 2

LITERATURE REIVEW

In this chapter, I will first briefly review traditional travel cost models. The main focus will then shift to the different approaches for modeling demand for multiple-destination recreation trips before finally reviewing portfolio-based discrete choice studies.

2.1 Traditional Travel Cost Models and Recreation Literatures

When it comes to measuring the economic value of recreational use of non-market resources, the Travel Cost Model (TCM) is the most commonly used method. The idea of the travel cost method was first proposed by Hotelling (1949) in an unpublished letter to the National Park Service regarding the recreational use of U.S. national parks. The basic idea behind this method is that although there is no price for any non-market resources, such as trips to national parks, the cost of reaching the site can serve as a good proxy "price" for this non-market good. Given this assumption, a traditional demand function based on this "price" and the number of trips taken and/or sites chosen to visit can be easily forged, allowing measurement of individuals' willingness to pay for the recreational use of non-market resources. Travel cost models are thus generally used when measuring the economic value of site access and quality changes in recreational sites.

The earlier and simpler version of the travel cost model is a single-site model, which is still widely used in the modern literature. The single-site model examines the demand for the recreational trips to a given site over a period of time. Just like the demand for market goods, it assumes that when trip costs increase the quantity demanded (the number of trips taken to the given site) decreases. Earlier applications of the TCM model were almost exclusively based on zonal data (Trice & Wood 1958; Clawson & Knetsch 1966). Areas around a single site were first defined as different geographic zones and then the travel costs from the center of each zone to the site were treated as the proxy "price" of the recreational use of the site. Over the years, the TCM has developed considerably. Starting in the 1970s researchers began to replace aggregated zonal data with individual-level data (Brown and Nawas 1973). This allows a more precisely measured insight into individual demand. In the late 1980s, single-site models took another leap when researchers began using truncated dependent variables, treating the trip counts as a continuous variable (Shaw 1988; Hellerstein & Mendelsohn 1993; Haab & McConnell 1996).

Another commonly used type of travel cost model is the RUM model. This approach became popular in the 1980s (Bockstael et al. 1984; Carson et al. 1986). Instead of estimating a demand function, the RUM model begins with a utility function. It focuses on an individual's choice of which site to visit among a number of possible sites. The site choice is based on the attributes of all sites and trip costs to get to each site, with each individual choosing the site that maximizes their utility. In this way, a full set of sites is incorporated, instead of focusing on only one site's characteristics. Phaneuf, Herriges and Kling (2000) present a generalized version of the RUM model. Their generalized corner solution model not only accounts for recreationist site choices but also the number of trips taken to each site. Over the years other authors presented refinements of the RUM model. For example, Train (1998) introduced simulated probability models and mixed logit/random parameter logit into this framework. As discussed in more detail in Chapter 4, this greatly reduces the restrictions from the independence of irrelevant alternatives (IIA) assumption and allows for preference heterogeneity.

2.2 Approaches for Modeling Multi-Destination/Multi-Purpose Trips

Over the years, numerous studies attempted to measure the economic value of recreational sites using the TCM. When using the TCM, the accuracy of the estimates relies on the validity of the assumption that individuals only make single destination trips and that the recreation site visited is the sole purpose for their trip. Therefore, single-site/single-purpose trips has been the main focus of this section of the literature.

However, in many cases multi-destination or multi-purpose trips, during which travelers visit more than one site or have purposes other than just visiting a recreational site, are quite common, especially when visitors travel considerable distances to reach the area. A survey conducted by the National Park Services in 1982 on visitation to Bryce Canyon National Park shows that 71% of Bryce Canyon visitors also visited Zion National Park and that 58% of them also visited the Grand Canyon (Haspel & Johnson 1982).

Potential violations of the single-site/single-purpose trip assumption in the TCM model have made it difficult to truly estimate the value of some recreation sites. Many studies have avoided this issue by either simply excluding multi-destination/multi-purposes trips from the sample or by treating them as single destination trips. Smith and Kopp (1980) discuss the spatial limits of the travel cost recreational demand model. They point out that as more origin zones are included in the sample, the assumption that each trip is single-purpose/single-destination, along with several other assumptions, becomes increasingly untenable. They suggest that a formal test for the stability of the estimated parameters should be performed in order to identify the spatial limits to the model, and that the sample should then be restricted to only the origin zones within these spatial limits to exclude all potential multi-destination/multi-purpose trips.

This sort of ad hoc solution only works if the proportion of multi-destination visitors is relatively small. Previous studies suggest that simply omitting multi-destination/multi-purpose or treating all recreational trips as single-site/single-purpose oriented could easily produce biased estimates of the consumer surplus derived from recreation sites. Haspel and Johnson (1982) showed that treating multi-destination trips as single-destination trips tends to overstate the value of the site. Loomis et al. (2000) also found that estimated consumer surplus per person per trip increases when multiple destination trips are included. Although their 95% confidence intervals suggest that the increase in estimated per trip consumer surplus is not significant, there is still a significant overestimation of total site values. Mendelsohn et al. (1992) note the importance of finding the correct way to measure the value of multi-destination trips when this type of trip is prevalent in the sample. The omission of close substitutes tends to underestimate the value for any site that is frequently part of a multi-destination trip.

In addition to these issues, Kuosmanen, Nillesen and Wesseler (2004) point out a subtler problem that arises from ignoring multi-destination trips. They believe that single destination vacationers may have different demographic profiles than multi-destination travelers. People who make single-destination trips are mostly people who live closer to the recreation spot, who may be systematically different from people who live further away. Therefore, omitting multiple-destination groups may leave some important demographic features under-represented.

The extant literature has suggested several ways to solve the multi-destination trips problem. One potential solution is to correct the estimation bias by assigning a fraction of the total travel cost to the evaluated site and then using weighted/adjusted travel costs for demand estimation. Haspel and Johnson (1982) divide the round-trip travel cost by the number of stops within the trip, assuming that all major destinations are equally spaced and valued, and then use the average willingness to pay (WTP) to travel to all destinations within the trip as the proxy travel cost to Bryce Canyon. They estimate the per-vehicle WTP for visiting Bryce Canyon is $91.

However, this method of disaggregating total joint costs is very arbitrary and cannot be consistently applied in most cases. For instance, compared to the rest of the sites individual might visit during a trip to Bryce Canyon, Zion National Park and Grand Canyon National Park are located considerably closer to Bryce Canyon. Therefore, most visitors choose to pay a visit to those two parks when visiting Bryce Canyon. Simply dividing the total trip cost by the number of stops would greatly overestimate the demand for Bryce Canyon because some of the demand for a stop at Bryce Canyon is derived from demand for the other two sites. To overcome this, Haspel and Johnson grouped the three national parks as one single destination, which brought the estimated consumer surplus of visiting Bryce Canyon down to $69 per trip.

Finding the correct proportion of the trip costs to allocate to the evaluated site is inevitably a challenge for this approach. It is necessary to define some systematic way to allocate trip costs across sites. Some authors recognized that the importance of each stop to an individual depends on far more than just the distance between stops. They suggest using some quantifiable variable, such as time spent at each site/objective, to value the importance of individual sites within a trip and allocate the trip costs accordingly (Knapman & Stanley 1991; Yeh, Haab & Sohngen 2006). Other studies use subjective values such as visitors' stated preferences for different sites as a measure of the importance of each site (Kuosmanen, Nillesen & Wesseler 2004; Martinez-Espineira & Amoako-Tuffour 2009). However, both of these methods have their limitations. Given that there's no uniform measure for consumers' subjective values, it is very difficult to accurately estimate site values based on their provided information. Conversely, quantifiable variables, such as nights spent on site, may not accurately reflect the importance of each site. Certain sites might be the main reason that individuals decided to make the trip at all while still not being the site at which they spent the largest amount of time.

Parsons and Wilson (1997) propose another approach to multi-purpose/multi-destination trips. They develop a single recreation demand model that incorporates multi-destination/ multi-purpose trips. In their model, incidental trips are treated as complements to primary purposes trips. Parsons and Wilson included a dummy variable in their regression to indicate trips with incidental consumption. The dummy indicator is able to capture the shift of the demand curve that occurs when there are multi-destination/multi-purpose trips involved. Their estimation results suggest that omitting the incidental consumption variable tends to slightly underestimate the value of the lost sites. Loomis et al. (2000) further expanded Parsons and Wilson's recreation demand model by separating joint consumption trips from incidental trips.

2.3 Portfolio-Based Discrete Choice

Instead of assigning the portion of travel costs to each stop of the trip and estimating the demand for each site separately, Mendelsohn et al. (1992) suggest an alternative way of analyzing multi-destination trips. They develop a demand system in which each combination of major sites visited is treated as a single commodity with its own demand function. They sampled at one site only (among four possible sites) and worked with zonal data. They estimate an inverse demand function for each site combination using the trip costs to all sites in the bundle and number of trips taken. By including the prices of different single sites and combinations of multiple sites in the same demand function, they are able to capture the substitution effects from consumers choosing between alternative site bundles. The authors emphasize that the loss of a single site will affect the prices of all the alternative bundles that included the closed site. Shutting down one site means removing all choice alternatives that include that site. Similar to the approach we adopt in this study, they measure the value of a given site as the value of the demand system with the site less what the value of the demand system would have been without the site, in effect measuring how much the site "adds" to the system. Their estimation results show that omitting alternative multiple trip choices can lead to a sizeable underestimation of site values. Consumer surplus from Bryce Canyon per trip per person increases from $9.47 dollars to $16.8 dollars when multiple sites bundles are taken into consideration, a roughly 77% increase.

The main limitation of Mendelsohn's method is the difficulty of deciding which sites should be included in the empirical work. Exactly which sites are feasible options for a trip is not always obvious, which means the set of alternatives can be hard to identify. More practically, some authors have pointed out that the size of the alternatives set increases exponentially as the number of individual sites increases, which can make estimation of the inverse demand system very challenging (Kuosmanen, Nillesen & Wesseler 2004). Unlike their analysis, mine is done in a discrete-choice setting using random utility theory, has sampling at all relevant sites and is based on individual level data instead of zonal data.

One of the studies that proposed the concept of bundling in a discrete choice model setting is the work by Tay, McCarthy and Fletcher (1996). They conceive household recreational decisions as a choice between numerous portfolios, each of which encompasses several individual choice dimensions. They point out that individual's travel decisions involve a complex process of simultaneous or sequential choices amongst different destinations, trip lengths, trip frequencies, and travel modes. Before Tay, McCarthy and Fletcher, Adler and Ben-Akiva(1979) considered a portfolio-based discrete choice model in a transportation context and Atherton, Ben-Akiva and McFadden(1990) and Train, McFadden and Ben-Akiva(1987) did so in a phone-services context.

Chapter 3

SAMPLE, SURVEY AND DATA

In this chapter, I discuss the design of this study, the data collection procedures, and provide descriptive statistics for the participation data.

3.1 Study Design

As noted in Chapter 1, trips to National Parks have three distinct characteristics that separate them from other recreation trips: (i) trips to national parks are usually taken no more than once in a year, and many are even once- or twice-in-a-life-time events, (ii) they are mostly multi-destination/multi-purpose trips, and (iii) visitors to these parks come from all over the U.S. (and the world), thus even sampling throughout the entire U.S would only have an extremely small probability of successfully contacting a potential visitor on a random phone call or mail survey. In order to collect a sufficient amount of data for the analysis, it is necessary to rely on on-site sampling methods. As discussed at length in the existing literature, on-site sampling can easily cause sampling bias if not treated correctly. Thus, the data gathering process faced not only the challenge of getting an acceptable distribution of respondents across the parks of interest but also the challenge of collecting data that would make it possible to control for selection bias arising from on-site sampling.

The preliminary on-site reconnaissance was first taken in nine national parks and monuments in the four states region. During this reconnaissance, the research team conducted informal interviews with visitors, rangers, and park managers in order to determine the general nature of trips to national parks in the southwest. Based on the information this reconnaissance gathered on sampling logistics and visitation patterns, the research team decided on sampling at seven national parks: Grand Canyon, Arches, Bryce Canyon, Zion, Mesa Verde, Petrified Forest, and Canyonlands. Figure 3.1 shows the geographic distribution of these parks.

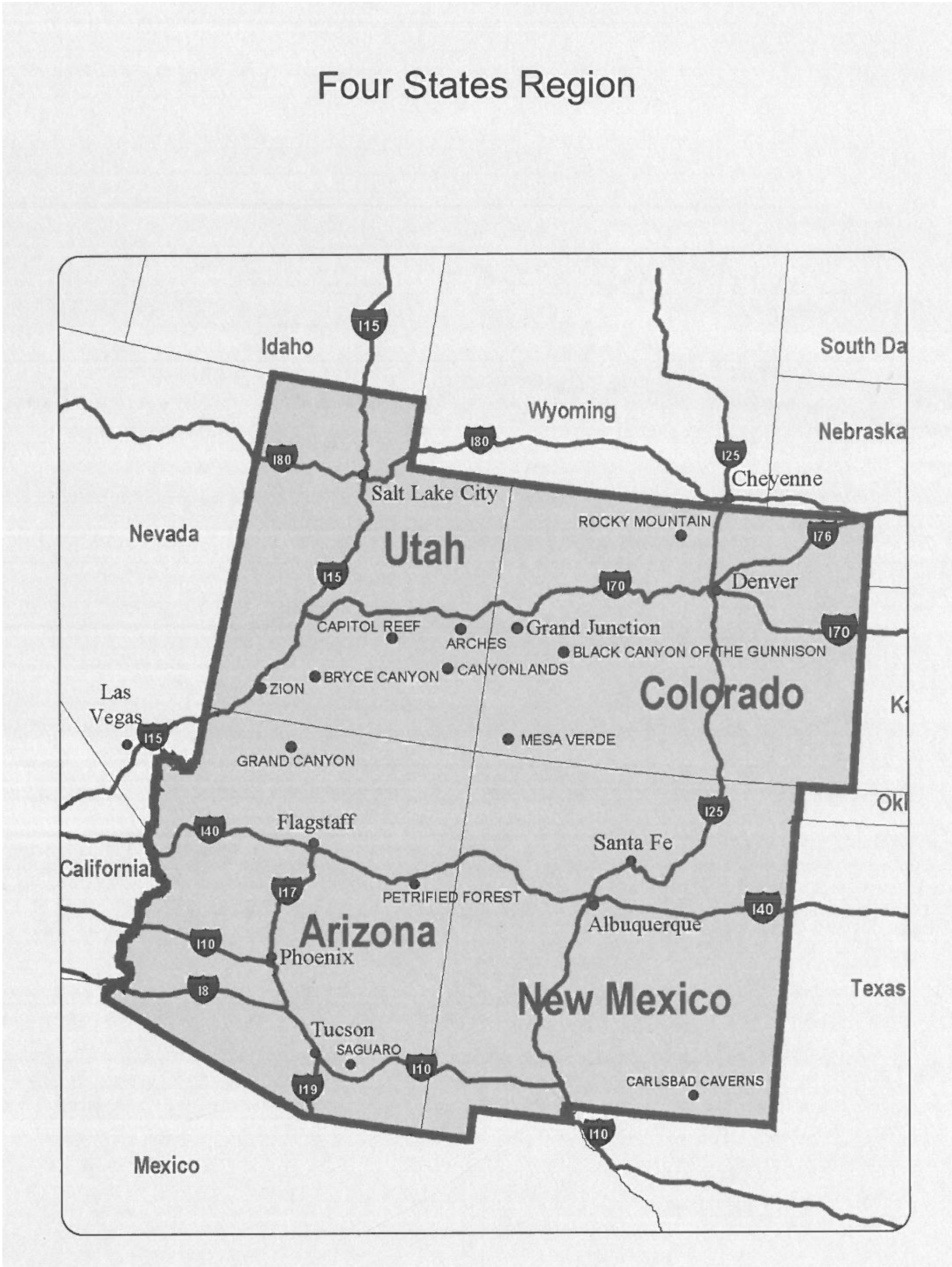


Figure 3.1: National Parks Sampled in the Four States Region

These are major national parks in the region; they are well known for their spectacular scenic vistas and offer a variety of additional attractions including archeological sites, geological features, hiking trails, and opportunities for wildlife viewing and nature appreciation. In addition, as shown in figure 3.1, these parks are spatially clustered together and therefore form natural bundles that are often chosen to visit by travelers during a single trip to the region. After the survey was designed, the research team first pre-tested it at Arches National Park, where visitors were randomly intercepted while entering the visitor center and were asked questions about the clarity of individual questions and the general flow of the survey. Eighty-six respondents completed this pre-test and the survey instruments were revised based on their comments.

The final survey was 13 pages and 54 questions long and contained 6 sections and one foldout map of the Four States Region. The survey is presented in full in Appendix A. Section A of the survey gathered general information on the respondents' trips. Respondents were asked to focus only on the trip they were taking when the research group interviewed them. This section gathered information about the respondents' arrival and departure dates in the Four States Region, as well as whether they currently lived in the Four States Region and the type(s) of vehicles used during the trip. If they were residents in the region, they were asked to circle the day they left home to begin the trip and the day they returned home at the end of the trip. Along with the travel dates, people were also asked to mark their entry and exit point in the region on the foldout map (O for entry, X for exit). If they made any brief side trips outside the region during the visit, they were instructed to only mark the final departure date and point from the region. Finally, respondents were asked to report the total number of days they spend in the region, excluding days spent on side trips outside the region. In this analysis, a "trip" is defined by the entry and exit into the four states region – time and money spent on the trip, parks visited on the trip, and other activities involved. Time and money spent to reach the region (e.g. airfare, bus/train expenses) were excluded.

The second section (B) of the survey focused on the details of visits to the seven National Parks in our analysis. Respondents were asked to report whether they paid a visit to the park, the number of separate times they entered the park, and the number of days spent onsite. The third section (C) of the survey collected information on respondents' visits to places other than the seven National Parks. This section provided lists of other national parks, national monuments, national historical parks, and national recreation areas in the Four States Region by state and asked respondents to identify which (if any) they had visited. Respondents were also asked if they had visited any of the 13 major cities in the area as part of their trip – Santa Fe, Las Vegas, Salt Lake City, Park City, etc. The fourth section (D) gathered information on any side trips (e.g. visiting friends or relatives or business/work stops) that respondents may have taken. This helped to provide a complete picture of any multi-destination/multi-purpose trips.

The fifth section (E) collected information on the characteristics of the respondents' parties during their trips. Respondents were first asked to describe the group they traveled with – whether they were traveling alone, with family, with friends, or with business associates. They were then asked about the number of people in the vehicle when they were interviewed and the age composition of their groups. This section also asked questions on lodging choices (e.g. hotels, camping, or stayed with friends/relatives) and the number of nights camping to help with calculations on lodging expenses. This section also included a small bank of questions on the importance of different activities during the trip to the region, such as biking, viewing scenery, driving scenic highways, nature study, exploring the visitor centers, and hiking. This section finishes with contingent valuation questions on respondents' maximum willingness to pay to visit the park. Respondents were first presented with the current entry fee for that park and were assured that the Park Service was not thinking of increasing the fee; they were asked to choose the highest amount they would have paid to visit the park during this trip, given a payment card (a list of numbers) starting from the current entrance fee. The following question then examined the factor that was the most important to the respondents when they chose this amount.

The last section (F) of the survey gathered information on basic demographic characteristics, such as age, gender, education, employment status, and household income, as well as the amount of flexibility that the respondents had in planning their trips.

3.2 Sampling and Survey Implementation

The survey was implemented in two stages. First, individuals were interviewed at entrances to all seven national parks in order to identify the eligibility of the respondents for the survey. Second, a survey was mailed to eligible individuals who had agreed to participate in the study.

Recruitment for the survey was done at the entrances to each park between June 15 and June 23, 2002. Each of the seven parks was sampled on two weekdays and two weekend days (except for the Grand Canyon which was sampled on three weekdays and three weekends) during this nine-day period. Appendix C Table C1 presents a more detailed description of this interview process. The first-stage survey presented during this recruitment stage consisted of a brief set of questions designed to evaluate the respondent's eligibility for a mail survey (see Appendix B for more detail). Respondents were randomly selected at each park entrance and all non-commercial vehicles entering the park were considered eligible for this initial survey. The goal was to target 200 people per day at each park entrance. The target sampling rate at each gate varied according to traffic flow and safety concerns (full target sampling rates are presented in Appendix C Table C2). At the gates of the relatively less popular National Parks (for example, Canyonlands and Petrified Forest) the target was to interview every vehicle (i.e., the sampling rate is 1-in-1). At the more popular National Parks, interviewing every vehicle was practically impossible and the target sampling rates are set to 1 in every 3 or 4 vehicles. The lowest target sampling rate (1 in every 9 vehicles) is at one of the entrances to Grand Canyon National Park, where the traffic flow is the highest among all of the parks in our analysis. For some parks, the target sampling rates also varied depending on the number of gates open to vehicles entering the park. The actual sampling rates varied from the target ones due to a variety of reasons, but can be calculated using information collected by the National Park Service throughout the course of the initial survey; specifically, the total number of vehicles passing through each park entrance on each day and daily summaries of cash register data from each park entrance (full details are presented in Appendix C Table C3 - Table C15). Obtaining actual sampling rates is crucial for on-site sampling correction (a topic covered in more detail in Chapter 5).

The eligible respondents recruited for the next stage of the survey were the adults (18 or older) in the vehicle with the most recent birthdays, who had to be U.S. citizens. The overall response rate (the number of recruits over the total number of eligible respondents) for the initial survey is 96%, with Grand Canyon and Canyonlands as the highest at 99% and Mesa Verde as the lowest at 92%. In total, 4,836 respondents were recruited to participate in the next stage of the survey.

The mail survey was conducted in July and August 2002. The first contact was done on July 17, 2002, when survey booklets and introduction letters were mailed to all 4,836 individuals who agreed to participate. On July 24, a reminder postcard was sent to all respondents and a second set of survey and introduction letters were mailed on August 14 to the 2,654 participants who had not yet responded. Among the 4,836 participants, 3,311 completed the mail survey, giving a response rate of 68%. The response rate by park ranged from a low of 63% for Petrified Forest to a high of 75% for Arches. The overall response rate for the survey (the total number of completed surveys over the total number of eligible individuals interviewed at the gates) was 65%. Among all parks, Arches had the highest overall response rate, 72%, while Petrified Forest had the lowest overall response rate, 60% (see Appendix C Table C16 for more detail). Of the 3,311 unique mail surveys, 592 were dropped due to respondents staying in second homes or with friends/relatives, having missing information on any critical variable (including entry/exit points, household income, national parks visited during the trip, or flexibility on planning the trip), or having non-adult respondents, unusual travel modes, an unusually long time spend on site (more than 60 days), or repeated respondents from the same vehicle, leaving 2,719 completed surveys used in our analysis.

3.3 Summary Statistics and Data Preparation

In this section, I present an overview of the data, including individual characteristics and trip statistics and then discuss the data preparation process. To correct for choice-based sampling bias, all data in our analysis are weighted so that observations may be interpreted as coming from a random draw of visits to the region (see Chapter 5 for a detailed discussion on this weighting process).

3.3.1 Summary Statistics - Participation Data

Table 3.1 presents a set of frequency distributions for the demographic data – age, education, employment, income, and gender. It shows that most (30%) National Park visitors are in their 40s, with an average age of 48. More than half of the sampled population has an education level of college graduate and above (36% college graduate and 26% graduate school). Most of the respondents are employed full time (62%) although there is also a large share of retirees (18%). 55% of the respondents are male and the most reported household income is in the range of $50,000-75,000.

Table 3.2, table 3.3 and table 3.4 present park visitation statistics. Table 3.2 presents a frequency distribution for the number of sites (among our set of seven) visited by respondents. As the table shows, around 38% of the respondents choose to visit more than one national park during a single trip, and 4 respondents visited all seven national parks on their trip. Table 3.3 shows the rank of visitation by Park. Grand Canyon is the most popular park among the seven with visitations by 63% of the sample, followed by Zion with visitations by 31% of the sample. Conversely, Canyonlands was the least popular site, with visitations by only 8% of the respondents. Table 3.4 ranks the popularity of the chosen portfolios, restricting the table to the top 25 most popular portfolios. A trip to Grand Canyon by itself is the most chosen portfolio (39%), followed by Zion (9%). The most popular multiple-park portfolio is a combination of Bryce Canyon and Zion (6%). As the table shows, the top chosen portfolios are mostly groups of parks that are located substantially close to one another, for example, Portfolio – Bryce Canyon and Zion, Portfolio – Grand Canyon and Petrified Forest, and Portfolio - Bryce Canyon, Grand Canyon and Zion.

In our analysis, time and money spend on the trip, starting at the entry point of the region and ending at the exit point of the region, constituted a respondent's travel cost. All of the other costs of reaching the region were excluded. Therefore, entry and exit points to the region are crucial for travel cost calculation. Table 3.5 presents a simple summary statistics of entry and exit points reported by state. 72% of the sampled population is from outside the region. Most respondents entered and exited the region in Arizona (respectively 32% and 30%).

Table 3.6 presents summary statistics of other relevant trip data. The average trip length for visiting the national parks in the region was 6.6 days. The average number of national parks (among the seven in our study) visited was 1.7 parks and the average number of other national parks visited is only 0.2, showing that the set of seven national parks in our study is a good reflection of visitors' choices. Table 3.6 also shows that respondents tended to stop by at other national attractions or cities in the region, with the average number of sites visited as, respectively, 0.9 and 1.5. The trip statistics also suggest that most people were traveling in groups. The average party size is 3.2, with an average of 2.4 adults and 0.8 children in the group. 82.2% of the visitors were traveling with family and 16.9% were with friends. Most respondents chose to stay in hotels overnight (71.7%). 32.9% of respondent visited family or friends during the trip and 9.6% of respondents claimed that they also made stops for work or business reasons in the region, which suggests that some trips in our sample are also multi-purpose trips. Finally, these data also show that 15.5% of respondents potentially rented cars for their trip. This is based on the assumption that if the reported entry and exit points are cities (suggesting that individuals arrived in the region either by train, bus, or airplane), then they would have needed to rent cars to travel between the national parks during their visits.

3.3.2 Data Preparation – Travel Cost

Travel cost is a critical variable for any travel cost model, as it explicitly converts the subjective values people have for trips to national parks into monetary terms. The construction of travel costs, in most cases, involves a number of judgment calls. In this section, I explain how travel cost is measured in this study in detail. First, recall that the unit used in this study is a party/household (a group of individuals in an interviewed vehicle) with an average of 2.4 adults and 0.8 children. Each party is making a single trip to visit at least one of the seven national parks in the four states region. Due to the setting of the portfolio model (explained in more detail in the following chapter), travel costs in this study focus only on the expenses incurred within the region. In other words, the travel cost for each party is computed from the time they enter until they depart the region. Any travel expenses incurred outside of the region (i.e., airfare to reach and depart from the region) is considered to be constant across all portfolios for a given party. From tables 3.5 and 3.6, we can see that 28% of the sample lives in the region, about 15.5% took mass transportation (bus, train, or airplane) to enter the region, and the remaining 56.5% drove to the area. The variation in entry and exit points generates variation in travel cost across the portfolios, as does the number of sites in a portfolio. There are 41 unique entry/exit points marked by respondents on the survey. See Appendix C Table C17 for more detail.

The travel cost for household *i* visiting portfolio *k* (where *m* = 1, … , 7 denotes individual parks) is measured as follows:

$$
\begin{aligned}
Travel\ Cost_{ik} = {} & \alpha_{v_i} \cdot distance_{ik} && (Transit\ Cost\ ) \\
& + \textstyle\sum_{m=1}^{7} \delta_m^E \cdot d_{mk} && (Entrance\ Fees) \\
& + \{income_i/250/3\} \cdot time_{ik} && (Time\ Cost) \\
& + \theta_k^F \cdot [adults_i + kids_i/2] \cdot lodging\ mode_i && (Food\ Cost) \\
& + \theta_k^L \cdot rooms_i \cdot lodging\ mode_i && (Lodging\ Cost)
\end{aligned}
\tag{1}
$$

$\alpha_{v_i} = per\ mile\ vehicle\ cost\ for\ vehicle\ type\ v\ used\ by\ household\ i$

$\delta_m^E = per\ vehicle\ entrance\ fee\ for\ park\ m$

$\theta_k^F = per\ person\ food\ cost\ for\ portfolio\ k$ (see equation 3 below)

$\theta_k^L = per\ room\ lodging\ cost\ for\ portfolio\ k$ (see equation 4 below)

$d_{mk} = 1\ if\ site\ m\ is\ in\ portfolio\ k, and\ 0\ if\ not$

$distance_{ik} = travel\ distance\ between\ parks\ in\ portfolio\ k\ for\ household\ i$

$income_i = annual\ household\ income\ in\ 2002\$$

$time_{ik} = average\ number\ of\ days\ respondent\ i\ spent\ visiting\ parks\ in\ portfolio\ k$ $(see\ equation\ 2\ below)$

$adults_i = number\ of\ adults\ traveling\ in\ household\ i$

$kids_i = number\ of\ children\ (< 18\ years\ old)\ traveling\ in\ household\ i$

$rooms_i = number\ of\ rooms\ rented\ by\ household\ i.$

$$lodging\ mode_i = 1\ if\ respondent\ i\ chose\ hotel\ , and\ 0.5\ if\ respondent\ i\ chose\ camping$$

For each respondent who lived in the four states region, the driving distances and times are calculated for the 127 feasible portfolios using the routing software Milemaker and the respondent's zip code. For respondents who are non-residents in the region, the driving distances and times for each portfolio are calculated using Milemaker, conditioned on their reported entry point into the region and exit point from the region. The times and distances were calculated for the fastest driving route that would allow the respondents to minimize their transit costs to visit all parks in their portfolio. I consider the specific order in which the parks are visited to be irrelevant. The per-mile vehicle costs are computed based on the type of vehicles they used for their trip. For respondents who chose more than one type of vehicles during the trip (< 2%), the per-mile vehicle cost is calculated using the sum across all of the vehicle types selected, implicitly assuming that these vehicles are used as a group, rather than switching between them. It appears that most of the second/third vehicles accompanied are RVs (69%). The rates used for the per-mile vehicle cost computation are from 2002 American Automobile Association (AAA) data, which include fuel, maintenance and tire wear (see table 3.7 for details). The AAA driving costs data do not include data for vehicle types such as motorcycles and RVs. For those, it is assumed that the driving cost of a motorcycle is 8/11 of the cost of a small car and that an RV's driving cost is 3 times the driving cost of a small car. Table 3.7 shows that most parties traveled in trucks or SUVs (33%).

The second part of the travel cost is from park entrance fees. Each party is assumed to pay the relevant entrance fee (δ_m^E) for one vehicle for each park in the portfolio (see table 3.8 for the entrance fee for each park). If the sum of the total costs across all parks visited exceeds $50 for a portfolio, we assume that the party instead purchased a single $50 group park pass. The entrance fee varies among different portfolios but stays constant for all individuals with the same portfolio.

The total cost also included the opportunity cost of time spent on the trip. To convert the time cost into monetary terms, we assume the opportunity cost of a day to be one-third of the household's annual income ($income_i$) divided by the assumed number of working days per year (50 weeks × 5 days per week=250 days). The length of a trip in days to visit portfolio k is

$$time_{ik} = \frac{\{\sum_{m=1}^{7} days_m \cdot d_{mk}\} \cdot 8 + traveltime_{ik}}{10} \qquad (2)$$

$days_m = average\ number\ of\ days\ respondents\ stayed\ at\ park\ m\ while\ visiting\ the\ area$

$d_{mk} = 1\ if\ park\ m\ is\ in\ portfolio\ k, and\ 0\ if\ not$

$traveltime_{ik} = travel\ time\ in\ hours\ between\ parks\ in\ portfolio\ k\ for\ household\ i$

Respondents reported their number of days at each park in ½ day increments (½ day, 1 day, 1 ½ days, etc.). The average number of days at each park ($days_m$) is the average for all trips to that park by all respondents. We assume 8 hours for each day of onsite time, thus $\{\sum_{m=1}^{7} days_m \cdot d_{mk}\} \cdot 8$ gives the total amount of hours spent onsite for all parks in each portfolio. $Traveltime_{ik}$ is also computed using Milemaker and is measured in hours. The total amount of time spent during the trip (onsite time + travel time between parks) is divided by 10 to convert hours to days, assuming that a full day of traveling and onsite time cannot be a full 24-hour day (in this case I assume it contains 10 hours). In other words, I assume one overnight stay for every 10 hours of onsite plus travel time.

The last two components of the travel cost are meal expenses and lodging expenses. These are computed using the federal government's per diem rates for nearby cities (see table 3.8 for details). The per diem rate is a good proxy for costs in these areas and accurately picks up variation in the costs across all parks.

The per person food cost for portfolio k, shown as θ_k^F in equation (1), is

$$\theta_k^F = \frac{\sum_{m=1}^{7} \delta_m^F \cdot days_m \cdot d_{km}}{SumDays_k} \cdot time_{ik}$$

OR

$$\theta_k^F = \sum_{m=1}^{7} \delta_m^F \cdot \frac{days_m \cdot d_{mk}}{SumDays_k} \cdot time_{ik} \quad (3)$$

$\delta_m^F = federal\ government\ per\ diem\ rate\ for\ food\ for\ town\ closest\ to\ park\ m$

$days_m = average\ number\ of\ days\ respondents\ stayed\ at\ park\ m\ while\ visiting\ the\ area$

$d_{mk} = 1\ if\ park\ m\ is\ in\ portfolio\ k, and\ 0\ if\ not$

$Sumdays_k = \sum_{m=1}^{7} days_m \cdot d_{mk}$

$time_{ik} = The\ length\ of\ a\ trip\ in\ days\ to\ visit\ portfolio\ k$

The per diem per person food cost by portfolio is first weighted by average onsite time at each park. Then, I use total trip length for each portfolio (onsite time + travel time) times the per diem per person food cost to compute the per person food cost for each portfolio. As shown in equation (1), adults are assumed to pay the full meal per diem and children pay ½ the meal per diem over days spent visiting portfolio *k*. Also, respondents staying in hotels/motels are assumed to pay full per diem while campers pay ½ per diem.

The per room lodging cost for portfolio k, noted as θ_k^L in equation (1) can be expressed as

$$\theta_k^L = \frac{\sum_{m=1}^{7} \delta_m^L \cdot days_m \cdot d_{km}}{SumDays_k} \cdot nights_{ik}$$

OR

$$\theta_k^L = \sum_{m=1}^{7} \delta_m^L \cdot \frac{days_m \cdot d_{mk}}{SumDays_k} \cdot nights_{ik} \qquad (3)$$

$\delta_m^L = federal\ government\ per\ diem\ rate\ for\ lodging\ for\ town\ closest\ to\ park\ m$

$days_m = average\ number\ of\ days\ respondents\ stayed\ at\ park\ m\ while\ visiting\ the\ area$

$d_{mk} = 1\ if\ park\ m\ is\ in\ portfolio\ k, and\ 0\ if\ not$

$Sumdays_k = \sum_{m=1}^{7} days_m \cdot d_{mk}$

$nights_{ik} = int(time_{ik}) = number\ of\ nights\ spend\ during\ a\ trip\ to\ visit\ portfolio\ k$

Lodging cost per room per diem is also weighted using the average onsite time at each park. The per room lodging cost for each portfolio can be computed by multiplying this cost by the average total number of nights households are in the area when visiting portfolio k. The number of nights is the integer of number of days spent during a trip to visit portfolio k. Assuming that two adults share one room and the number of children does not to affect the number of rooms, $rooms_i$ can then be computed as $ceil(adults_i/2)$ (i.e. < 2 adults implies 1 room, 2 to 4 adults implies 2 rooms, 4 to 6 adults implies 3 rooms, and so forth). Similar to the method used for meal expenses, I assume that respondents staying in hotels/motels pay full per diem and campers pay ½ per diem.

Table 3.9 shows the decomposition of the resulting total travel cost. As the table shows, the mean travel cost to all portfolios across all parties in the sample is \$1,698. The possible total costs among all portfolios range from \$40 to \$9,978. The mean travel cost across all chosen portfolios is \$897. Among all costs, the opportunity time cost account for the largest share of the average travel cost at 32%, followed by food cost at around 30%.

Table 3.1: Demographic Data (n = 2719)[1]

	Mean or Percent of sample	Number of Respondents	Percentage of Respondents
Age	48 years		
<=20 years		24	1%
21~30 years		277	11%
31~40 years		441	17%
41~50 years		776	30%
51~60 years		567	22%
61~70 years		328	13%
71~80 years		132	5%
81~90 years		18	1%
Education Level			
Less than high school		3	<1%
Some high school		22	1%
High school or GED		251	9%
Technical or trade school degree		118	4%
Some college		637	23%
College graduate		975	36%
Graduate school		711	26%
Employment Status			
Full time		1695	62%
Part time		249	9%
Work in household		145	5%
Unemployed		58	2%
Retired		501	18%
Student		71	3%
Other		1	<1%
Household Income[2]	$ 72481		
Less than $15,000 per year		91	3%
$15,000 to $20,000 per year		54	2%
$20,000 to $30,000 per year		198	7%
$30,000 to $40,000 per year		281	10%
$40,000 to $50,000 per year		291	11%
$50,000 to $75,000 per year		676	25%

Table 3.1 continued

	Mean or Percent of sample	Number of Respondents	Percentage of Respondents
$75,000 to $100,000 per year		541	20%
$100,000 to $150,000 per year		403	15%
More than $150,000 per year		184	7%
Male	55%		

[1] Households with missing income are excluded from the sample because income is needed to estimate the value of time in our models – 215 observations were dropped for this reason. Also, the number of observations is less than 2716 for some of the other variables due to item non-response.
[2] The mean is calculated using the midpoints of the income categories (using $150,000 for the highest group).
[3] Due to rounding some percentages may not add up to 100%.

Table 3.2: Number of Parks (Among Set of Seven) Visited

Number of Parks Visited by Respondent	Number of Respondents	Percent of the Sample
1	1695	62%
2	547	20%
3	260	10%
4	128	5%
5	61	2%
6	25	1%
7	4	<1%
Total	2719	100%

Table 3.3: Visitation by Park

Parks	Visitors	% of the Sampled Visitors
Grand Canyon	1715	63%
Zion	851	31%
Bryce Canyon	590	22%
Arches	430	16%
Mesa Verde	419	15%
Petrified Forest	338	12%
Canyonlands	216	8%

Table 3.4: Most Frequently Chosen Portfolios[1]

Portfolio Group	Visitors	Percent of the Sample
Grand Canyon	1070	39%
Zion	243	9%
Mesa Verde	191	7%
Bryce Canyon, Zion	153	6%
Grand Canyon, Petrified Forest	131	5%
Bryce Canyon, Grand Canyon, Zion	118	4%
Arches	103	4%
Grand Canyon, Zion	66	2%
Arches, Canyonlands	51	2%
Bryce Canyon	46	2%
Petrified Forest	33	1%
Grand Canyon, Mesa Verde	32	1%
Bryce Canyon, Grand Canyon	25	1%
Grand Canyon, Mesa Verde, Petrified Forest	24	1%
Bryce Canyon, Grand Canyon, Petrified Forest, Zion	23	1%
Arches, Grand Canyon	20	1%
Arches, Bryce Canyon, Canyonlands, Grand Canyon, Zion	19	1%
Arches, Bryce Canyon, Zion	18	1%
Arches, Bryce Canyon, Grand Canyon, Zion	18	1%
Arches, Canyonlands, Mesa Verde	18	1%
Arches, Mesa Verde	17	1%
Arches, Zion	16	1%
Bryce Canyon, Grand Canyon, Mesa Verde, Zion	12	<1%
Arches, Bryce Canyon, Grand Canyon, Mesa Verde, Zion	11	<1%
Arches, Bryce Canyon, Canyonlands, Zion	10	<1%
All others	253	9%
Total	2719	100%

[1]Only 111 out of 127 possible portfolios were chosen by respondents.

Table 3.5: Entry and Exit Points in the Four State Region

Entry/Exit States	Entry Points		Exit Points	
	Number of Respondents	% of the Respondents	Number of Respondents	% of the Respondents
Arizona	861	32%	817	30%
Colorado	69	3%	68	3%
New Mexico	362	13%	352	13%
Utah	654	24%	708	26%
Residents in the Four States Region	774	28%	774	28%
Total	2719	100% *	2719	100%*

* Due to rounding percentages may not add up to 100%.

Table 3.6: Trip Statistics

	Min	Mean	Max	SD
Number of Days in Area	1	6.6	148	5.7
Number of National Parks (Among the 7 parks) Visited	1	1.7	7	1.1
Number of Other National Parks Visited	0	0.2	4	0.5
Number of Other National Attractions Visited	0	0.9	15	1.5
Number of Other Cities Visited	0	1.5	10	1.4
Party – Size	1	3.2	16	1.4
Number of Children in Party	0	0.8	14	1.2
% Renting Cars[1]	-	15.5%	-	0.4
% Staying in Hotels	-	71.7%	-	0.4
% Visiting family/friends during the trip	-	32.9%	-	0.5
% Business trips	-	9.6%	-	0.3
% Traveling alone	-	5.3%	-	0.2
% Traveling with family	-	82.2%	-	0.4
% Traveling with friends	-	16.9%	-	0.4
% Traveling with business associates	-	1.2%	-	0.1

[1] There is no direct information on whether the respondents rented cars. It is assumed that if respondents took mass transportation (bus, train, airplane, etc.) to enter the region, then most likely they rented cars to visit parks in the region.

Table 3.7: Vehicle Cost Per-Mile[1]

Type of Vehicle	% of Sample[2]	2002 Cost-per-mile (cents)
Small Car	10%	10.6
Mid-sized Car	20%	11.8
Full-sized Car	14%	13.0
Van	19%	11.0
Truck/SUV	33%	11.6
Motorcycle	< 1%	7.7 (8/11 × Small Car)
RV	6%	31.8 (3 × Small Car)

[1]American Automobile Association. (2002). *Your Driving Cost 2002* [Pamphlet]. Costs include fuel, maintenance and tires.
[2]Some respondents chose more than one type of vehicles; therefore the percentages in the sample do not necessarily add up to 100%.

Table 3.8: Lodging, Food Cost, Entrance Fees and Average Time on Site by Park

Parks	2002 Lodging Per-Day[1]	2002 Food Per-Day[1]	2002 Entrance Fees Per Vehicle[2]	Average Time on Site
Arches	$87	$38	$10	0.9 days
Bryce Canyon	57	38	20	1.2
Canyonlands	87	38	10	0.9
Grand Canyon	103	46	20	1.6
Mesa Verde	67	34	10	1.0
Petrified Forest	65	38	10	0.6
Zion	57	38	20	1.3

[1] Federal government per diem rates for the towns closest to each park. U.S. General Services Administration – Per Diem Rates Look-Up in 2002 (http://www.gsa.gov/portal/category/100120).
[2]If total entrance fee for a portfolio is greater than $50, then we assume they purchased a $50 park pass.

Table 3.9: Per Party Travel Cost

	Total Cost	Transit Cost	Lodging Cost	Food Cost	Entrance Fee	Time Cost[1]
All Portfolios						
Mean	$1698	$178	$421	$510	$42	$547
Min	40	2	0	13	10	8
Max	9978	1616	4178	4473	50	2586
SD	843	117	286	322	11	366
% of Total[2]	100%	10%	25%	30%	2%	32%
Chosen Portfolios						
Mean	$897	$92	$212	$281	$26	$287
Min	82	2	0	20	10	9
Max	8999	1616	2785	4327	50	2555
SD	653	104	192	226	13	254
% of Total[3]	100%	10%	24%	31%	3%	32%

[1] Time costs are opportunity costs calculated using travel time (transit time between parks + on site time) times 1/3 of household income.

Chapter 4

THEORY AND METHEDOLOGY

In this chapter, I present the theoretical foundations of and methodology behind this study on national park visitors' choice behavior. General park visitation choices can be represented as a standard discrete choice, where individuals are considered to be facing a choice among a set of alternatives, including the alternative of choosing no item from the set. From the decision maker's perspective, all of these alternatives are mutually exclusive. In other words, when the decision maker chooses one alternative none of the other alternatives can be chosen. In addition, the set of alternatives is both finite and exhaustive (i.e., all possible alternatives are included in a finite set). Each alternative in the set can be described by a set of attributes. When making the choice, the individual evaluates all the attributes, making their decisions based on the implied trade-offs between each alternative. Given the nature of this type of situation, discrete choice modeling is the optimal methodology for this analysis.

In the following sections, I first explain Random Utility Theory – the underlying assumption behind discrete choice modeling. I then present the method used in this analysis to address the multi-destination issue in this discrete choice framework, followed by the empirical specifications for these models.

4.1 Random Utility Theory

Random Utility Maximization Theory (RUM) is the most fundamental assumption underlying the discrete choice models. It argues that people are always governed by utility-maximization behavior when making consumption decisions. According to RUM theory, decision makers receive different levels of utility from different alternatives in the choice set, and it is always in their best interests to choose the one yield the highest possible utility. The utilities they derive from each alternative may depend on different attributes of the alternatives as well as the unique characteristics of the decision makers.

Random Utility Models, in a functional form, can be specified as follows. A decision maker, denoted by i, faces a set of J alternatives, with the quantity of these alternatives denoted by vector A_i. Let p_i denote the vector of prices for each alternative for individual i. Let B_i denote a vector of the quantities of other commodities consumed (with the price vector for other commodities normalized to 1) and y_i denote the total income of individual i. Given the information set, decision maker i faces the following problem to maximize her utility (U_i) from consuming A_i and B_i:

$$Max\ U_i(A_i, B_i)$$
$$s.t. \quad p_i \cdot A_i + B_i \leq y_i \tag{4}$$

The conditional indirect utility function is the solution to the above constrained utility maximization problem and is given by:

$$V_i = max\{U(A_i, B_i) \mid p_i \cdot A_i + B_i \leq y_i, A_i \geq 0, B_i \geq 0\}. \tag{5}$$

As established in the beginning of the chapter, for discrete choice problems all alternatives in the choice set are mutually exclusive, so one and only one alternative can be chosen per choice occasion. A_i, therefore, is a vector of $(J-1)$ 0s for all of the non-chosen alternatives and a 1 for the chosen alternative j $(j = 1, 2, \dots, J)$. Solving the utility maximization problem, conditioning on alternative j being chosen and applying Roy's identity, the conditional indirect utility function can be written as:

$$V_{ij} = \mathrm{V}\left(A_{ij}, y_i - p_{ij}\right) \forall j, \quad (6)$$

where A_{ij} represents the chosen alternative j $(j = 1, 2, \dots, J)$ and $(y_i - p_{ij})$ represents the residual disposable income available after spending p_{ij} on alternative j. As established earlier, each alternative in the choice set can be described as a set of attributes and characteristics of that alternative and consumers derive their utility from the attributes of the alternative instead of the alternative itself (Lancaster 1966). Therefore, equation (6) can be specified as:

$$V_{ij} = \mathrm{V}\left(q_{ij}, y_i - p_{ij}\right) \forall j, \quad (7)$$

where $q_{ij} \; \forall j$ is the vector of characteristics associated with alternative j. The conditional indirect utility function is thus a function of the attributes of alternative j exclusively (note that no other alternative's attributes are present in V_{ij}). This indicates that once alternative j is chosen, other alternatives' attributes would have no effect on an individual's utility (Bockstael & McConnell, 2007).

Consider now the problem faced by the researcher. The decision maker's utility function is known only by the decision maker; it cannot be directly observed by the researcher. The researcher observes some attributes, $x_{ij} \epsilon \, q_{ij} \; \forall j$, that affect the individual's decision making (but not all attributes) and some characteristics of the decision maker, z_i (other than income y_i). This makes it possible for the researcher to specify a function that relates these observed attributes/characteristics to the decision maker's utility:

$$V_{ij} = \mathrm{V}\left(x_{ij}, y_i - p_{ij}, z_i\right) \ \ \forall j . \tag{8}$$

This utility function is often called the representative utility or the deterministic component of the utility function. Since there are aspects that the researches cannot observe easily, we denote all the unknown factors as ε_{ij}. Therefore, the true utility function can be expressed as

$$V_{ij}^* = V_{ij} + \varepsilon_{ij} \quad \forall j . \tag{9}$$

where $V_{ij} = \mathrm{V}\left(x_{ij}, y_i - p_{ij}, z_i\right)$.

From equation (9), we can see that the decision maker would only choose alternative j if and only if $V_{ij}^* > V_{ik}^* \ \forall \ k \neq j$. Since ε_{ij} is the stochastic component of the utility function, the researcher's prediction on whether the decision maker chooses alternative j is a probabilistic matter rather than a deterministic one. The distribution of the random component ε_{ij} will therefore have a great effect on the prediction's accuracy. Denote the joint density of the random component as $f\ (\varepsilon_{ij}) \ \forall j$. The probability of individual i chooses alternative j is then given by:

$$P_{ij} = \ Prob\left(V_{ij}^* > V_{ik}^* \ \forall \ k \neq j\right)$$

$$= Prob\left(V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik} \forall\, k \neq j\right)$$
$$= Prob\left(\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik} \;\forall\, k \neq j\right)$$
$$= Prob\left(\varepsilon_{ik} - \varepsilon_{ij} < \mathrm{V}\left(x_{ij}, y_i - p_{ij}, z_i\right) - \mathrm{V}(x_{ik}, y_i - p_{ik}, z_i) \;\forall\, k \neq j\right) \qquad (10)$$

Assume that the representative utility function takes a linear form (i.e., the attributes and individual characteristics are linearly additive). The representative utility function then becomes:

$$V_{ij} = \beta \cdot x_{ij} + \alpha \cdot \left(y_i - p_{ij}\right) + \delta \cdot z_i \;\forall\, i, j, \qquad (11)$$

where β and δ are vectors of parameters for the alternative's attributes and individual characteristics respectively, and α is the marginal utility of income. By substituting equation (11) into equation (10), the probability of choosing alternative j becomes:

$$\begin{aligned} P_{ij} = Prob\,\Big(\varepsilon_{ik} - \varepsilon_{ij} & \\ & < \left(\beta \cdot x_{ij} + \alpha \cdot \left(y_i - p_{ij}\right) + \delta \cdot z_i\right) \\ & - \left(\beta \cdot x_{ik} + \alpha \cdot \left(y_i - p_{ij}\right) + \delta \cdot z_i\right) \forall\, k \neq j\Big) \\ = Prob\,(\hat{\varepsilon_\iota} < \beta \cdot \hat{x_\iota} + \alpha \cdot \hat{p}_i\,), & \end{aligned} \qquad (12)$$

where $\hat{\varepsilon_\iota}$ is the difference in the unobserved utility, $\hat{x_\iota}$ is the differences between the alternatives' attributes, and $\hat{p}_i$ is the difference between alternatives' prices.

As equation (12) shows, the probability of choosing alternative j is the probability that $\hat{\varepsilon_\iota}$ is less than $(\beta \cdot \hat{x_\iota} + \alpha \cdot \hat{p})_i$. Using the density function $f\ (\varepsilon_{ij})$, this probability can be specified as a cumulative distribution:

$$P_{ij} = Prob(\widehat{\varepsilon_i} < \beta \cdot \widehat{x_i} + \alpha \cdot \hat{p}_i)$$
$$= \int I(\widehat{\varepsilon_i} < \beta \cdot \widehat{x_i} + \alpha \cdot \hat{p}_i) \cdot f(\varepsilon_i) \cdot d_{\varepsilon_i} \qquad (13)$$

where $I(\cdot)$ is an indicator function, which equals 1 if $\widehat{\varepsilon_i} < \beta \cdot \widehat{x_i} + \alpha \cdot \hat{p}_i$, or, in a more general form, $\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik}\ \forall\ k \neq j$, and 0 otherwise.

Equation (13) shows that the choice probabilities depend on two factors – the differences among the alternatives' attributes ($\widehat{x_i}$ *and* $\hat{p}_i$) and the distribution of the random component, $f(\varepsilon_i)$. Note that it is the *differences* among alternative characteristics, rather than their absolute values, which affect the probability of selecting any alternative. In other words, attributes that do not vary across alternatives have no effect on the probabilities of selecting between those alternatives (Haab & McConnell, 2002). The choice probabilities also depend on the specification of the density function $f(\varepsilon_i)$. That is, different discrete choice models can be developed using different assumptions about the distribution of the error term ε_i (Train, 2009). The specifications of different discrete choice models will be discussed in the later sections of this chapter.

4.2 Site-Portfolio Approach

In this section, I present the theoretical model that not only describes a party's choice on national park visitation, but also addresses the multi-destination issue in trips to southwestern national parks. Unlike traditional site choice models which model people's decision when facing a set of single parks, we consider that each party is making a choice among a set of portfolios of parks drawn from the set of seven national parks. As established in Chapter 3, due to the characteristics of trips to national parks in the southwest the sample for this study is randomly selected onsite, which suggests that the choice of the observed portfolios is conditioned on the party's taking a recreational trip to at least one of the seven parks. An adjustment for the choice based sampling will be discussed in the next chapter.

While people do visit other parks and make side trips, this study focuses on these seven national parks, which are major destinations in the four states region. The portfolios may contain only one park, all seven parks, or any other combination of the seven parks. When constructing the portfolios, only the combinations of parks matter. The sequence of parks being visited is assumed to be irrelevant. This assumption is necessary due to practical limitations. It was possible to learn the sequence of sites visited by asking respondents to map out the route they had or would take in the survey. However, considering all combinations of parks and all possible routes to visit these parks would make it infeasible to make the set of choices/portfolios exhaustive. Instead, it is assumed that the party will simply visit the parks in the order that minimizes travel cost. Therefore, if there are M national parks in the region, the set of portfolios can be described as:

$$A = [\{1\}, \{2\}, \dots \{M\}, \{1,2\}, \{1,3\}, \dots, \{1, \dots, M\}]. \tag{14}$$

There are thus in total $K = 2^M - 1$ (for notational purposes, alternatives in this set of possible portfolios will be indexed by "k" instead of "j" in the previous section) portfolios from which the party can choose from. In this study, as there are seven parks of interest, the choice set is a set of 127 portfolios.

According to Random Utility Theory, the party chooses the portfolio of parks that maximizes its utility subject to its budget constraint. The conditional indirect utility function for party i choosing portfolio k ($k = 1, 2, \ldots, 127$) can be specified as:

$$V_{ik}^{*} = \mathrm{V}(x_k, y_i - p_{ik}, z_i) + \varepsilon_{ik} \ \forall\, k, \tag{15}$$

where y_i is the party's relevant income constraint, z_i is a vector of the demographic characteristics of respondent i, p_{ik} is the travel cost of portfolio k for party i conditional on a particular entry and exit point to the region, x_k represents a vector of observed attributes associated with the k^{th} portfolio (note that the attributes are only associated with the portfolio, not with the individuals; a full definition of x_k will be presented in the following section), and ε_{ik} is a stochastic component that captures all the unobserved factors that may contribute to the decision making. The travel cost here includes transit cost, time cost, lodging and food cost, and entrance fee(s). Since the fixed cost of entering and exiting the region does not vary across portfolios for any given party, it is not included in the travel cost.

When each party faces a choice among the *K* (=127) portfolios, they compare among the set of *K* conditional indirect utility functions and choose the alternative that yields the highest utility. Each party $i's$ choice is then defined as if they are solving the following problem:

$$\underset{k \in S}{Max} \quad V_{ik}(x_k, p_{ik}, y_i, z_i), \tag{16}$$

where *S* is the set of *K* portfolios.

4.3 Specific Models

Different choice models can be derived by using different assumptions about the distribution of the unobserved component ε_i. In this section, I present the three empirical versions of choice models used in this study: (i) Standard Logit with Additive Site Utilities (ASU-SL), (ii) Mixed Logit with Additive Site Utilities (ASU-MXL), and (iii) Portfolio Specific Constants as Utilities (PSC).

4.3.1 Additive Site Utilities Models – SL Model and MXL Model

Recall from the previous section that individual $i's$ choice can be specified as $V_{ik}^{*} = \mathrm{V}(x_k, y_i - p_{ik}, z_i) + \varepsilon_{ik}\ \forall\ k$. For the standard logit model with additive site utilities, it is important to understand how each park contributes to the indirect utility. Therefore, the alternative attributes vector x_k is specified as a vector $(x_{k\cdot})$ of M (= 7 parks in this study) index variables, where:

$$x_{k\cdot} = (x_{k1,} \dots, x_{km})' \tag{17}$$

and $x_{km} = 1$ if park m is in the k^{th} portfolio, and 0 otherwise.

Assuming that the park index variables are additively separable from the remaining deterministic components of the indirect utility, the conditional indirect utility function of individual i choosing portfolio k can be written as:

$$V_{ik}^{*}(\beta, \gamma) = \beta x_{k\cdot} + f(\,(y_i - p_{ik}),\ z_i, \gamma) + \varepsilon_{ik}\ \forall\ k, \tag{18}$$

where β and γ are vectors of unknown parameters, with $\beta x_{k\cdot} = \sum_{m=1}^{M} \beta_m x_{km}$. As x_{km} is a vector of index variables which only equals 1 if park *m* is in portfolio *k*, the conditional indirect utility function simplifies to:

$$V_{ik}^{*}(\beta,\gamma) = \sum_{m \in S_k} \beta_m + f(\,(y_i\ - p_{ik}),\ z_i, \gamma) + \varepsilon_{ik}\ \forall\, k, \qquad (19)$$

where S_k is the set of parks in portfolio *k*. In this way, each park *m* contributes to an individual's utility by adding its parameter β_m to the total utility when park m is in the k^{th} portfolio; this is why the model is named "Additive Site Utilities." Consider the following example; for portfolio $k = 1$ which includes only Arches ($m = 1$), the utility entry for individual i visiting this portfolio is $V_{i1}^{*} = \beta_1 + f(\,(y_i\ - p_{i1}),\ z_i, \gamma) + \varepsilon_{i1}$, so the portfolio utility includes a "utility hit" only from Arches. Say, if individual i visited portfolio $k = 10$, which contains both Arches ($m = 1$) and Grand Canyon ($m = 4$), then the utility entry for individual i visiting portfolio 10 is $V_{i10}^{*} = \beta_1 + \beta_4 + f(\,(y_i\ - p_{i10}),\ z_i, \gamma) + \varepsilon_{i10}$. In this case, there are two "utility hits," from Arches and Grand Canyon separately.

4.3.1.1 Standard Logit Model

The easiest and most widely used discrete choice model is the standard logit model. The logit formula was first developed by Luce (1959) based on the assumption of independence from irrelevant alternatives (IIA) property, and then completed by McFadden (1974). It is derived under the assumption that the unobserved components (ε_i) are independently and identically distributed (IID) type-I extreme values, which suggests that the unobserved components are uncorrelated over all alternatives and have the same variance. The density function for each unobserved component of utility is then given by:

$$f(\varepsilon_{ik}) = e^{-\varepsilon_{ik}} e^{-e^{-\varepsilon_{ik}}}, \tag{20}$$

and the cumulative distribution is then given by:

$$F(\varepsilon_{ik}) = e^{-e^{-\varepsilon_{ik}}}. \tag{21}$$

The probability of individual i choosing portfolio k, based on the choice probability derived in the first section of this chapter, can be written as:

$$\begin{aligned} P_{ik} &= Prob\left(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \; \forall \, j \neq k\right) \\ &= Prob\left(\varepsilon_{ij} < \varepsilon_{ik} + V_{ik} - V_{ij} \; \forall \, j \neq k\right) \\ &= \int I(\varepsilon_{ij} < \varepsilon_{ik} + V_{ik} - V_{ij} \forall \, j \neq k) \cdot f\,(\varepsilon_{ik}) \cdot d_{\varepsilon_{ik}}. \end{aligned} \tag{22}$$

Since the error terms are independent from one another, the probability that $\varepsilon_{ij} < \varepsilon_{ik} + V_{ik} - V_{ij}$ is true for all $j \neq k$ is the product of the individual cumulative distribution for each ε_{ij} $(j \neq k)$ evaluated at $\varepsilon_{ik} + V_{ik} - V_{ij}$, or:

$$I(\cdot) = \prod_{j \neq k} F\left(\varepsilon_{ij} < \varepsilon_{ik} + V_{ik} - V_{ij}\right) = \prod_{j \neq k} e^{-e^{-(\varepsilon_{ik} + V_{ik} - V_{ij})}}.$$

(23)

By substituting equations (20) and (23) into equation (22) and engaging in some algebraic manipulation, a closed form expression of the choice probability can be written as[2]:

$$P_{ik} = \frac{e^{V_{ik}}}{\sum_j e^{V_{ij}}} = \frac{e^{\beta x_{k\cdot} + f\left((y_i - p_{ik}),\, z_i, \gamma \right)}}{\sum_j e^{\beta x_{j\cdot} + f\left((y_i - p_{ij}),\, z_i, \gamma \right)}}. \tag{24}$$

The indirect utility functions are usually considered to be linear; that is, it is assumed that the individual characteristics and disposable income are linearly additive in $f\left((y_i - p_{ij}),\ z_i, \gamma \right)$. Thus the choice probability can be written as:

$$P_{ik} = \frac{e^{\beta x_{k\cdot} + \gamma_1 (y_i - p_{ik}) + \gamma_2 z_i}}{\sum_j e^{\beta x_{j\cdot} + \gamma_1 (y_i - p_{ij}) + \gamma_2 z_i}} = \frac{e^{\beta x_{k\cdot} + \gamma_1 p_{ik}}}{\sum_j e^{\beta x_{j\cdot} + \gamma_1 p_{ij}}}. \tag{25}$$

Equation (25) shows that the choice probability of individual i choosing portfolio k only depends on the portfolio attributes (the parks in portfolio k and the price of visiting portfolio k) and not on any individual characteristics.

As the logit choice probability takes a closed form, equation (19) can be estimated using the traditional maximum – likelihood method. Assuming that the sample is an exogenous random draw[3] and that visitors’ choices on which portfolio to

---

[2] Derivation of these logit probabilities can be found in Trains (2009).

[3] As noted in Chapter 3, the sample selected for this study is not exogenous. It is instead a choice-base sample. However, after the weighting procedure to correct for

visit are independent from one another, the probabilities of each visitor in the sample choosing the portfolio she was observed to choose is:

$$L(\beta^*) \ = \ \prod_{i=1}^{I} \prod_k (P_{ik})^{I_{ik}}, \qquad (26)$$

where β^* is a vector of all the parameters in the model and I_{ik} is an index variable which equals 1 if visitor i chose portfolio k and 0 otherwise. Since only one portfolio can be chosen at each choice occasion, $\prod_k (P_{ik})^{I_{ik}}$ is simply the probability of the chosen alternative. The log likelihood function is thus:

$$LL(\beta^*) \ = \ \sum_{i=1}^{I} \sum_k I_{ik} ln(P_{ik}). \qquad (27)$$

and $LL(\beta^*)$ is maximized with respect to β^* to obtain the parameter estimates for the model.

The IID type-I extreme value assumption on the error term makes the calculation of the choice probabilities much simpler. This is almost certainly why the logit model is the most popular basic choice model. However, this simplicity and convenience come at a cost of restrictions on modeling realistic choice occasions (Train, 2009). First, the logit model implies proportional substitution across alternatives. The independent irrelative alternatives assumption implies that the ratio of choice probabilities between two alternatives always remains the same regardless of changes in the attributes of any alternatives other than the two, or $P_{ik}/P_{in} =$

---

choice-based issue (will be discussed in Chapter 5), the weighted sample can be considered an exogenous random draw.

$\frac{e^{V_{ik}}}{\sum_j e^{V_{ij}}} \Big/ \frac{e^{V_{in}}}{\sum_j e^{V_{ij}}} = e^{V_{ik}}/e^{V_{in}} = e^{V_{ik} - V_{in}}$. This is very unrealistic in any real choice scenarios. Second, the IID assumption suggests that unobserved factors are independent over time in repeated choice situations, that is, multiple decisions made by the same choice makers are uncorrelated over time. In this study, we only model a one-time choice decision among visitation decisions; thus, this limitation would not have a great effect on our results. Last but not the least, the basic logit model fails to capture taste varieties among individuals. It assumes all choice makers have homogeneous preferences; that is, individuals have the exact same tastes over each attribute of the alternatives. This is very unrealistic when it comes to real decision-making scenarios. For example, low-income households may be more concerned about trip costs than high-income households. People who have more flexible time may also be more sensitive to trip costs. Another limitation of the standard logit model is that it cannot account for the correlation among error terms associated with portfolios that have common sites. For example, having Bryce Canyon (BC) in the portfolio almost certainly matters more to some visitors than others. There may be certain features in Bryce Canyon which greatly appeal to some visitors, in which case having it in the portfolio will have a more substantial impact on their utility than it would for other visitors. Then for these parties, all portfolios containing BC will have a higher than average "utility hit" of BC and therefore higher than average error terms for all these portfolios. Therefore, all portfolios that contain BC will have correlated error, which violates the IIA assumption.

One simple way to solve the heterogeneous taste issue is to modify the standard logit model by including interaction terms. The most commonly used method is to interact alternative attributes with individual characteristics. Consider the possibility that the variables in $f(\,(y_i\ - p_{ik}),\ z_i, \gamma)$ are no longer linearly additive, instead taking the form $\gamma \cdot p_{ik} \cdot z'_i$, where z'_i is a vector of $[1, y_i,\ z_i]$ and γ is the corresponding coefficient vector $[\gamma_p, \gamma_{py}, \gamma_{pz}]$. The conditional indirect utility function can then be specified as:

$$V_{ik}^{*}(\beta, \gamma) = \beta x_{k\cdot} + \gamma \cdot p_{ik} \cdot z'_i + \varepsilon_{ik}\ \forall\, k, \tag{28}$$

and the choice probability of individual i choosing portfolio k becomes:

$$P_{ik}\ =\ \frac{e^{\beta x_{k\cdot} + \gamma \cdot p_{ik} \cdot z'_i}}{\sum_j e^{\beta x_{j\cdot} + \gamma \cdot p_{ij} \cdot z'_i}}. \tag{29}$$

Introducing this interaction term into the utility function can capture heterogeneous preferences on alternative prices due to certain individual characteristics. In this way, the preference heterogeneity is explained systematically. For instance, consider the interaction term between household income and portfolio prices, $\gamma_{py}\ (p_{ik} \cdot y_i)$. This will capture any heterogeneous preference on the alternative price due to differences in household income. The marginal utility of portfolio prices is

$$\partial V_{ik}^{*} \big/ \partial p_{ik} = \gamma_p + \gamma_{py} \cdot y_i, \tag{30}$$

where γ_p captures the average marginal utility of portfolio prices over all parties and $\gamma_{py} \cdot y_i$ adjust the average marginal utility due to household incomes. We would expect that higher income parties are less sensitive to portfolio prices. Therefore, γ_{py} should be negative.

Although the standard logit model with interaction terms can accommodate some taste variances and capture the heterogeneities that relate to observed characteristics of the decision makers, it still does not solve the issue of having correlation among the error terms due to common parks in the portfolios. Several more flexible models have been developed to avoid the limitations of standard logit, such as the probit model, nested logit model, and mixed logit model (McFadden & Train, 2000).

4.3.1.2 Mixed Logit Model

The Mixed logit (or random parameter logit) model is a highly flexible model which further generalizes the logit model while relaxing some of the restrictions of the standard logit model's IIA assumption for the error terms. The technique behind it was developed by McFadden and Train (2000) and is by far the most widely accepted generalized form of the logit model. It allows for full preference heterogeneity and correlation in error terms, and under this version of the logit model the substitution pattern is no longer necessarily a fixed proportion.

MXL Model Specification

To allow for full preference heterogeneity and correlation in error terms, the mixed logit model allows the parameters associated with the explanatory variables to vary across the population according to some probability distribution. Recasting the standard logit model (28) within a mixed logit framework, the conditional indirect utility function changes very subtly by adding a party-specific subscript to β_m, giving a condition utility function of:

$$V^*_{ik}(\beta,\gamma) = \sum_{m=1}^{M} \beta_{im} x_{km} + \gamma \cdot p_{ik} \cdot z'_i + \varepsilon_{ik} \;\; \forall\, k, \tag{31}$$

where ε_{ik} is still a random term that is IID extreme value. Instead of estimating an average "utility hit" β_m which is constant across parties, the mixed logit model estimates a β_{im} which is variable by party i. This specification allows the contribution of a particular park to the conditional indirect utility to vary over parties, thus capturing tastes variations for each park. Each party is still assumed to know their own preference β_{im}. The researcher, however, while aware of variation in preferences, cannot directly observe β_{im}. The variation can be represented as a mean effect plus a deviation from the mean, where the deviation varies over parties. The preference heterogeneity of travel cost is still systematically represented using the interaction terms defined in the previous section. The coefficients of travel cost and travel cost interaction terms are fixed across parties, assuming that parties' travel cost preference variations only come from the observed individual characteristics.

As the error term is IID extreme value, the choice probability of individual i choosing portfolio k conditional on knowledge of the vector $\beta_{i\cdot} = [\beta_{i1}, \dots, \beta_{iM}]$ still follows the traditional logit specification:

$$P_i(k \mid \beta_{i\cdot}, \gamma) = \frac{e^{V_{ik}}}{\sum_j e^{V_{ij}}} = \frac{e^{\beta_i x_{k\cdot} + \gamma \cdot p_{ik} \cdot z'_i}}{\sum_j e^{\beta_i x_{j\cdot} + \gamma \cdot p_{ij} \cdot z'_i}}\,. \tag{32}$$

However, since the researcher does not know $\beta_{i\cdot}$, the choice probability cannot be conditional on $\beta_{i\cdot}$. $\beta_{i\cdot}$ is a random variable with a density function of $f(\beta_{i\cdot}|\psi)$, where ψ is a set of parameters representing the distribution of $\beta_{i\cdot}$. The unconditional choice probability of individual i choosing portfolio k can now be specified as an integral of the conditional choice probability over all possible values of the unknown $\beta_{i\cdot}$:

$$P_i(k \mid \psi, \gamma) = \int P_i(k \mid \beta_{i\cdot}, \gamma)\, f(\beta_{i\cdot}|\psi) d\beta_{i\cdot}. \tag{33}$$

Intuitively, this is the probability that party i chooses portfolio k conditional on a prior knowledge of the distribution of $\beta_{i\cdot}$.

The researcher must specify a distribution for the coefficients that satisfies his expectation of the choice behavior. The most commonly used distributions are the normal, lognormal, triangular, and uniform distributions. The lognormal distribution is generally used when the same sign of the parameter is expected for every decision maker; for example, travel cost coefficient is usually expected to be negative for all decision makers. When the sign is uncertain, the normal distribution is usually used, with $\beta_{i\cdot} \sim (b, \sigma)$ where mean b and standard deviation σ are estimated. The triangular and uniform distribution are normally applied to cases where the researcher needs to bound both sides of the distributions to avoid unreasonably large coefficients draw from the tails for some decision makers (Hensher & Greene, 2003; Train, 2009). With the uniform or triangular distribution, the coefficients were bounded between $b - s$ and $b + s$ in both cases, where b and s are, respectively, the mean and spread. The only difference is that with the uniform density, the coefficients are distributed uniformly within the bounds, while with the triangular distribution, the coefficients first rises linearly from $b - s$ to b and then decreases linearly to $b + s$. Hensher & Greene (2003) and Train (2009) both discussed the choice of density functions in more detail.

Since parties' preferences on having certain parks included in the portfolio are not necessarily positive or negative, it is reasonable to assume that the random parameters $\beta_{i\cdot}$ follows a normally distributed density function of $f(\beta_{i\cdot}|b,\sigma)$, where b is a vector of means for each park and σ is the standard deviation that varies across parties. This mixed logit model framework allows a pattern of correlation across the portfolios sharing common site(s). At the same time, this framework allows the "utility hit" for a given park in the portfolio to vary stochastically across decision makers.

We can specify equation (31) in a slightly different form to illustrate the correlations across portfolios. Note that while the utility specification in (31) is written as a random parameters model, it can equivalently be viewed as an error component specification (Train, 2009). Given that β_{im} is randomly distributed, it can be decomposed as the following:

$$\beta_{im} = b_m + \mu_{im} \tag{34}$$

where b_m is the mean and μ_{im} is the deviation from the mean. The indirect utility function can then be rewritten as:

$$\begin{aligned} V_{ik}^{*}(\beta,\gamma) &= \sum_{m=1}^{M} b_m x_{km} + \gamma \cdot p_{ik} \cdot z'_i + \sum_{m=1}^{M} \mu_{im} x_{km} + \varepsilon_{ik} \ \forall\, k, \\ &= \sum_{m=1}^{M} b_m x_{km} + \gamma \cdot p_{ik} \cdot z'_i + \widetilde{\varepsilon_{ik}} \ \forall\, k, \end{aligned} \tag{35}$$

where

$$\widetilde{\varepsilon_{ik}} = \sum_{m=1}^{M} \mu_{im} x_{km} + \varepsilon_{ik} = \sum_{m\in S_k} \mu_{im} + \varepsilon_{ik}\,, \tag{36}$$

where, again, S_k denotes the set of sites in portfolio k. Written in this form, it is clear that $\widetilde{\varepsilon_{ik}}$`s are no longer independent; they are correlated across the portfolios that include the same site m. All portfolios with site m share the same component μ_{im}. A positive random component μ_{im} suggests that the unobserved characteristics of party i make it prefer site m more than the average party does. Conversely, a negative μ_{im} suggests that their unobserved characteristics make party i enjoy site m less than the average party does.

A simple specification of this mixed logit model is to assume that within each portfolio, there is no correlation across sites, i.e., $\mu_{im} \sim N(0, \sigma_m^2)$ with $Cov(\mu_{im}, \mu_{in}) = 0 \ \forall \ n \neq m$. In this case, the variance of the indirect utility associated with the choice of portfolio k comes from two stochastic components, $\varepsilon_{ik} \sim N\ (0, \sigma_\varepsilon^2)$ and $\sum_{m=1}^{M} \mu_{im} x_{km}$. For the system as a whole, the disturbance covariance matrix is equal to:

$$\Omega_{ik} = \sigma_\varepsilon^2 \cdot \mathrm{I}_T + \ X_{k\cdot} \cdot W \cdot X'_{k\cdot}\,, \tag{37}$$

where I_T is an identity matrix, $X_{k\cdot}$ is a vector of the portfolio's attributes, and the variance covariance matrix $W = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_M^2 \end{bmatrix}$. The variance can be expressed as

$$Var(V_{ik}) = \sum_{m=1}^{M} \sigma_m^2\, x_{km}^2 + \sigma_\varepsilon^2 = \sum_{m \in S_k} \sigma_m^2 + \sigma_\varepsilon^2, \tag{38}$$

Thus, even though σ_ε^2 remains identical across portfolios, the variances of the whole stochastic term are no longer identical due to $\sum_{m \in S_k} \sigma_m^2$. With different set of sites included in the portfolios, $\sum_{m \in S_k} \sigma_m^2$ changes based on the specific portfolio, causing the overall variance to change along with it.

Beyond simply having nonidentical variance terms, this specification also allows the covariances across portfolio to no longer equal to zero, even though it is assumed that there is no correlation across sites (i.e., $Cov(\beta_{im}, \beta_{in}) = 0 \; \forall \; n \neq m$) and $Cov(\varepsilon_{ik}, \varepsilon_{ij}) = 0 \forall \; j \neq k$). Any pair of portfolios that contain common sites will have non-zero covariances. The covariance between portfolios is given by:

$$Cov(V_{ik},\, V_{ij}) = \sum_{m=1}^{M} \sigma_m^2 x_{km} x_{jm} = \sum_{m \in S_k \cap S_j} \sigma_m^2 \,. \tag{39}$$

Consider the following simple example. Suppose that there were three sites available to visit; then, the matrix of correlations among the $(2^3 - 1)$ portfolios would be as follows:

$$\begin{array}{c} \{1\} \\ \{2\} \\ \{3\} \\ \{1,2\} \\ \{1,3\} \\ \{2,3\} \\ \{1,2,3\} \end{array} \begin{bmatrix} \sigma_1^2+\sigma_\varepsilon^2 & & & & & & \\ 0 & \sigma_2^2+\sigma_\varepsilon^2 & & & & & \\ 0 & 0 & \sigma_3^2+\sigma_\varepsilon^2 & & & & \\ \sigma_1^2 & \sigma_2^2 & 0 & \sigma_1^2+\sigma_2^2+\sigma_\varepsilon^2 & & & \\ \sigma_1^2 & 0 & \sigma_3^2 & \sigma_1^2 & \sigma_1^2+\sigma_3^2+\sigma_\varepsilon^2 & & \\ 0 & \sigma_2^2 & \sigma_3^2 & \sigma_2^2 & \sigma_3^2 & \sigma_2^2+\sigma_3^2+\sigma_\varepsilon^2 & \\ \sigma_1^2 & \sigma_2^2 & \sigma_3^2 & \sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_3^2 & \sigma_2^2+\sigma_3^2 & \sigma_1^2+\sigma_2^2+\sigma_3^2+\sigma_\varepsilon^2 \end{bmatrix} \tag{40}$$

In general, the more common sites the two portfolios share, the greater the correlation is between the portfolios. For instance, $Cov(V_{i,\{1,2\}},\, V_{i,\{2,3\}}) = \sigma_2^2$ and $Cov(V_{i,\{1,2,3\}},\, V_{i,\{2,3\}}) = \sigma_2^2 + \sigma_3^2$. Portfolios that share no common park, portfolio {1} and portfolios {3} for example, have a correlation equal to 0. From the correlation matrix, we can see the innovation of the portfolio choice model. The correlation structure of a traditional site choice model, where only one site is chosen at a time, is simply the left top corner of (40), $\begin{bmatrix} \sigma_1^2+\sigma_\varepsilon^2 & 0 & 0 \\ 0 & \sigma_2^2+\sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_3^2+\sigma_\varepsilon^2 \end{bmatrix}$. The rest of the correlation matrix (40) results from the combined choices of multiple sites.

The simple specification thus far suggests that the portfolios are only correlated when they share at least one common site, under the assumption that the μ_{im}'s themselves are uncorrelated across sites. However, it is possible that preferences over sites may be correlated due to common features in those sites. For instance, some national parks in the choice set provide well-designed hiking trails compared to other parks. A group of hikers would then prefer all these parks with good hiking opportunities and so the positive deviation μ_i associated with each of these good hiking sites should be correlated. This feature can be integrated into this mixed logit model by allowing for correlations among sites. The variance – covariance matrix of μ_{im} now becomes $W = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1M}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{M1}^2 & \cdots & \sigma_{MM}^2 \end{bmatrix}$. Herriges and Phaneuf (2002), modeling single site trips, use an error components approach to induce pair wise correlation among alternative sites.

Estimating MXL Model

Recall that for the standard logit model, the typical approach is to estimate the unknown parameters of the model by maximizing the log likelihood function in (27), where the probability of individual choosing alternative k, $P_{ik} = \frac{e^{\beta x_{k\cdot}+\gamma\cdot p_{ik}\cdot z'_i}}{\sum_j e^{\beta x_{j\cdot}+\gamma\cdot p_{ij}\cdot z'_i}}$, takes a closed form. In principle, we would want to take the same approach with the mixed logit model. However, consider the log likelihood function for the mixed logit model, which is given by:

$$LL(\psi,\gamma) = \sum_{i=1}^{N} I_{ik} \ln P_i(k\,|\psi,\gamma)$$
$$= \sum_{i=1}^{N} I_{ik} \ln(\int P_i(k\,|\beta_{i\cdot},\gamma)\, f(\beta_{i\cdot}|b,W)\, d\beta_{i\cdot}), \quad (41)$$

where $N(= i \times k)$ is the total number of observations, ψ, again, is the set of parameters (b, W) that defines the normal distribution of $\beta_{i\cdot}$, and $P_i(k\,|\beta_{i\cdot},\gamma) = \frac{e^{\beta_i x_{k\cdot} + \gamma \cdot p_{ik} \cdot z'_i}}{\sum_j e^{\beta_i x_{j\cdot} + \gamma \cdot p_{ij} \cdot z'_i}}$. It will not be possible to maximize the log-likelihood in this form because there is no closed form for the integral. To address this problem, researchers have found several simulation methods that can satisfactorily evaluate this integral form (Gourieroux & Monfort, 1996; Greene, 2008; Train, 2009). In this analysis we use the following simulation estimation technique, known as maximum simulated likelihood estimation (MSL). MSL replaces the integral probability with an approximation simulated for any given value of ψ. It first randomly draws a value from the given distribution $f(\beta_{i\cdot}|b,W)$ and then uses this value to calculate the choice probability $P_i(k\,|\beta_{i\cdot}^r,\gamma)$. This process is then repeated many times and finally the results are averaged to get the simulated probability:

$$\breve{P}_i(k\,|\psi,\gamma) = \frac{1}{R}\sum_{r=1}^{R} P_i(k\,|\beta_{i\cdot}^r,\gamma), \quad (42)$$

where R is the number of random draws and $\beta_{i\cdot}^r$ is the random drawing value from $f(\beta_{i\cdot}|b,W)$, where $r = 1$ refers to the first draw. With a sufficient number of random draws, $\breve{P}_i(k\,|\psi,\gamma)$ is an unbiased estimator of $P_i(k\,|\psi,\gamma)$[4]. The properties of this

[4] Note that although $\breve{P}_i(k\,|\psi,\gamma)$ is an unbiased simulator of $P_i(k\,|\psi,\gamma)$, $\ln \breve{P}_i(k\,|\psi,\gamma)$ is not unbiased to $\ln P_i(k\,|\psi,\gamma)$. This is due to the fact that the log operation is not a linear transformation. The biased log estimator will enter the simulated log-likelihood function and cause biased estimates. Gourieroux and Monfort (1996) point out that if

unbiased approximation are very desirable: its variance decreases as R increases, it is strictly positive, it is twice differentiable in the parameters ψ and the variables x, and, finally, it always sums to one over alternatives (Train, 2009).

Inserting $\breve{P}_i(k \mid \psi, \gamma)$ into equation (41), the MSL then maximizes the following simulated likelihood function with respect to ψ and γ:

$$SLL(\psi, \gamma) \; = \; \sum_{i=1}^{I} \sum_k I_{ik} \ln \breve{P}_i(k \mid \psi, \gamma). \tag{43}$$

The maximum simulated likelihood estimator (MSLE) is the solution to the maximization problem. It is found by equating the derivatives to zero.

4.3.2 Portfolio Specific Constant Model (PSC)

An alternative model used to estimate the portfolio choices is the alternative specific constant model, in this case, the portfolio specific constant model. As discussed in the previous section, in choice utility models utility is only impacted by the differences between the alternative attribute levels, rather than the absolute levels themselves. The same applies to the alternative specific constants. Therefore, when including the constants for each alternative, one should be normalized to zero as the baseline, and the rest of the alternative constants can be interpreted relative to that

---

the number of random draws R rises at the same rate with the square root of the sample size N, then the simulation bias disappears and MSL is consistent. Train (2009) also mentions that if R rises faster than $\sqrt{N}$ then the MSL is not only consistent but also efficient. However, if R is fixed, then the MSL is no longer consistent, which is the main limitation of the MSL method.

normalized alternative. In other words, with K alternatives, there can only be $K - 1$ constants in the model.

The reason to model portfolio specific utility is to introduce the unobserved correlation across sites in each portfolio. The conditional indirect utility function of party i choosing alternative k can be written as:

$$V_{ik}^{*}(\alpha, \gamma) = \sum_{k=1}^{K-1} \alpha_k A_k \ + \gamma \cdot p_{ik} \cdot z'_i + \varepsilon_{ik} \ \forall \ k. \tag{44}$$

where A_k is a dummy variable for alternative k and α_k is the corresponding parameter. Since only one alternative can be chosen per choice occasion, the summation of the multiple alternatives can be reduced to a single constant, equal to α_k – the portfolio specific constant.

In the additive site utility models, i.e., the SL and MXL, the utilities are additive in a sense that each park in portfolio k contributes to the overall utility of choosing portfolio k by adding a "utility hit" of β_m (or β_{im} in the MXL model). The size of the "hit" from park m is irrelevant to the presence of other parks in the portfolio. However, this may not hold in real site choice scenarios. For example, if two parks which have features that complement each other are in the same portfolio, the combination of the two should give a bigger "utility hit" than the sum of the two parks' separate "hits." On the other hand, if two parks happen to be substitutes to one another, then having both of them in the same portfolio may lower the utility of that portfolio, i.e., the combined "utility hit" may be less than the sum of the two separate "hits." The complementarity case might include sites that satisfy a diversity of interests, for instance, one that has canyons and another that has special wildlife. The substitution case might have sites that both include canyons or are otherwise similar. The PSC specification introduces combined sites effects in the utility function and let the combination of sites interact in such a way that the "utility hit" for one site varies depending on the other sites in the portfolio.

The estimation of the PSC model follows the traditional maximum log-likelihood method for the standard logit model. The log-likelihood function is given by:

$$LL(\theta) \ = \ \sum_{i=1}^{I} \sum_{k} I_{ik} \ln \frac{e^{\alpha_k + \gamma \cdot p_{ik} \cdot z'_i}}{\sum_j e^{\alpha_k + \gamma \cdot p_{ij} \cdot z'_i}}, \tag{46}$$

where θ is the set of parameters. At the maximum of the likelihood function, the derivative with respect to each parameter equals zero. For the alternative specific constants, the first-order condition is:

$$\sum_i \sum_k (I_{ik} - P_{ik}) = 0. \tag{47}$$

Rearranging and dividing both sides by the number of observation N, equation (47) becomes:

$$\frac{1}{N} \sum_i \sum_k I_{ik} = \frac{1}{N} \sum_i \sum_k P_{ik}. \tag{48}$$

The left hand side of the equation is the share of people in the sample who are observed choosing alternative k, while the right hand side of the equation is the predicted share for alternative k. This brings up one of the shortcomings of the PSC model, which is it can only estimate the constants associated with the portfolios that have been chosen by at least one decision maker (Newman, Ferguson & Garrow, 2012).

Chapter 5

CHOICE-BASED SAMPLING

The previous discussion on logit model estimations was based on an assumption of an exogenous or random sample. However, the sample collected for this national park visitation study is not entirely exogenous. As explained in Chapter 3, due to the fact that visitors to these national parks are from all over the U.S. (and the world) and are mostly one-time visitors, the probability of contacting a real visitor through random phone calls or mail is extremely small. A random sample would thus need to be extremely large and prohibitively costly to assure a reasonable amount of park visitors being selected. Therefore, instead of randomly sampling people all over the U.S., the sample was selected on-site at each national park of interest, which makes the sample endogenously stratified. This type of sample is usually referred to as a choice-based sample.

The concept of choice-based sampling was first considered by Warner (1963) in the context of transportation demand. He pointed out that in the case of a hypothetical choice of transportation mode problem, the sample selected are usually from the group of existing travelers who had chosen one of the modes being considered. This sampling method is considerably less costly and can efficiently collect sufficient amount of data for those infrequently chosen alternatives. The application of choice-based sampling has become widely used in areas other than transportation decision problems.

However, as Warner (1963) cautioned, even though choice-based sampling is less costly and more efficient in certain ways, it can be problematic when it comes to estimation. In a choice-based sample, the probability of a member entering the sample now depends on the outcome of the decision makers' choice, instead of being fully random. Take this national parks study as an example, since the data collection was done on-site at each national park gate, the probability of an individual being recruited for the survey is based on the portfolio of parks the person chose to visit. A person who picks the portfolio of visiting all seven parks and spends a day at each park during the two-week survey period is more likely to be included in the sample than someone who chose to visit only one park for a day and then left the region. This means that the correct likelihood function will depend upon both the standard choice probability and the probability that a given observation enters the sample. Thus, the sampled parties' visitation behaviors are derived from a different probability distribution from the one that exists for the general population. The sampling distributions associated with these observations are no longer random, but rather weighted/size-biased. If not treated carefully, this type of endogenous stratification can result in biased parameter estimates and misleading welfare measures (Shaw 1988; Englin & Shonkwiler 1995).

Most site choice studies encounter choice-based sampling in a relatively simpler context than the Southwestern National Park study. Traditionally, the alternatives in a site choice study are individual sites instead of portfolios of sites. In the rest of this section, I first outline the problem in the simpler case where the alternatives are single sites, along with treatments designed for this type of choice-based sampling. I then present the modified method used to correct for our choice-base sampling, where the alternatives are portfolios of sites.

5.1 Choice-based Sampling When the Alternatives Are Single Sites

Consider a simple case where decision makers are facing a choice between only two sites, A and B, and suppose that 85% of the population choose A and 15% choose B. Now suppose that for practical reasons, 50% of the sample is randomly selected from people who choose A and 50% of the sample is randomly chosen from people who choose B. The selected sample is therefore not an accurate reflection of the population; people who choose B are over-sampled by 0.5/0.15, while people who choose A are under-sampled by 0.5/0.85.

One straightforward way to correct for this type of sampling is to use the weighted exogenous sampling maximum likelihood (WESML) estimator designed by Manski and Lerman (1977). The WESML estimator is simply:

$$w(j) = {}^{H(j)}\!\big/_{S(j)}, \tag{49}$$

where $H(j)$ is the population probability of choosing alternative j, and $S(j)$ represents sample share who choose alternative j. The weights $w(j)$ are therefore non-negative constants. To make this estimator more specific to a simple site choice problem, define $H(j) = N_j/N$, where N_j is the total number of choices of site j in the population and N is the total number of choices made in the population among all sites and define $S(j) = S_j/S$ where S_j is the number of individuals sampled at site j and S is the total number of individual sampled at all sites. If $S(j) = H(j)$, there is no bias introduced by on-site sampling. However, when $S(j) \neq H(j)$, as one would expect with on-site sampling, the sample choices share does not accurately reflect the actual pattern of choices in the general population. At sites where $S(j) > H(j)$, visitors are over-sampled and at sites where $S(j) < H(j)$, visitors are under-sampled. To fix the over/under-sampling, it is necessary to re-weight each observation by the WESML estimator $w(j)$. Continuing the example defined at the beginning of this section, people who choose B are over-sampled by a factor of 0.5/0.15 and therefore will be corrected by "weighting down" by 0.3 = $H(j)/S(j)$ = 0.15/0.5. Similarly, those who choose A are under-sampled by 0.5/0.85 and therefore need to be "weighted up" by a factor of 1.7 = $H(j)/S(j)$ = 0.85/0.5. The weight estimator then directly enters the log likelihood function. The weighted log likelihood function can be stated as:

$$LL(\theta) = \sum_{i=1}^{S} \frac{H(j_i)}{S(j_i)} \ln P(j_i|z_{ij}, \theta), \tag{50}$$

where $P(j_i|z_{ij}, \theta)$ is the probability of individual *i* choosing alternative j. The estimates obtained through maximizing this log likelihood function are consistent and asymptotically normal (Manski & Lerman, 1977). To use this approach, it is important that both the sampling shares $S(j)$ and the population probabilities $H(j)$ are either already known or, if not known, can be acquired through interviews with a random sample of the population.

5.2 Choice-based Sampling When Alternatives Are Portfolios of Sites

The basic approach to the portfolio-choice based sampling is the same as the site-choice based sampling, where the weight is now defined as $H(k)/S(k)$, with $k = 1, \ldots, 127$ indexing the 127 portfolios. The sample shares $S(k)$ are simply the share of parties surveyed that choose portfolio k. However, when it comes to computing the population choice probabilities, the portfolio choice model faces a more complex choice-based sampling issue. The fact that alternatives in the choice set are no longer singles sites but combinations of sites (portfolios), makes it much harder to obtain the population probability $H(k)$ for two reasons. First, we do not directly observe which portfolios parties chose, only the sites where they were interviewed. Second, parties choosing multiple-sites portfolios, by definition, can show up at any sites in their portfolios.

To deal with this complex portfolio-choice based sampling problem consider the following approach. First, assume that parties were interviewed at all parks on the same day[5]. With cash register data - a summary of the cash register tallies maintained by National Park Service at park entrances - the probability of a party in our population (group of parties that visited one of the seven national parks on "the day of sampling") being at site m, denoted $G(m)$, can be computed. Specifically, $G(m)$ can be expressed as:

$$G(m) = \sum_{k \in Z_m}^{K} \phi(m|k) \cdot H(k), \qquad (51)$$

where Z_m is the set of all portfolios that contain site m, $\phi(m|k)$ is the likelihood of the party who choose portfolio k being counted as a visitor at site m on a given day during its trip, and $H(k)$ is (again) the proportion of the population choosing portfolio k that we ultimately want to obtain.

---

[5] Even though this is not the actual interview procedure we took during the survey period, the adjustment for this is minor and will be discussed later.

Now consider the term within the summation sign, i.e. $\phi(m|k) \cdot H(k)$. This term represents the joint probability of the occurrence of both events – choosing portfolio k and being counted as a visitor at site m on the sampling day. Using Bayes Rule, one can show that:

$$\phi(m|k) \cdot H(k) = \tau(k|m) \cdot G(m), \tag{52}$$

where $\tau(k|m)$ is proportion of parties choosing portfolio k conditioned on being observed entering site m on a given day. By rearranging equation (52), one can estimate the population proportion $H(k)$ as:

$$H(k) = \frac{\tau(k|m) \cdot G(m)}{\phi(m|k)}. \tag{53}$$

Now, the whole problem boils down to obtaining the values of $G(m)$, $\tau(k|m)$, and $\phi(m|k)$. All three are obtainable using a combination of data collected through mail surveys and cash register counts at all entrances of all seven parks. $G(m)$ can be obtained by determining how many parties in our population of national park visitors on the sampling day entered park m; given this data, $G(m)$ is simply the total number of entrants to park m on the given day divided by the total population. $\tau(k|m)$ is the proportion of parties at site m on the sampling day that chooses portfolio k. This is fairly easy to obtain since people sampled at site m indicate their choices of portfolios (the combination of sites visited) in their mailing surveys. Finally, $\phi(m|k)$ is the probability of parties being observed on site m conditioned on the choice of portfolio

k. The computation for $\phi(m|k)$ may not seem straightforward at this point but can be explained using the following simple example.

Assume there are only 2 sites of interest, A and B. The set of portfolio choices therefore is {A, B, AB}. To simplify the matter, assume that when site A is visited people always stay for 2 days and when site B is visited 3 days are spent on site. Assume our population is N=10,000, and among those 20% choose portfolio {A}($H(1) = .2$), 50% choose portfolio {B}($H(2) = .5$), and the remaining 30% choose portfolio {AB} ($H(3) = .3$). Individuals are assumed to visit the region over the course of a season, $T = 100$ days. Assume that the start days of these trips are spread evenly over the region; therefore, on any random day, the number of parties starting a trip to portfolio k should be $N_k = N * H(k)/T$. The actual trip pattern would be settled after day 5, and the visitations to the three portfolios would look as follows:

Table 5.1: Example Visitation Patterns - Population

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	…
Portfolio 1																
	(A 20*	A) 20				(A 20	A) 20				(A 20	A) 20				…
		(A 20	A) 20				(A 20	A) 20				(A 20	A) 20			
			(A 20	A) 20				(A 20	A) 20				(A 20	A) 20		
				(A 20	A) 20				(A 20	A) 20				(A 20	A) 20	
					(A 20	A) 20				(A 20	A) 20				(A 20	A) 20
Portfolio 2																
	(B 50	B 50	B) 50			(B 50	B 50	B) 50			(B 50	B 50	B) 50			…
		(B 50	B 50	B) 50			(B 50	B 50	B) 50			(B 50	B 50	B) 50		
			(B 50	B 50	B) 50			(B 50	B 50	B) 50			(B 50	B 50	B) 50	
				(B 50	B 50	B) 50			(B 50	B 50	B) 50			(B 50	B 50	B) 50
					(B 50	B 50	B) 50			(B 50	B 50	B) 50			(B 50	B 50
Portfolio 3																
	(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30	B 30	B 30	B) 30	…
		(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30	B 30	B 30	B) 30
			(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30	B 30	B 30
				(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30	B 30
					(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30	B 30	B 30	B) 30	(A 30	A 30

* The numbers underneath are N_k – the number of parties in the population taking trips to portfolio k.

On any random sampling day, using day 8 (the highlighted column) as a specific example, the number of parties one would see at each site and for each portfolio is summarized in table 5.2.

Table 5.2: Visitations On A Given Sampling Day

		Portfolio			
		1 (A)	2 (B)	3 (AB)	Subtotal
Site	A	40	0	60	100
	B	0	150	90	240
	Subtotal	40	150	150	340

Using table 5.2 and the population numbers one can easily compute the probabilities of interest, as shown in tables 5.3-5.5:

1) The population probability of entering site m on the sampling day: $G(m) = \frac{N_m}{N}$, where N_m is the number of parties that enter site m on the sampling day and N is the population number;

2) The proportion of parties at site m on the sampling day that chooses portfolio k: $\tau(k|m) = \frac{N_{km}}{N_m}$, where N_{km} is the number of parties that both choose portfolio k and entered site m on the sampling day;

3) The probability of parties entering site m on the sampling day conditioned on the choice of portfolio k: $\phi(m|k) = \frac{N_{km}}{N*H(k)} = \frac{D_{mk}*N_{k/T}}{N*H(k)} = \frac{D_{mk}*(\frac{N*H(k)}{T})}{N*H(k)} = \frac{D_{mk}}{T}$, where D_{mk} is the number of days spend on site m in portfolio k.

Table 5.3: $\boldsymbol{G(m)}$

	$G(m)$
$m = A$	100/10000 = 0.01
$m = B$	240/10000 = 0.024

Table 5.4: $\boldsymbol{\tau(k|m)}$

	$\tau(k\|m)$		
	$k = 1$	$k = 2$	$k = 3$
$m = A$	40/100 = 0.4	0	60/100 = 0.6
$m = B$	0	150/240 = 0.625	90/240 = 0.375

Table 5.5: $\boldsymbol{\phi(m|k)}$

	$\phi(m\|k)$	
	$m = A$	$m = B$
$k = 1$	40/2000 = 0.02	0
$k = 2$	0	150/5000 = 0.03
$k = 3$	60/3000 = 0.02	90/3000 = 0.03

The equality in equation (52) should hold for all sites in portfolio k. Since there are multiple estimates (when portfolio k contains more than one site) of $H(k)$, one can use an average $\overline{H}(k)$ instead of $H(k)$.

$$\overline{H}(k) = \frac{1}{c_k} \sum_{m \in A_k} \frac{\tau(k|m) \cdot G(m)}{\phi(m|k)}, \tag{54}$$

where A_k is the set of sites in portfolio k and c_k is the number of sites in portfolio k. According to tables 5.3-5.5, $\overline{H}(1) = \frac{0.4*0.01}{0.02} = 0.2$, $\overline{H}(2) = \frac{0.625*0.024}{0.03} = 0.5$, $\overline{H}(3) = \frac{1}{2} * \left(\frac{0.6*0.01}{0.02} + \frac{0.375*0.024}{0.03}\right) = \frac{1}{2} * (0.3 + 0.3) = 0.3$. These estimates perfectly match the hypothetical probabilities established in this example. As a matter of fact, equation (54) can be simplified to:

$$\bar{H}(k) = \frac{1}{c_k}\sum_{m \in A_k} \frac{\tau(k|m) \cdot G(m)}{\phi(m|k)} = \frac{1}{c_k}\sum_{m \in A_k} \frac{\frac{N_{km}}{N_m} \cdot \frac{N_m}{N}}{\frac{D_{mk}}{T}} = \frac{1}{c_k}\sum_{m \in A_k} \left(\frac{N_{km}}{D_{mk}} \cdot \frac{T}{N}\right), \quad (55)$$

Since all that matters in terms of weighted sample maximum likelihood (WESML) is the relative weights and $\frac{\mathrm{T}}{\mathrm{N}}$ is a constant term that does not vary over individuals, portfolios, or site, the computation of $\bar{\mathrm{H}}(\mathrm{k})$ can be further simplified to:

$$\bar{H}(k) = \frac{1}{c_k}\sum_{m \in A_k} \frac{N_{km}}{D_{mk}}, \quad (56)$$

This example demonstrates the general steps required to compute the weights. In practice, there are a number of other adjustments needed to account for other aspects of the sampling approach:

1) Differential Sampling Rates. As noted in Chapter 3, all seven national parks were sampled over a nine-day period with various sampling rates. At some parks, people were sampled at more than one gate and the sampling rates may vary across these different gates. See Table C2 for more detail. Given these differential sampling rates, the actual observed number of parties on the sampling day cannot be directly used for computing the three probabilities of interest. For instance, continuing the previous example, suppose that the sampling rate at site A is 1/4 and sampling rate at site B is 1/5. On an average day, given the sampling rate, the actual observed visitations would be:

Table 5.6: Example Visitation Patterns – Observed (Sample)

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	…
Portfolio 1																
	(A 5	A) 5				(A 5	A) 5				(A 5	A) 5				…
		(A 5	A) 5				(A 5	A) 5				(A 5	A) 5			
			(A 5	A) 5				(A 5	A) 5				(A 5	A) 5		
				(A 5	A) 5				(A 5	A) 5				(A 5	A) 5	
					(A 5	A) 5				(A 5	A) 5				(A 5	A) 5
Portfolio 2																
	(B 10	B 10	B) 10			(B 10	B 10	B) 10			(B 10	B 10	B) 10			…
		(B 10	B 10	B) 10			(B 10	B 10	B) 10			(B 10	B 10	B) 10		
			(B 10	B 10	B) 10			(B 10	B 10	B) 10			(B 10	B 10	B) 10	
				(B 10	B 10	B) 10			(B 10	B 10	B) 10			(B 10	B 10	B) 10
					(B 10	B 10	B) 10			(B 10	B 10	B) 10			(B 10	B 10
Portfolio 3																
	(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5	B 6	B 6	B) 6	…
		(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5	B 6	B 6	B) 6
			(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5	B 6	B 6
				(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5	B 6
					(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5	B 6	B 6	B) 6	(A 7.5	A 7.5

* The numbers underneath are S_k – the number of parties in the sample taking trips to portfolio k.

Given that the target sampling rates for the interviewed survey varied at different parks and entrances from 1-in-1 to 1-in-7 and the actual sampling rates inevitably varied from the target sampling rates due to various practical issues, it is more accurate to use the averaged actual sampling rates at seven parks. The actual sampling rates r^a_{mgt} can be calculated as:

$$r^a_{mgt} = \frac{N_{\cdot mgt}}{\tilde{N}_{\cdot mgt}}, \tag{55}$$

where m represents parks $m = 1, \ldots, 7$, g represents different gates, and t is the day on which the on-site sampling was conducted. $N_{\cdot mgt}$ denotes the total number of vehicles actually interviewed on that day and $\tilde{N}_{\cdot mgt}$ denotes the total number of vehicles entering each site/gate on a given day that were eligible for interview. $\tilde{N}_{\cdot mgt}$ can be obtained using the cash register data from the NPS. Using r^a_{mgt} and the observed (sample) counts of parties following portfolio k at each site (gate) on different days, one can easily recover the population number of individuals visiting site m (gate g) on day t who choose portfolio k, given by:

$$\tilde{N}_{kmgt} = \frac{N_{kmgt}}{r^a_{mgt}}. \tag{56}$$

2) Different Sampling Days. As established in Chapter 3, the national park on-site samplings were done on different days over a nine-day period. They did not occur on the same day, in the contrast to the assumption made earlier in this chapter. Each of the seven parks was sampled on two weekdays and two weekend days, except for Grand Canyon, which was sampled on three weekdays and three weekends. To adjust for this, it was necessary to simply aggregate all $\tilde{N}_{kmgt}$ over time and take the average, or:

$$\bar{N}_{kmg\cdot} = \frac{1}{T_{\cdot mg}} \sum_{t=1}^{T_{\cdot mg}} \tilde{N}_{kmgt}, \tag{57}$$

where $T_{\cdot mg}$ denotes the number of days where interviews occurred for site m (gate g). I considered a more complex adjustment, such as accounting for weather conditions. However, I decided such adjustments were not necessary, given that all sampling days were in a relatively short time period (9 days) and each park had an equal number of week and weekend days.

3) Multiple Locations for Site Sampling. For parks that were sampled at more than one gate, the number for counts are aggregated over all gates to obtain overall site visitation numbers, i.e., $\bar{N}_{km\cdot\cdot} = \sum_g \bar{N}_{kmg\cdot}$.

4) Variation of Park Entrances. Before adjusting for the variation in park entrances, it is worth emphasizing a number of assumptions in this work. First, parties' travel time during the trip were neglected, instead assuming that individuals do not spend a full day outside of one of the parks in the choice set. In other words, each party enters at least one park per day during its trip. Second, it is not only assumed that each party enters at least one park a day, but also that they enter only one park a day. Therefore, there are no multiple parks in a single day.

Even with the above assumptions, there is still another issue that needs to be taken into consideration, which is the number of entrances to the same park. In some cases, parties may enter multiple times to the same park within a day, while in other cases, parties may enter a park only once in an overnight or multiple-day visitation. When parties enter the same park multiple times in a day, they inflate the counts of visitations for park m and portfolio k. Conversely, if a party stays overnight or for multiple days at a park and only enter the park once at the beginning, the chance of observing them at the entrance on a random day decreases.

The following modification undoes the inflation/deflation of visitation counts due to the variation of park entrances and ensures that the visitation counts follow the "one and only one entrance" condition. In brief, a "weight" is introduced to correct for the variation of park entrances. Let $\overline{\omega}_{km\cdot}$ denotes the average ratio of $d_{km\cdot}$(the number of days at the park) to $\eta_{km\cdot}$ (the number of park entrances) for each park/portfolio combination. Then:

$$\breve{N}_{kmg\cdot} = \overline{\omega}_{km\cdot} \cdot \ \overline{N}_{kmg\cdot} = \frac{d_{km\cdot}}{\eta_{km\cdot}} \cdot \ \overline{N}_{kmg\cdot}. \tag{58}$$

For cases in which multiple-night stays are prevalent (i.e. $d_{km\cdot} > \eta_{km\cdot}$), $\overline{\omega}_{km\cdot}$ will inflate the number of counts to correct for the fact that there was a reduced chance of interviewing these parties. Conversely, for cases which involve multiple entrances on the same day to the same park (i.e. $d_{km\cdot} < \eta_{km\cdot}$), $\overline{\omega}_{km\cdot}$ deflates the number to account for the possibility of double-counting.

After all of these adjustments, the resulting estimates of $\overline{H}(k)$ could then be used in the WESML procedure to account for portfolio choice-based sampling. The likelihood function, with weights $\overline{H}(k)/S(k)$, becomes:

$$LL(\theta) = \sum_{i=1}^{S} \frac{\overline{H}(k)}{S(k)} \ln P(k_i|z_{ik}, \theta), \tag{59}$$

where $S(k)$ is the proportion of individuals surveyed that are found choosing portfolio k (i.e. $S(k) = S_k/S$) and $P(k_i|z_{ij}, \theta)$ is the relevant probability expression in the absence of choice-based sampling.

Figure 6.1 and Figure 6.2 show the trip patterns before and after the sample is weighted. In figure 6.1, number of parks visited changes significantly after weighting the sample. The sample selected on-site shows that only 29% of the trips to these national parks are single site trips and the remaining 71% are all multiple-site trips. With the weighted sample, which may be interpreted as coming from a random draw of visitors to the region during the two-week period in June, the percentage of single site trips significantly increases to 62%. This is primarily due to that there is a lower chance of sampling a visitor when he/she visits only one versus many sites. Therefore, in the weighted sample, all single site trips are "weighted up" and multiple-site trips are "weighted down" to represent the general population. Figure 6.2 shows the visitation by parks. The percentage of Grand Canyon visitations increases after the weighting while the percentage of visitations to other parks decreases. This is partially because many trips to Grand Canyon are single-site trips.

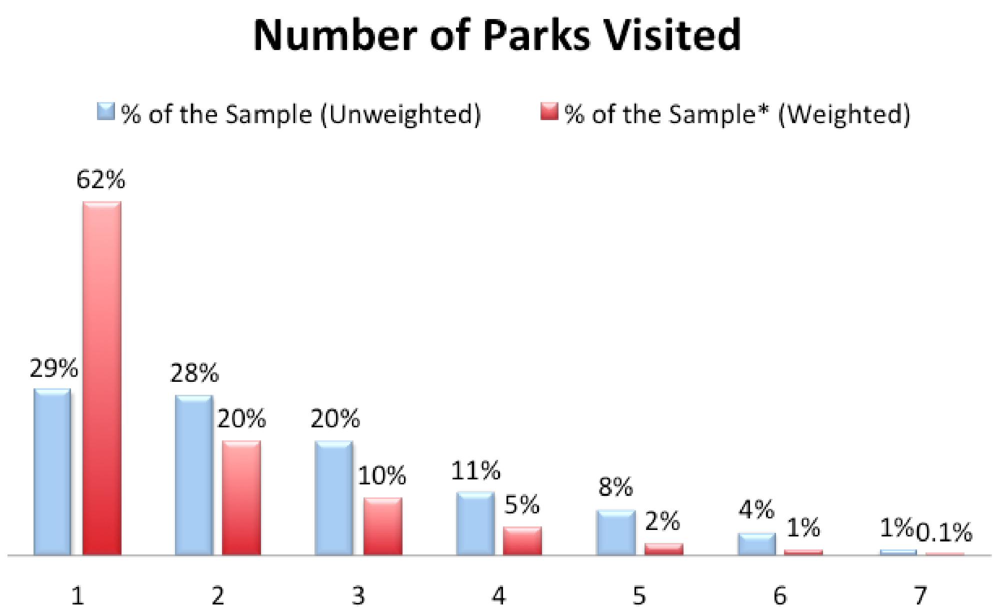


Figure 5.1: Number of Parks Visited

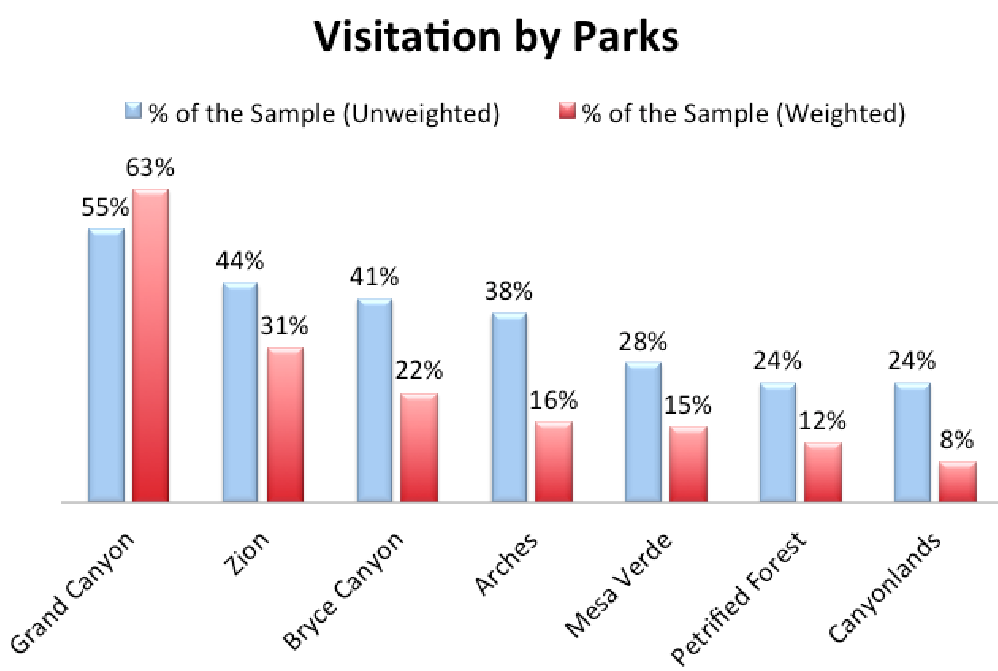


Figure 5.2: Visitation by Parks

Chapter 6

ESTIMATION RESULTS AND WELFARE ANALYSIS

In this chapter, I present the results of estimating the models discussed in Chapter 4 along with a welfare analysis based on those results. I estimate three models: (i) Standard Logit with Additive Site Utilities (ASU-SL), (ii) Mixed Logit with Additive Site Utilities (ASU-MXL), and (iii) Portfolio Specific Constants as Utilities (PSC). In Section 6.1, I briefly lay out the specification of the three models, and then present the coefficient estimates for each model separately. Section 6.2 presents the results of a welfare analysis based on the estimation results. This welfare analysis is one of the main purposes of this study; quantifying the true welfare losses to the public (potential park visitors) due to park closures. The welfare losses are calculated across a range of different scenarios, including single park closures and groups of park closures.

6.1 Model Specification

In Chapter 4 I discussed in detail the methodologies behind the models and their econometric properties. Therefore, here, I simply specify the composition of each model's indirect utility function. All three models include a travel cost variable and individual characteristics interacted with travel cost to pick up the heterogeneity in sensitivity to travel cost across user groups. The individual characteristics included are household income, flexible time (a dummy variable indicating when planning the trip to national parks in the southwestern region, whether the party could have chosen a longer trip or if they faced time constraints), whether the party visited recreational sites or cities other than the seven national parks[6], and potential car renter (a dummy variable indicating whether the party was likely to have rented a car given their entry and exit points to the region). In addition to the travel cost and demographic interaction variables, the ASU models include a set of site-specific constants for each park and the PSC model includes a set of portfolio-specific constants for each portfolio (see equations (28), (31), and (44)). Table 6.1 provides a list of variables used in the models with definitions for each variable.

[6] I considered an alternative method for accounting for the effects of visiting sites other than the seven national parks. Using the survey responses, it is possible to compute the number of other recreational sites and cities visited, a number which ranges from 0-23. The number of secondary sites visited could be grouped into four categories: visit no other places, visit 1-5 secondary sites, visit 6-10 secondary sites, and visit more than 10 secondary sites. Consider now that each party would then be facing a choice of a set of national parks (seven of our interest) and the number of secondary sites to visit (one of the four categories). An individual's choice set is then expended to 508 choices (127 x 4). These expended choice set models and the original choice set models provide qualitatively and quantitatively similar results. Therefore, I decided to proceed with the original 127 choice set models.

Table 6.1: Variable Definitions

Variable	Definition
Travel Cost	See Chapter 3 Section 3.3.2 for a detailed discussion of travel cost (thousands of 2002 dollars)
Flextime	= 1 if visitors could have chosen a longer trip to the Four States Region
Car Renter	= 1 if respondent did not live in the four states region and took mass transportation to enter and exit the Four States Region (a potential car renter)
Visited Other Sites	=1 if visited other recreational sites or cities
Income	Annual household income (thousands of 2002 dollars)
Arches	= 1 If Arches is included in the portfolio
Bryce Canyon	= 1 If Bryce Canyon is included in the portfolio
Canyonlands	= 1 If Canyonlands is included in the portfolio
Grand Canyon	= 1 If Grand Canyon is included in the portfolio
Mesa Verde	= 1 If Mesa Verde is included in the portfolio
Petrified Forest	= 1 If Petrified Forest is included in the portfolio
Zion	= 1 If Zion is included in the portfolio
PSC 1 ~ 127	Portfolio specific constant for each portfolio

6.1.1 ASU – Standard Logit Model

Table 6.2 presents the estimation results for the Standard Logit Additive Site Utility model. Most of the parameter estimates are significant with the expected signs. The significant negative coefficient of travel cost suggests that the probability of a party choosing a portfolio decreases when the associated trip expense is high. The interactive terms of travel cost with demographic variables further examine the travel cost effects, accounting for preference heterogeneity across visitors. These coefficient estimates also have the expected signs. Visitors with higher income or who also visit recreational sites and cities other than the seven national parks tend to be less sensitive to travel costs. In other words, the effect of travel cost on the probability of choosing a portfolio decreases for higher income groups or for those who visit secondary sites. Visitors who potentially rented cars during their visit are more sensitive to travel costs. As explained in more detail in Chapter 3, travel cost includes transit cost, but not any car rental cost. Travel cost increases with the number of parks visited and number of days spent on the trip. For car renters, the increase in total cost would be even higher with an increase of days or parks visited, due to the extra (unaccounted for) rental costs. However, the size of the car renter and income interactive terms are small relative to the absolute size of the travel cost coefficient, suggesting that the magnitude of these effects are not large. Whether visitors have more flexible time (i.e., could have taken a longer trip) does not have a significant effect on their sensitivity to travel cost.

The site-specific variables are all significant and the relative size of the coefficient estimates follows the order of observed visitation counts. Portfolios that contain Grand Canyon have relatively higher probabilities of being selected, followed by portfolios containing Zion. Canyonland is the least popular site with the lowest coefficient. Unexpectedly, some parks have negative signs on their site-specific parameters. As discussed in Chapter 4, each site in the portfolio is expected to contribute to the overall utility of the portfolio; relatively unpopular sites were expected to have small coefficients, not negative coefficients. Although travel costs increase as more sites are added, visits to national parks should generate utility that serves to offset these money and time costs. If certain site parameters are negative, the combination of these sites with any other sites causes a lower utility. For instance, Grand Canyon by itself ranks higher than the combination of Grand Canyon and Arches, and this (smaller) combination ranks higher than the grouping of Grand Canyon, Arches and Bryce Canyon. This could be due to the fact that the dominant type of trips observed are single park trips (62%), or it could be caused by certain substitute or complement effects among parks that are not well captured in the standard additive site utility model.

Table 6.2: Standard Logit Additive Site Utilities Model (SL Model)

Variable	Coefficients	z-statistics
Travel Cost (in $1000)	-6.493	-16.86
Cost * Flextime	0.045	0.34
Cost * Car Renter	-0.326	-1.88
Cost * Income (in $1000)	0.014	7.56
Cost * Visited Other Sites	2.622	10.44
Arches	-0.622	-6.72
Bryce Canyon	-0.330	-3.68
Canyonlands	-1.336	-13.64
Grand Canyon	1.953	15.79
Mesa Verde	-0.613	-6.24
Petrified Forest	-1.246	-16.04
Zion	0.156	1.8
Log-likelihood	-8237.0122	
Sample size	2719	

6.1.2 ASU – Mixed Logit Model

As explained in Chapter 4, the mixed logit model has the same specification as the standard logit model, except that the site-specific parameters are considered to be random with certain distributions. Random parameters can not only account for preference heterogeneity across visitors, but, more importantly, also accommodate correlations across alternatives. When site-parameters are treated as random variables, the mixed logit models account for correlation among portfolios that share the same national parks.

To determine if mixing is necessary, or in other words, if there is correlation among portfolios, I perform a Lagrange Multiplier (LM) test. The LM test for this purpose was first proposed by McFadden and Train (2000) and later summarized by Brownstone (2001). To perform this LM test, one needs to first construct a set of artificial variables (z_{ik}) for the variables that are assumed to have random coefficients. z_{ik} is constructed using the following formula:

$$z_{ik} = (x_{ik} - \overline{x_i})^2, \text{ with } \overline{x_i} = \sum_{k=1}^{127} x_{ik} P_{ik} \qquad (60)$$

where x_{ik} is a vector of variables that have random parameters relating to individual i and alternative k (in this study, site-specific variables vary only across portfolios and not individuals) and P_{ik} is the choice probability of the conditional logit model. One then re-estimates the conditional logit model with the set of artificial variables z_{ik}. The null hypothesis of no mixing of the variable is rejected if z_{ik} is significant.

After performing this LM test, I found that all *z* variables are highly significant, suggesting that all seven site-specific variables should have random parameters. I also conducted a Likelihood Ratio test to test the joint significance of the *z* variables, and the result suggests the *z* variables are jointly significant. Tables 6.3 and 6.4 present results for these tests.

Table 6.3: Standard (Conditional) Logit Model with Z variables

Variable	Coefficient	z-statistics
Travel Cost (in $1000)	-4.524	-8.47
Cost * Flextime	0.063	0.5
Cost * Car Renter	-0.209	-1.24
Cost * Income (in $1000)	0.010	5.25
Cost * Visited Other Sites	2.135	7.58
Arches	2.448	7.7
Bryce Canyon	0.910	3.24
Canyonlands	6.909	7.03
Grand Canyon	1.366	6.17
Mesa Verde	2.463	6.38
Petrified Forest	2.437	3.15
Zion	0.479	2.71
Z - Arches	-5.596	-10.12
Z - Bryce Canyon	-3.178	-5.73
Z - Canyonlands	-10.699	-8.57
Z - Grand Canyon	0.917	3.06
Z - Mesa Verde	-5.428	-8.93
Z - Petrified Forest	-5.541	-5.18
Z - Zion	-2.336	-4.44
Log-likelihood	-8073.513	
Sample size	2719	

Table 6.4: Likelihood Ratio Test

Model	df	LL(null)	LL(model)	AIC	BIC
SL	12	-13181.74	-8237.012	16498.02	16627.05
SL with Zs	19	-13181.74	-8073.513	16185.03	16389.32
Likelihood-ratio test			LR chi2(7) = 327.00 (Prob > chi2 = 0.0000)		

The next step for estimating the mixed logit model is to specify the distribution of the random parameters. As discussed in Chapter 4, random parameters can take a number of predefined functional forms, such as the normal, lognormal, triangular, and uniform distributions. Since none of the random parameters in this model are expected to have a specific sign, the lognormal distribution would not be appropriate. I thus restricted my tests to the normal, triangular, and uniform distributions.

The normal distribution is most commonly chosen distribution for random parameters without expected signs. As the result will not necessarily be independent of the number of random draws in the simulation, I estimated the model using different number of Halton draws, ranging from 100 to 1000. The differences between the parameter estimates became quantitatively smaller as the number of draws increased. Between 750 and 1000 draws, the difference was almost negligible. The result of the mixed logit model with normal distribution and 1000 Halton draws is presented in table 6.5, column 1. Travel cost is again highly significant with a negative sign. All of the interaction terms also have the same signs as the standard logit model. The only thing that changes is the order of the site-specific dummies and their significance. Grand Canyon is still the site with the highest “utility hit”. Petrified Forest, however, takes the place of Canyonlands as the least popular site, with the least (or most negative) “utility hit”. Among the mixing parameters, only Grand Canyon and Petrified Forest have significant standard deviations, suggesting that only portfolios that both have Grand Canyon or both have Petrified Forest are correlated. The weakness of the normal distribution is the unbounded tails. This creates the possibility of behaviorally unacceptable draws for the coefficient from the tails. Note that the standard deviation of Grand Canyon is almost twice the size of its mean. This creates even longer tails and causes the sign of parameter estimates of Grand Canyon to change frequently with different draws. As a result, the welfare estimates for Grand Canyon with different numbers of random draws vary more compared to other park welfare estimates.

One of the commonly used method to constrain the draws to more reasonable and behaviorally acceptable ranges is to use the triangular distribution, where both ends of the distribution are bounded. With mean b and spread s, the distribution is bounded within the range of $[b - s, b + s]$ and reaches its peak of $1/s$ at b. Random draws from this distribution can be created as $\beta = b + s\left(\sqrt{2\mu} - 1\right)$ if $\mu < 0.5$ and $\beta = b + s\left(1 - \sqrt{2(1 - \mu)}\right)$ if $\mu > 0.5$, where μ is a random draw from a standard uniform distribution. However, when applying the triangular distribution to all seven site-specific dummies, the model fails to converge.

The uniform distribution is another way to bind the upper and lower bound of the distribution and is often used when the variable is a dummy variable. Since the seven site-specific variables are all dummies, I also tested the uniform distribution and found that the parameter estimates are very similar to the estimates of normal distribution. Although the ends of the distribution are bounded, the spread of the Grand Canyon estimates becomes even bigger (a standard deviation almost three times the mean).

To further constrain the ranges of the coefficients, one may use truncated or constrained distributions (Hensher and Greene, 2003). Truncated distributions restrict the standard deviation or spread to be a function of the mean. The constraint specification can be applied to any distribution. For example, a triangular distribution specified as $\beta_i = b + sv_i$, where v_i is the random variable, can be constrained by setting $s = zb$, where z is a coefficient of variation taking any positive value. The distribution then becomes $\beta_i = b + zbv_i$. z is generally expected to lie between 0 and 1. One commonly used z value is 1, which set the standard deviation or spread equal to the mean. With a truncated triangular distribution, constraining $b = s$ binds the parameter estimates to be consistent with the same sign. The range is [0, $2b$] if b is positive, and [2b, 0] if b is negative.

In the mixed logit model, Grand Canyon is the only site-specific dummy with a standard deviation significantly larger than its mean. Since every random parameter may have its own distribution, I choose to constraint only the distribution of the Grand Canyon coefficient, while allowing the other sites to have normal distributions. With a truncated triangular distribution with $b = s$, the sign of the Grand Canyon coefficient is constrained to be either always positive or negative. The estimation result is presented in table 6.5, column 2. The parameter estimates become more similar to the standard logit model.

I also tested the extreme case of constraining the Grand Canyon coefficient by fixing it. Column 3 of table 6.5 presents the estimates of the model where Grand Canyon is treated as having a fixed parameter while others are random and follow normal distributions. The estimation results turn out to be almost identical to the standard logit model, with all standard deviations being insignificant. This could be due to the fact that a large percentage of our population (63%) visited Grand Canyon and 62% of our population visited only one national park during their trip.

Table 6.5: Mixed Logit Additive Site Utilities Model (MXL Model)[1]

Variable	All Random (Normal)[2]		All Random (Uniform)[3]		All Random (Normal + Triangular)[4]		Grand Canyon Fixed	
	Coef.	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**
Travel Cost (in $1000)	-9.258	-16.57	-9.255	-23.02	-7.326	-22.03	-6.493	-16.86
Cost * Flextime	0.246	1.29	0.244	1.88	0.073	0.67	0.045	0.34
Cost * Car Renter	-0.419	-1.67	-0.417	-2.49	-0.332	-2.3	-0.326	-1.88
Cost * Income (in $1000)	0.024	8.96	0.024	13.17	0.016	10.78	0.014	7.56
Cost * Visited Other Sites	3.266	8.11	3.283	11.57	2.809	13.15	2.622	10.44
Grand Canyon	6.542	2.87	6.530	3.78	2.373	18.85	1.953	15.79
Arches	-0.155	-1.45	-0.161	-1.84	-0.442	-5.28	-0.622	-6.72
Bryce Canyon	0.144	1.32	0.131	1.24	-0.154	-1.95	-0.33	-3.68
Canyonlands	-0.843	-7.58	-0.851	-7.01	-1.148	-11.68	-1.336	-13.64
Mesa Verde	-0.087	-0.76	-0.093	-1.01	-0.419	-4.81	-0.613	-6.24
Petrified Forest	-2.755	-3.29	-4.682	-2.37	-1.096	-14.17	-1.246	-16.04
Zion	0.644	5.92	0.639	7.72	0.336	4.35	0.155	1.8
SD								
Grand Canyon	12.94	1.9	16.715	2.6	2.373	18.85	-	-
Arches	0.001	0.1	0.008	0.02	0.002	0.01	0.002	0.9
Bryce Canyon	0.076	0.7	0.346	0.28	0.001	0	0.005	0.75
Canyonlands	0.04	0.25	0.149	0.08	0.009	0.02	0.001	0.13

Table 6.5 Continued

Mesa Verde	0.003	0.42	0.044	0.11	0.006	0.02	0.002	0.28
Petrified Forest	2.967	3.57	7.689	2.75	0.014	0.03	0.017	1.14
Zion	0.015	1.21	0.007	0.02	0.005	0.02	0.016	0.91
Log-likelihood	-8151.35		-8150.20		-8217.14		-8237.01	
Sample size	2719		2719		2719		2719	

Note: [1.] All mixed logit models are estimated using 1000 halton draws. [2.] All random parameters assumed to follow normal distribution [3.] In this mixed logit model, the random parameter of Grand Canyon is assumed to follow triangular distribution with mean equals to its spread, while the random parameters of other parks are assumed to follow normal distribution.

6.1.3 Portfolio Specific Constant Model

As explained in Chapter 4, the Portfolio Specific Constant Model includes a set of portfolio specific constants instead of site-specific dummies. Compared to the Additive Site Utility models, the Portfolio Specific Constant model is relatively less constrained. For the ASU models each individual site is considered to contribute independently to a person's overall trip utility. In other words, the effect of any single park on utility is constant, regardless of the presence of any other parks in the portfolio. For instance, the contribution of visiting Grand Canyon during the trip is identical, regardless of whether one also visited Arches, Zion, or any other parks. However, this would not be true if the parks included in the portfolio are either substitutes or complements to one another. With the PSC model, each portfolio is represented by its own constant term, which implicitly allows for interactive effects between parks. Consider the following example; if both Zion and Arches provide similar "services" (e.g., both offer opportunities to view wildlife or hike canyons), then one park would easily be viewed as the substitute of the other. In that case, having both parks in the portfolio should generate less overall utility than the addition of the two individual "utility hits." Therefore, if parks in a portfolio are substitute for each other, one should expect the PSC constant to be lower than the sum of same utilities from the ASU Model. If the parks have features that complement each other, the PSC constant should be higher than the sum. If the effect is negligible, the sum from the ASU Model should be close to the corresponding constant from the PSC Model.

The estimation results are presented in table 6.6. To keep the table manageable, table 6.6 only lists the range of the estimates of the 111 observed portfolios' specific constants[7]. Note that the travel cost coefficient (-4.985) is relatively lower (in absolute value) compared to the one in the ASU-Standard Logit Model (-6.493). This could be due to the fact that, with the PSC Model, the portfolio specific constants are able to pick up the complementary effects among parks that are located close to one another. The ASU models' site-specific dummies cannot capture such effects, which thus may well be picked up in (or be contaminating) the travel cost coefficient for the ASU Model. For example, Bryce Canyon and Zion are located fairly close to one another and the portfolio {BC, ZI} is one of the top chosen portfolios. The closeness between the two parks causes the travel cost of the portfolio to be relatively lower compared to other combinations of two parks (all other factors held constant). Since ASU site dummies do not account for the complementary effects between the two sites, the model attempts to use travel costs to explain why portfolio {BC, ZI} is more popular, which biases the travel cost coefficient upwards. Another way of thinking about this issue is that there are synergies between nearby parks which are masked by the imposed additivity in the ASU Model. Finally, the demographic interaction terms are largely insensitive to whether the specification is PSC or ASU. The signs and significance of those coefficients in the PSC Model are similar to the ones in the ASU Model.

---

[7] As shown in table 3.4, only 111 out of 127 portfolios were chosen by respondents. Therefore, only 111 portfolio specific constants are included in the model for estimation.

Table 6.6: 111 Portfolio Specific Constant Model (PSC Model)

Variable	Coefficients	z-statistics
Travel Cost (in $1000)	-4.985	-13.15
Cost * Flextime	0.050	0.56
Cost * Car Renter	-0.227	-1.83
Cost * Income (in $1000)	0.010	7.52
Cost * Visited Other Sites	2.235	8.32
PSC 2 ~ 127 (PSC1 as the baseline)	-17.449~ 2.850	-
Log-likelihood	-7152.5593	
Sample size	2719	

[*] The range of coefficient estimates of the 111 portfolio dummies, with PSC4 being the highest and PSC16 being the lowest. See Appendix C Table C18 for the full set of PSC estimates.

Table 6.7 lists the top 20 PSC constant estimates. The additive site utility model suggests that in many cases combinations of several parks rank lower than single park portfolios (due to the negative coefficient estimates of several parks). However, with the PSC Model, the top ranking portfolio constant estimates are in fact mostly portfolios which contain more than one park. This could be taken as more evidence that the PSC Model does a better job catching complementary effects between parks. One consistency between the ASU and PSC models is that visiting Grand Canyon alone (PSC4) ranks the highest among all portfolios.

To further explore the correlation between parks, I estimated the second stage regression of the fitted PSC's on a model with dummies for the included sites and pairwise interactions between parks. These results are presented in table 6.8. Among the 21 pairwise dummies, 5 pairs are significant with positive signs, indicating that those pairs of parks are positively correlated (i.e., serve as complements to one another). They are {Bryce Canyon, Zion}, {Arches, Canyonlands}, {Bryce Canyon, Canyonlands}, {Grand Canyon, Petrified Forest}, and {Mesa Verde, Petrified Forest}. If one checks the map of the Four States Region (Figure 3.1), all of these pairs appear to be the ones which are geographically located close to each other. Visiting a combination of parks near each other allows a household to spend more time onsite at each park instead of traveling between parks and as a result further boost visitors' utility.

Table 6.7: Top 20 Portfolio Specific Constant Estimates

Portfolio Specific Constant	Coefficient (Compare to PSC1)	# of Parks in the Portfolio
PSC4	2.792	1
PSC121	2.357	6
PSC50	2.023	3
PSC101	1.935	5
PSC122	1.877	6
PSC127	1.836	7
PSC107	1.417	5
PSC24	1.301	2
PSC92	1.216	4
PSC18	1.170	2
PSC7	0.932	1
PSC91	0.878	4
PSC70	0.842	4
PSC124	0.774	6
PSC25	0.715	2
PSC118	0.628	5
PSC5	0.572	1
PSC60	0.503	3
PSC106	0.477	5
PSC67	0.394	4

Table 6.8: Second Stage OLS Regression of the PSC Model[1]

	Site-Specific and Pairwise Dummies	
	Coefficient	**t-statistics**
Constant	0.700	0.5
Arches	-0.103	-0.08
Bryce Canyon	-2.878	-2.25
Canyonlands	-3.791	-2.95
Grand Canyon	1.059	0.85
Mesa Verde	-3.172	-2.47
Petrified Forest	-2.133	-1.66
Zion	-0.475	-0.38
Pairwise Dummies		
AR - BC	0.453	0.47
AR - CA	2.746	2.76
AR - GC	-1.138	-1.14
AR - MV	1.605	1.61
AR - PF	-0.194	-0.2
AR - ZI	-0.523	-0.53
BC - CA	2.301	2.37
BC - GC	0.795	0.82
BC - MV	-0.386	-0.4
BC - PF	0.746	0.77
BC - ZI	3.228	3.33
CA - GC	-0.207	-0.21
CA - MV	0.394	0.4
CA - PF	-0.092	-0.09
CA - ZI	-0.016	-0.02
GC - MV	0.247	0.25
GC - PF	1.939	1.98
GC - ZI	0.252	0.26
MV - PF	1.740	1.76
MV - ZI	1.395	1.43
PF - ZI	-1.146	-1.18

[1.] 111 (the chosen portfolios) PSC estimates are used to fit the second stage regression.

6.2 Welfare Analysis

The ultimate purpose behind estimating all of the above models is to provide a means for estimating the welfare losses associated with temporary closure of the national parks in this study. Note that any welfare loss estimate obtained from these data will be an underestimate of true losses, due to the population from which these losses are extrapolated. As this population only includes people who actually take trips to at least one of the seven national parks, there is no information on parties that were considering such a trip and changed their trip plans because they learned of the closure beforehand.

In this section, I examine the welfare losses due to park closures using the estimation results from the models above. The first part of the welfare analysis concentrates on two types of scenarios: individual park closures and groups of parks closures. I estimate per-party, per-adult, and per-person welfare losses due to park closures. The second part of the analysis focuses on estimating the loss-to-trip ratio for individual park closures. Finally, I estimate the aggregated values of park closures using park data on visitation rates.

6.2.1 Per-trip Welfare Loss for Park Closures

In the RUM setting, welfare changes can be calculated using the indirect utility function. The indirect utility function is the maximized value of the utility function. Assuming the error term follows the IID type 1 extreme value distribution, the expected maximum utility can be expressed as:

$$E_i(\max_k(U_{ik})) = E_i(\max_k(V_{ik} + \varepsilon_{ik})) = \ln(\textstyle\sum_{k=1}^{127} \exp(V_{ik})) + C, \qquad (60)$$

where C is the Euler's constant[8] from solving with the extreme value distributional assumption for the error term and can be ignored when measuring changes of utilities. The per-trip welfare loss can be calculated by assessing the change in utility (consumer surplus) that would occur if all feasible portfolios containing the closed parks were eliminated from the choice set. For instance, if Arches is closed then all portfolios containing Arches become unavailable. If more than one park is closed, portfolios containing any of the closed parks will be excluded from the choice set. To convert the utility to dollar terms (assuming that utility is linear in income), simply divide the difference in expected maximum utility by the marginal utility of income, which is the travel cost related coefficients in this model. This is a version of the well-known log-sum-difference formula. For the standard logit model, the log-sum differences per party for the loss of one/multiple parks is:

$$EV_i^{SL} = \frac{\{ln \sum_{k \in A_{-n}} \exp(\sum_{m=1}^{7} \beta_m x_{km} + \gamma \cdot p_{ik} \cdot z_i') - ln \sum_{k \in A} \exp(\sum_{m=1}^{7} \beta_m x_{km} + \gamma \cdot p_{ik} \cdot z_i')\}}{-\gamma \cdot z_i'}, \quad (62)$$

where A is the full set of 127 portfolios, A_{-n} is the set of portfolios excluding the closed n parks, γ is the travel cost coefficient, and z_i' is a vector of 1 and individual characteristics. For the PSC model the log-sum differences per party for the loss of one/multiple parks is:

[8] A proof of equation (61) can be found in Haab and McConnell (2002).

$$EV_i^{PSC} = \frac{\{ln \sum_{k \in A_{-n}} exp(\alpha_k + \gamma \cdot p_{ik} \cdot z_i') - ln \sum_{k \in A} exp(\alpha_k + \gamma \cdot p_{ik} \cdot z_i')\}}{-\gamma \cdot z_i'}. \tag{63}$$

In the mixed logit model, all β_ms become variable across the population. Therefore, the log-sum term becomes:

$$\int \ln(\sum_k exp(\sum_{m=1}^{7} \beta_{im} x_{km} + \gamma \cdot p_{ik} \cdot z_i')) f(\beta|b, W) d\beta \tag{64}$$

where $f(\beta|b, W)$ is the density function of the random parameter β_{im}. The welfare loss estimated based on the mixed logit model can be expressed as:

$$EV_i^{MXL} = \frac{\{\int \ln(\sum_{k \in A_{-n}} exp(\beta_{i \cdot} x_{k \cdot} + \gamma \cdot p_{ik} \cdot z_i')) f(\beta|b,W) d\beta - \int \ln(\sum_{k \in A} exp(\beta_{i \cdot} x_{k \cdot} + \gamma \cdot p_{ik} \cdot z_i')) f(\beta|b,W) d\beta\}}{-\gamma \cdot z_i'}, \tag{65}$$

where $\beta_{i\cdot}$ and $x_{k\cdot}$ are vectors of β_{im} and x_{km} for $m = 1, \ldots 7$ respectively. Since there is no closed form for the equation above, welfare losses are usually computed as the average of the monetized log-sum differences over all sampled individuals over certain numbers of random draws. Equation (65) then becomes:

$$EV_i^{MXL} = \frac{1}{R} \sum_{r=1}^{R} \frac{\{ln \sum_{k \in A_{-n}} exp(\beta_{i\cdot}^r x_{k\cdot} + \gamma \cdot p_{ik} \cdot z_i') - ln \sum_{k \in A} exp(\beta_{i\cdot}^r x_{k\cdot} + \gamma \cdot p_{ik} \cdot z_i')\}}{-\gamma \cdot z_i'}, \tag{66}$$

where R is the number of random draws and $\beta_{i\cdot}^r$ is the vector of coefficients from the r^{th} draw.

For all three equations, the numerator reflects the difference between the log sum over all portfolios except those including the park(s) being valued for loss and the log sum over all 127 portfolios. These values are all per trip welfare losses for each individual and conditional on the person making a trip to the four states region.

Table 6.9 presents welfare loss estimates based on different models. It shows both individual park losses and groups of parks losses. For groups of parks closures, I picked three portfolios which contained popular groups of parks.

The values are also reported in per-trip per-party, per-trip per-adult, and per-trip per-person formats. The per-trip per-party values are calculated by simply using the sum of the individual per trip value divided by number of parties. The per-trip per-adult and per-trip per-person values divide the values in equations (62), (63), and (66) by the number of adults/people (adults and children) in the party and are computed by enumerating over the sample. Having the values in different units is useful when transferring values to other parks, where aggregate visitation data may count all people, all adults, or all parties. Since the number of adults and children varies across parties, these averages are not simple transformations of each other. The per-adult values are about half as large as the per-party values, since the average number of adults per-party is near two, and the per-person values are about one third of the value of the per-party values, since the average number of people including children is about three.

The PSC welfare estimates are larger than the SL estimates for every single park closure and group park closures. This result is primarily driven by the small absolute value of travel cost in the PSC Model. The values from the PSC Model range from a low of $21 per party for Canyonlands to $217 for Grand Canyon. Zion, as expected, has the second largest value at \$88. The group loss for closing the group of three parks – Grand Canyon, Zion and Bryce Canyon, runs the largest (\$408). The estimates from the SL Model follow the same order as the ones from the PSC Model.

For the MXL Model, I choose to use the estimates from the All Random (Normal + Triangular) model, where all site-specific parameters are random and Grand Canyon by itself has a different distribution (triangular) from the others (normal). The first two MXL Models in table 6.5 both have standard deviations for Grand Canyon which are significantly larger than their means, causing the unstable estimates of welfare loss for Grand Canyon with different number of random draws. The last MXL model, which constrains Grand Canyon to have a fixed parameter, gives almost exactly the same estimates as the SL Model.

Table 6.9: Per-trip Welfare Loss for Park Closures (2002$)

Single Park Closures	Per-Party Welfare Loss			Per-Adult Welfare Loss			Per-Person Welfare Loss		
	SL Model	PSC Model	MXL Model[1]	SL Model	PSC Model	MXL Model[1]	SL Model	PSC Model	MXL Model[1]
Arches	$29	$42	$26	$15	$22	$14	$12	$17	$11
Bryce Canyon	40	60	35	21	32	18	16	25	14
Canyonlands	14	21	12	7	11	6	6	9	5
Grand Canyon	159	217	161	84	113	85	65	88	66
Mesa Verde	27	39	24	14	20	12	11	15	10
Petrified Forest	20	31	18	10	16	9	8	13	7
Zion	59	88	52	31	46	27	24	36	21
Multiple Parks Closures[2]									
Group I - Bryce Canyon, Grand Canyon and Zion	$295 (258)[2]	$408 (365)	$279 (248)	$155 (136)	$211 (191)	$146 (131)	$120 (106)	$164 (149)	$114 (102)
Group II- Grand Canyon and Petrified Forest	185 (179)	235 (248)	184 (179)	97 (94)	122 (129)	97 (94)	76 (73)	95 (101)	75 (73)
Group III - Arches and Canyonlands	43 (43)	48 (63)	38 (38)	22 (22)	25 (33)	20 (20)	17 (17)	20 (26)	15 (16)

[1.] Calculated using the parameter estimates from the All Random (Normal + Triangular) Model in Table 6.5 with 5000 random draws. [2.] Values in the parenthesis are the sum of corresponding individual park losses.

6.2.2 Loss-to-trip Ratio and Aggregated Welfare Loss for Park Closures

In interpreting the values in table 6.9, it is important to keep in mind that these are "per trip to the four states region," whether the trip destination includes the lost park or not. In aggregating these values to total annual losses over all users, the number of parties traveling to all seven parks should be multiplied by the per party value. Another welfare measure that is commonly used is the loss-to-trip ratio. These are "per trip to a specific park." In this case, aggregating the per-trip value to total annual losses over all users is accomplished by multiplying by the total number of parties traveling to the park of interest. This is often used in natural resource damage assessment where one knows the total number of trips lost to a specific park or parks and seeks the per trip value for that park(s). The loss-to-trip ratio is calculated as:

$$lttr_m = \sum_{i=1}^{2719} EV_i \,/\, \sum_{i=1}^{2719} \lambda_m, \tag{67}$$

where $\sum_{i=1}^{2719} EV_i$ is the total welfare loss during the sampling period that resulted from single/multiple park closures and $\sum_{i=1}^{2719} \lambda_m$ is the weighted total number of trips taken to site m (in this case it's also the number of parties/adults/people that visited site m during the sampling period). Table 6.10 shows the loss-to-trip values per-party/per-adult/per-person. All parks have similar loss-to-trip values, expect for Grand Canyon.

Table 6.10: Loss-to-trips Ratio for Individual Park Closures (2002$)

Single Park Closures	Per-Party Loss-to-trips Ratio			Per-Adult Loss-to-trips Ratio			Per-Person Loss-to-trips Ratio		
	SL Model	PSC Model	MXL Model[1]	SL Model	PSC Model	MXL Model[1]	SL Model	PSC Model	MXL Model[1]
Arches	$185	$265	$165	$83	$119	$74	$62	$89	$56
Bryce Canyon	183	278	162	80	122	71	58	88	52
Canyonlands	172	259	153	79	119	70	62	93	55
Grand Canyon	252	345	255	111	152	113	80	110	81
Mesa Verde	176	250	156	80	114	71	56	80	50
Petrified Forest	161	248	143	71	109	63	50	77	44
Zion	189	280	167	85	125	75	61	90	54

[1.] Calculated using the parameter estimates from the All Random (Normal + Triangular) Model in Table 6.5 with 5000 random draws.

The last welfare measure considered in this study is the aggregated welfare loss for individual park closures over the season (June 2002). It can be calculated using park data on visitation rates:

$$AgWL_m = lttr_{m,per-person} * N_m, \tag{67}$$

where N_m is the total visitors to park m during the month of June 2002. Visitation data is obtained from National Park Service Use Statistics[9]. Table 6.11 presents aggregated welfare loss for each park based on the different model estimates. These values range from $4.1 million for Canyonlands to $55.2 million for Grand Canyon using the PSC Model estimates.

---

[9] National Park Service Visitor Use Statistics. Recreation Visitors by Month by Parks. https://irma.nps.gov/Stats/SSRSReports/Park%20Specific%20Reports/Recreation%20Visitors%20By%20Month%20(1979%20-%20Last%20Calendar%20Year). A detailed explanation of visitor use counting procedures is also available at the website of National Park Service Visitor Use Statistics

Table 6.11: Aggregated Welfare Loss for Individual Park Closures and Total Visitors by Park

Single Park Closures	Aggregate Loss in June 2002 (Millions of 2002$)			Total Visitors in June 2002 (Thousands)
	SL Model	PSC Model	MXL Model[1]	
Arches	$6.3	$9.0	$5.6	101.1
Bryce Canyon	7.5	11.4	6.7	129.2
Canyonlands	2.7	4.1	2.4	44.0
Grand Canyon	40.4	55.3	40.9	502.2
Mesa Verde	4.5	6.4	4.0	80.2
Petrified Forest	4.2	6.5	3.7	84.3
Zion	20.0	29.5	17.6	329.4

[1.] Calculated using the parameter estimates from the Grand Canyon Random Model in Table 12 with 1000 random draws.

Chapter 7

CONCLUSION

The main objective of this dissertation is to provide a new methodology for addressing the multiple-site visitation issue in the Travel Cost Model. Since the inception of the TCM, multiple-site visitations have been an issue that has been neglected by most researchers. As Myrick Freeman mentioned in his book:

> "In implementing the CK technique[10], it must be assumed that the primary purpose of the recreation trip is to visit that site. When trips involve purposes other than visiting the site[11], at least some portion of the total travel cost is a joint cost which cannot be allocated meaningfully to the visit." A. Myrick Freeman III (1979), pp. 202

Over the years, several approaches have been proposed to address the issue of allocating travel costs when multiple-site visitations are involved; however, these approaches have proven to be problematic for generalized application. Rather than attempting to divide total travel cost among sites, the portfolio-based strategy I propose instead considers bundling the sites and treating each bundle/portfolio as a single choice. Compared to other approaches this approach makes the model relatively more applicable. In cases where the trips are dominantly single-site visits, such as day trips for fishing or beach recreation, conventional travel cost models continue to be a

---

[10] The "CK technique" refers to the Clawson-Knetsch travel cost method of demand estimation.

[11] Visiting two or more sites or to visit a relative en route.

valid approach. However, the portfolio-based approach can be applied to a much broader swathe of potential trips. For example, this approach can estimate costs for day trips for bird-watching (where viewers often move from one viewing site to another during the day), overnight trips where multiple recreation locations are visited, or trips to national parks in other countries where parks also cluster in certain regions. Although the data collection necessary for this type of application is relatively more time and labor consuming (as it usually needs to be conducted on site) and the weighting of the data to correct for sampling presents a non-trivial complication relative to conventional travel cost modeling approaches, as this dissertation has demonstrated both issues can be overcome using conventional surveys and econometric methods.

This dissertation also provides estimates for the welfare losses that park users would incur in the event of a short-term closure of one or more national parks in the four states region. This portfolio-based approach is conducted in a utility-theoretical framework capable of generating per trip measures of value for the closure of individual sites or group of sites. The per party per trip welfare losses for closing individual parks range from $12 for the least popular park - Canyonlands to $161 for the most popular park – Grand Canyon (in 2002 dollars). The estimated per party loss-to-trip ratio of individual park closures ranges from $143 to $255 (in 2002 dollars). These results provide useful information to assist in the assessment of current management and policy actions regarding national park closures due to natural disasters or environment hazards, such as wildfires, avalanches, oil spills, health and safety issues raised by abandoned mine lands, etc. In many environmental hazards related cases, instead of shutting down the entire park only portions of the national parks are closed. More research is necessary to determine the welfare losses from these partial closures. One simple solution is to adjust the full welfare losses using the trip cancelation rates due to partial park closures.

There are a few potential improvements which can be made in future studies. First, the per-trip value estimated in this study is confined to short-term impacts. Given the data collection method, all participants in this study are individuals who actually traveled to one or more of the national parks of interest. Thus, the data exclude any individuals who find out about a park closure in advance and cancel their entire trips. For future studies, combining an onsite portfolio choice survey with an offsite national survey focusing on individuals' participation decision to get the rates of use can successfully incorporate these individuals' participation decisions into the model and thus obtain per-trip values that are no longer confined to "short term" impacts. Another option is to include stated preference (SP) questions in the survey to collect information on how visitors adjust their trips if they become aware of the park closures before starting the trip. Second, the portfolio model presented in this dissertation only included national park dummies. Future studies can incorporate an additional set of site characteristics to allow for values for characteristics – such as environmental quality (water, land cover, etc.), the presence of wildlife, the number of trails, and other amenities. This empirical improvement could be easily incorporated within the realm of random utility theory and feasible based on practical data collection.

REFERENCES

Adler, Thomas & Ben-Akiva, M. (1979). A Theoretical and Empirical Model of Trip Chaining Behavior. *Transportation Research Part B: Methodological*, 13 (3), 243-257.

Atherton, T., Ben-Akiva, M., & McFadden, D. (1990). Micro-simulation of local residential telephone demand under alternative service options and rate structures. *In Telecommunications Demand Modeling: An Integrated View (Fontenay A., Shugad M. and Sibley D., Eds)* North Holland, New York, pp. 137-163.

Bockstael, N. E., Hanemann, W. M. & Strand, I. E. (1984). Measuring the benefits of water quality improvements using recreation demand models. Report presented to the U.S. Environmental Protection Agency. College Park, MD: University of Maryland.

Bockstael, N., & McConnell, K. (2007). *Environmental and resource valuation with revealed preferences : A theoretical guide to empirical models*, Springer.

Brown, W. G., & Nawas, F. (1973). Impact of aggregation on the estimation of outdoor recreation demand functions. *American Journal of Agricultural Economics*, 55(2), 246-249.

Brownstone, D.,. (2001). Discrete Choice Modeling for Transportation. *Travel Behavior Research : The Leading Edge.*, 97-124.

Carson, R. T., Hanemann, W. M., & Wegge, T. C. (2009). A nested logit model of recreational fishing demand in Alaska. *Marine Resource Economics, 24*(2), 101-129.

Clawson, M., & Knetsch, J. L. (1966). *Economics of outdoor recreation.* Baltimore: Johns Hopkins University.

Englin, J., & Shonkwiler, J.S.,. (1995). Estimating social welfare using count data models: An application to long-run recreation demand under conditions of endogenous stratification and truncation. *Revieconstat the Review of Economics and Statistics, 77*(1), 104-112.

Freeman, M. A.,. (1993). *The measurement of environmental and resource values: Theory and methods.* Washington, D.C.: Resources for the Future.

Gourieroux, C., & Monfort, A. (1996). *Simulation-Based Econometric Methods*. Oxford; New York: Oxford University Press.

Greene, W. H.,. (2008). *Econometric Analysis*. Upper Saddle River, N.J.: Prentice Hall.

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*(1), 89-102.

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

Haspel, A. E., & Johnson, F. R. 1982. Multiple Destination Trip Bias in Recreation Benefit Estimation. *Land Economics, 58*(3), 364-372.

Hellerstein, D. M., & Mendelsohn, R. (1993). A theoretical foundation for count data models. *American Journal of Agricultural Economics, 75*(3), 604-611.

Hensher, D. A., & Greene, W. H. (2003). The Mixed Logit Model: The State of Practice. *Transportation: Planning - Policy - Research – Practice, 30*(2), 133-176.

Herriges, J. A., & Phaneuf, D. J. (2002). Inducing Patterns of Correlation and Substitution in Repeated Logit Models of Recreation Demand. *Amerjagriecon American Journal of Agricultural Economics,* 84 (4), 1076-1090.

Hotelling, H.,. (1949). *An economic study of the monetary valuation of recreation in the National Parks.* Washington, DC: U.S. Department of the Interior, National Park Service and Recreation Planning Division.

Knapman, B., & Stanley, O. (1991). *A Travel Cost Analysis of the Recreation use Value of Kakadu National Park.* Resource Assessment Commission Inquiry, AGPS, Canberra.

Kuosmanen, T., Nillesen, E., & Wesseler, J. (2004). Does Ignoring Multidestination Trips in the Travel Cost Method Cause a Systematic Bias? *Australian Journal of Agricultural and Resource Economics,* 48 (4), 629-651.

Loomis, J. B., Yorizane, S., & Larson, D. M. (2000). Testing Significance of Multi-Destination and Multi-Purpose Trip Effects in a Travel Cost Method Demand Model for Whale Watching Trips. *Agricultural and Resource Economics Review,* 29 (2), 183-191.

Luce, R. D.,. (1979). *Individual choice behavior : A theoretical analysis.* Westport, Conn.: Greenwood Press.

Manski, C. F., & Lerman, S. R. (1977). The Estimation of Choice Probabilities from Choice Based Samples. *Econometrica Econometrica*, 45 (8), 1977-1988.

Martínez-espiñeira, R. & Amoako-tuffour, J. (2009). Multi-Destination and Multi-Purpose Trip Effects in the Analysis of the Demand for Trips to a Remote Recreational Site. *Environmental Management, 43*(6), 1146-61.

McFadden, D.,. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (ed.), *Frontiers in econometrics* (pp. 105-142). Academic Press: New York.

McFadden, D., & Train, K. (2000). Mixed MNL Models for Discrete Response. *JAE Journal of Applied Econometrics, 15*(5), 447-470.

Mendelsohn, R., Hof, J., Peterson, G., & Johnson, R. (1992). Measuring Recreation Values with Multiple Destination Trips. *American Journal of Agricultural Economics, 74*(4), 926-933.

Newman, J., Ferguson, M., & Garrow, L. (2012). Estimating discrete choice models with incomplete data. *Transportation Research Record* 2302: 130-137

Parsons, G. R., & Wilson, A. J. (1997). Incidental and Joint Consumption in Recreation Demand. *Agricultural and Resource Economics Review, 26*(1), 1-6.

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*(1), 83-92.

Shaw, D.,. (1988). On-site samples' regression: Problems of non-negative integers, truncation, and endogenous stratification. *Journal of Econometrics, 37*(2), 211-223.

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

Tay, R., McCarthy, P. S., & Fletcher, J. J. (1996). A portfolio choice model of the demand for recreational trips. *Transportation Research Part B: Methodological, 30*(5), 325-337.

Train, K.,. (1998). Recreation demand models with taste differences over people. *Land Economics, 74*(2), 230-239.

Train, K.,. (2009). *Discrete Choice Methods with Simulation*. Cambridge; New York: Cambridge University Press.

Train, K. E., McFadden, D. L., & Ben-Akiva, M. (1987). The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *The Rand Journal of Economics, 18*(1), 109-123.

Trice, A. H., & Wood, S. E. (1958). Measurement of recreation benefits. *Land Economics, 34*(3), 195-207.

Warner, S. L. (1963). Multivariate regression of dummy variates under normality assumptions. *Journal of the American Statistical Association, 58*(304), 1054-1063.

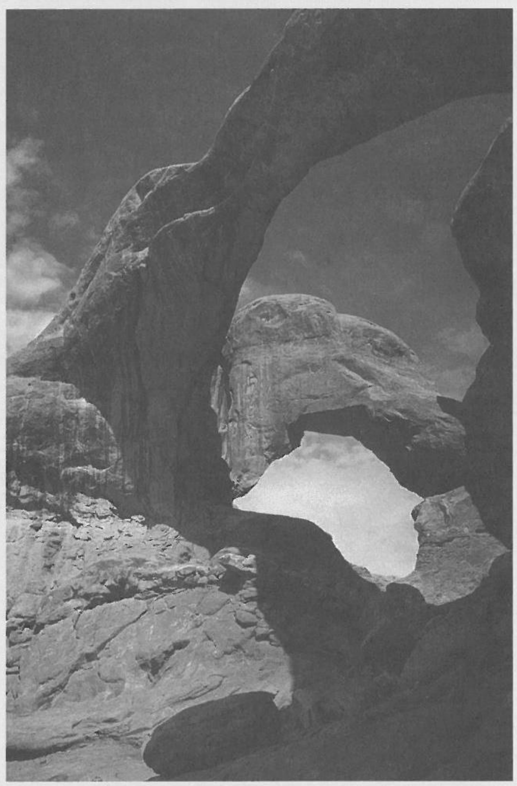
Yeh, C., Haab, T. C., & Sohngen, B. L. (2006). Modeling multiple-objective recreation trips with choices over trip duration and alternative sites. *Environmental and Resource Economics*, *34*(2), 189-209.

Appendix A

SURVEY QUESTIONNAIRE

SOUTHWEST NATIONAL PARK
VISITOR SURVEY

Tell us about your trip!



National Park Service Photo

PLEASE FOLD OUT FRONT COVER OF BOOKLET TO VIEW MAP OF STUDY REGION

Four States Region

SECTION A Arrival and Departure

This questionnaire is about your recent trip to visit national parks in Arizona, Colorado, New Mexico, and/or Utah (see fold-out map on the inside front cover). We refer to these four states in the questionnaire as the "Four States Region."

In the past few months you may have taken more than one trip to visit national parks in this region. When you answer the following questions we want you to focus only on the trip you were taking when we approached you in June about taking this questionnaire.

First, we need to have information about when you entered and left the Four States Region during this trip and by what means of travel.

1 On the calendar below please circle the day you *first arrived* in the Four States Region on this trip. (If you already live in the Four States Region, please circle the day you left home to begin the trip.)

2 Please circle the day you *left* the Four States Region at the end of this trip. If you made any brief side trips *outside* of the region during this trip, please tell us *only* about your *final* departure from the region. (If you already live in the Four States Region, please circle the day you returned home at the end of the trip.)

MAY						
S	*M*	*T*	*W*	*T*	*F*	*S*
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

JUNE						
S	*M*	*T*	*W*	*T*	*F*	*S*
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

JULY						
S	*M*	*T*	*W*	*T*	*F*	*S*
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

AUGUST						
S	*M*	*T*	*W*	*T*	*F*	*S*
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Other arrival date (if not on calendar)________________________

Other departure date (if not on calendar)________________________

3 Was this your first trip to visit national parks in the Four States Region? (Circle number)

1 Yes
2 No

4 Do you currently live in the Four States Region (shaded portion of the map on the inside front cover)?

1 Yes → Skip to question **7**
2 No

5 It is important for our study to know *where you first entered* the Four States Region (shaded portion of the map on the inside front cover). Please **circle** on the map on the inside front cover the place on the border where your vehicle first crossed into the Four States Region.

If you came by *bus, train, or airplane*, please **circle** the city in the Four States Region where you began the driving part of your trip. (For example, if you flew to Denver and rented a car, you would circle Denver on the map.)

6 Where did you *finally leave* the Four States Region? Using the map on the inside front cover, please **place an "X"** on the place on the border where your vehicle finally left the Four States Region. (If you made any brief side trips outside the region during your visit, please tell us *only* about your *final* departure from the region.)

If you left by *bus, train, or airplane*, please **place an "X"** on the city where you ended the driving part of your trip to the Four States Region.

7 When you traveled around *in* the Four States Region what kind of vehicle did you use or rent?

1 Small car
2 Mid-sized car
3 Full-sized car
4 Van
5 Truck/SUV
6 Motorcycle
7 RV
8 Other (please describe) ______________________________

8 What is the total number of days you spent in the Four States Region itself during this trip? (If you left the region temporarily to make side trips during your stay do not include those days in your total.)

________total days in the Four States Region during this trip

SECTION B National Park Visits

Next we would like to ask you whether you visited any of the following seven national parks during your trip in the Four States Region. The seven national parks are listed in alphabetical order. *Please answer the first question for each national park even if you did not visit it.*

ARCHES NATIONAL PARK

9 Did you visit Arches National Park during this trip?

1 Yes
2 No → Skip to question **12**

10 How many *separate times* did you enter Arches National Park during this trip? (For example, if you left the park to get lunch then reentered through the gate you would have entered the park two times.)

1 I (we) entered the park only once
2 I (we) entered the park a total of ________ times

11 Approximately how much time in total did you spend *inside* Arches National Park during this trip?

1 ½ day
2 1 day
3 1 ½ days
4 2 days
5 More than 2 days → _______ days.

BRYCE CANYON NATIONAL PARK

12 Did you visit Bryce Canyon National Park during this trip?

1 Yes
2 No → Skip to question **15**

13 How many *separate times* did you enter Bryce Canyon National Park during this trip? (For example, if you left the park to get lunch then reentered through the gate you would have entered the park two times.)

1 I (we) entered the park only once
2 I (we) entered the park a total of _______ times

14 Approximately how much time in total did you spend *inside* Bryce Canyon National Park during this trip?

1 ½ day
2 1 day
3 1 ½ days
4 2 days
5 More than 2 days → _______ days.

CANYONLANDS NATIONAL PARK

15 Did you visit Canyonlands National Park during this trip?

1 Yes
2 No → Skip to question **18**

16 How many *separate times* did you enter Canyonlands National Park during this trip? (For example, if you left the park to get lunch then reentered through the gate you would have entered the park two times.)

1 I (we) entered the park only once
2 I (we) entered the park a total of ________ times

17 Approximately how much time in total did you spend *inside* Canyonlands National Park during this trip?

1 ½ day
2 1 day
3 1 ½ days
4 2 days
5 More than 2 days → _______ days.

GRAND CANYON NATIONAL PARK

18 Did you visit Grand Canyon National Park during this trip?

1 Yes
2 No → Skip to question **21**

19 How many *separate times* did you enter Grand Canyon National Park during this trip? (For example, if you left the park to get lunch then reentered through the gate you would have entered the park two times.)

1 I (we) entered the park only once
2 I (we) entered the park a total of ________ times

20 Approximately how much time in total did you spend *inside* Grand Canyon National Park during this trip?

1 ½ day
2 1 day
3 1 ½ days
4 2 days
5 More than 2 days → _______ days.

MESA VERDE NATIONAL PARK

21 Did you visit Mesa Verde National Park during this trip?

1 Yes
2 No → Skip to question **24**

22 How many *separate times* did you enter Mesa Verde National Park during this trip? (For example, if you left the park to get lunch then reentered through the gate you would have entered the park two times.)

1 I (we) entered the park only once
2 I (we) entered the park a total of ________ times

23 Approximately how much time in total did you spend *inside* Mesa Verde National Park during this trip?

1 ½ day
2 1 day
3 1 ½ days
4 2 days
5 More than 2 days → _______ days.

PETRIFIED FOREST NATIONAL PARK

24 Did you visit Petrified Forest National Park during this trip?

1 Yes
2 No → Skip to question **27**

25 How many *separate times* did you enter Petrified Forest National Park during this trip? (For example, if you left the park to get lunch then reentered through the gate you would have entered the park two times.)

1 I (we) entered the park only once
2 I (we) entered the park a total of _______ times

26 Approximately how much time in total did you spend *inside* Petrified Forest National Park during this trip?

1 ½ day
2 1 day
3 1 ½ days
4 2 days
5 More than 2 days → _______ days.

ZION NATIONAL PARK

27 Did you visit Zion National Park during this trip?

1 Yes
2 No → Skip to question **30**

28 How many *separate times* did you enter Zion National Park during this trip? (For example, if you left the park to get lunch then reentered through the gate you would have entered the park two times.)

1 I (we) entered the park only once
2 I (we) entered the park a total of ________ times

29 Approximately how much time in total did you spend *inside* Zion National Park during this trip?

1 ½ day
2 1 day
3 1 ½ days
4 2 days
5 More than 2 days → _______ days.

SECTION C Other Places Visited

Now we would like to ask you about other national parks, national monuments, national historic parks, national recreation areas, and cities in the Four States Region you may have visited during this trip.

Please answer "yes" to the following four questions only if your visit was planned and not a short rest-stop or a brief stop on your way to somewhere else.

30 Did you make *planned* visits to any of the following national parks in the Four States Region during this trip? (Circle all that apply)

1 Saguaro National Park (Arizona)
2 Black Canyon of the Gunnison National Park (Colorado)
3 Rocky Mountain National Park (Colorado)
4 Carlsbad Caverns National Park (New Mexico)
5 Capital Reef National Park (Utah)

31 Did you make *planned* visits to any of the following national monuments, national historic sites, national historic parks, or national recreation areas in the Four States Region during this trip? (Circle all that apply)

ARIZONA

1 Canyon de Chelly National Monument
2 Casa Grande Ruins National Monument
3 Chiricahua National Monument
4 Coronado National Monument
5 Fort Bowie National Historic Site
6 Glen Canyon National Recreation Area
7 Hubbell Trading Post National Historic Site
8 Lake Meade National Recreation Area
9 Navajo National Monument
10 Organ Pipe Cactus National Monument
11 Pipe Spring National Monument
12 Montezuma Castle National Monument
13 Sunset Crater Volcano National Monument
14 Tonto National Monument
15 Tumacacori National Historic Park
16 Tuzigoot National Monument
17 Walnut Canyon National Monument
18 Wupatki National Monument

COLORADO

19 Bent's Old Fort National Historic Site
20 Colorado National Monument
21 Curecanti National Recreation Area
22 Dinosaur National Monument
23 Florissant Fossil Beds National Monument
24 Great Sand Dunes National Monument
25 Hovenweep National Monument

NEW MEXICO

26 Aztec Ruins National Monument
27 Bandelier National Monument
28 Capulin Volcano National Monument
29 Chaco Culture National Historic Park
30 El Malpais National Monument
31 El Morro National Monument
32 Fort Union National Monument
33 Gila Cliff Dwellings National Monument
34 Pecos National Historic Park
35 Petroglyph National Monument
36 Salinas Pueblo Missions National Monument
37 White Sands National Monument

UTAH

38 Cedar Breaks National Monument
39 Glen Canyon National Recreation Area
40 Golden Spike National Historic Site
41 Hovenweep National Monument
42 Natural Bridges National Monument
43 Rainbow Bridge National Monument
44 Timpanogos Cave National Monument

Other (please specify)____

32 Did you make *planned* visits to any of the following cities during <u>this trip</u>? (Circle <u>all</u> that apply)

1 Flagstaff, Arizona
2 Phoenix, Arizona
3 Tucson, Arizona
4 Denver, Colorado
5 Durango, Colorado
6 Las Vegas, Nevada
7 Albuquerque, New Mexico
8 Santa Fe, New Mexico
9 Taos, New Mexico
10 Moab, Utah
11 Park City, Utah
12 Salt Lake City, Utah
13 Other (please specify) ____________________

SECTION D Other Planned Stops

33 Did you make *planned* visits with friends or relatives *in the Four States Region* during <u>this trip</u>? (Circle number)

1 Yes
2 No → Skip to question **35**

34 Please list each place in the region where you made these planned visits:

City/State ____________________

City/State ____________________

City/State ____________________

35 Did you make *planned* stops for work or business reasons *in the Four States Region* during <u>this trip</u>? (Circle number)

1 Yes
2 No → Skip to question **37**

36 Please list each place in the region where you made these planned stops:

City/State ____________________

City/State ____________________

City/State ____________________

SECTION E Your Party

37 How would you describe the group you were traveling with in the Four States Region? (Circle all that apply)

1 I was traveling alone → Skip to question **40**
2 I was traveling with family
3 I was traveling with friends
4 I was traveling with business associates
5 Other (please describe)____________________

38 How many people were in your vehicle?

_____ Total number of people in vehicle including children

39 How many of the people in your vehicle were…

_____ Children up to 12 years old
_____ Children 13 – 18 years old
_____ Adults 19 – 30 years old
_____ Adults 31 – 50 years old
_____ Adults 51 – 64 years old
_____ Adults 65 or older
_____ TOTAL in vehicle (should equal response to question **38**)

40 Which type of lodging did you use *most frequently* when you were in the Four States Region? (Circle one number)

1 Hotels and motels
2 RV or vehicle camping
3 Tent camping
4 Stayed with friends/relatives
5 None; it was a day trip
6 Other (please describe) __________________

41 How many nights, if any, did you camp in the Four States Region during this visit (including tent camping, RV camping, etc.)?

_____ Nights

42 Did anyone in your party use any of the following National Park passes during this trip? (Circle all that apply)

1 National Parks Pass
2 Golden Eagle
3 Golden Age Passport
4 Golden Access Passport
5 None
6 Not sure
7 Other (please describe): ____________________

43 How important to you personally was each of the following activities during your trip to the Four States Region? (Circle a number for each activity)

	Extremely Important	Very Important	Somewhat Important	Not Important
Biking	1	2	3	4
Viewing scenery; driving scenic highways	1	2	3	4
Nature study	1	2	3	4
Exploring the visitor centers	1	2	3	4
Hiking	1	2	3	4

44 This question is about what a visit to Bryce Canyon National Park is worth to you.

Visitors to Bryce Canyon National Park currently pay an entry fee of $10 per vehicle for a seven-day pass. The Park Service is not currently thinking of increasing this fee. In this question we use entry fee increases only to learn how much visiting Bryce Canyon National Park is worth to you.

> Some people would not pay more than $10 and would go elsewhere if the fee were higher.
>
> Other people would pay more to visit the Park, if necessary, because it has this much value to them.

What is the highest amount you would have paid to visit Bryce Canyon National Park during this trip? (Circle the amount) (If you have a multi-park pass, please answer as if your pass were not valid for Bryce Canyon)

$20 (current fee)	$22	$24
$26	$28	$31
$34	$37	$40
$45	$50	$60 or more

45 What was the most important factor to you in answering question **44**? (Circle one number)

1 I selected the highest amount that I would be willing to pay to visit the park.
2 I selected an amount less than I would really pay because I want the entry fee to be fair and affordable to everyone.
3 I selected an amount less than I would really pay because we already pay for national parks through taxes.
4 I selected an amount less than I would really pay in order to keep the entry fee low.
5 I selected an amount higher than I would really pay because national parks need more funding.
6 Other (please describe) ____________________________

SECTION F About Yourself

Your answers to these questions are important for our statistical analysis.

46 In what year were you born? 19____

47 What is the highest level of education that you have completed?

1 Less than high school
2 Some high school
3 High school or GED
4 Technical or trade school degree
5 Some college
6 College graduate
7 Graduate school

48 Which of the following best describes your employment status? (Circle one number)

1 Employed full time
2 Employed part time
3 Work in the household (for example, raise children)
4 Unemployed
5 Retired
6 Student
7 Other (please describe) ______________________

49 When you were planning your trip to national parks in the Four States Region, could you have chosen a longer trip or were you restricted by your job, limited vacation time, family commitments, etc.?

1 Yes, I could have chosen a longer trip
2 No, I could not have chosen a longer trip → Skip to **51**

50 What is the longest possible trip you could have taken to visit national parks in the Four States Region? (Circle one number)

1 Two to three days
2 One week
3 Two weeks
4 Three weeks
5 Four weeks
6 More than four weeks

51 How much did you enjoy your visit to Bryce Canyon National Park?

1 A lot
2 Quite a bit
3 Somewhat
4 Not very much
5 Not at all

52 What was your approximate total household income *before* taxes for the year 2001? Just your best estimate. Include money from jobs, rent, pensions, social security, etc. *This information will be kept confidential. It is very important to our economic analysis.*

1 Less than $15,000 per year
2 $15,000 to $20,000 per year
3 $20,000 to $30,000 per year
4 $30,000 to $40,000 per year
5 $40,000 to $50,000 per year
6 $50,000 to $75,000 per year
7 $75,000 to $100,000 per year
8 $100,000 to $150,000 per year
9 More than $150,000 per year

53 What is your gender?

1 Male
2 Female

54 Do you have any other comments about your trip?

____

____

____

____

____

____

Thank you for participating in our survey! Your opinions and information about your trip are of great importance to us. Please return this booklet to us in the enclosed self-addressed, stamped envelope.

Appendix B

INTERCEPT SURVEY

<u>Intercept Interviewer Script</u>

Hi! Welcome to __________ National Park.

(PURPOSE OF THE STUDY)

We are conducting a survey in order to learn more about your trip to the four corners region.

How many people are in your vehicle today? => RECORD ANSWER ON CARD

How many of you are U.S. citizens or Native Americans? => RECORD ANSWER ON CARD

IF NONE, THEN THANK PERSON AND TERMINATE INTERVIEW.

How many of the U.S. citizens or Native Americans are age 18 or older? => RECORD ANSWER ON CARD

IF NONE, THEN THANK PERSON AND TERMINATE INTERVIEW.

Which of the adult U.S. citizens or Native Americans in the car most recently celebrated a birthday?

ADDRESS REMAINDER OF INTERVIEW TO THE ADULT U.S. CITIZEN WHO HAD THE MOST RECENT BIRTHDAY.

We would like to mail you a short survey to learn more about your trip after you return home. Could you please write your address on this card so that we can mail you the survey? HAND CARD AND PEN TO RESPONDENT.

Thank you for your cooperation. Enjoy your stay in _____ NATIONAL PARK!

Appendix C

TABLES

Table C1: Intercept Survey Recruitment Schedule

SOUTHWESTERN PARKS INTERCEPT SURVEY RECRUITMENT SCHEDULE (2002)									
	6/15 Sat.	6/16 Sun.	6/17 Mon.	6/18 Tues.	6/19 Wed.	6/20 Thur.	6/21 Fri.	6/22 Sat.	6/23 Sun.
Arches	X	X	X	X					
Canyonlands Island Entrance						X		X	
Canyonlands Needles Entrance							X		X
Bryce	X	X	X	X					
Zion - South Entrance						X		X	
Zion - East Entrance							X		X
Mesa Verde	X	X	X	X					
Petrified Forest North Entrance						X		X	
Petrified Forest South Entrance							X		X
Grand Canyon Desert View Entrance	X		X						
Grand Canyon South Rim Entrance		X		X					
Grand Canyon North Rim Entrance							X	X	

Table C2: Target Sampling Rates

Exhibit A-1 TARGET SAMPLING RATES	
Park Entrance	**Target Sampling Rate[a]**
Arches Main Entrance	1 in 4
Bryce Canyon Main Entrance	1 in 4, 1 in 5, or 1 in 6[b]
Canyonlands Island Entrance	1 in 1
Canyonlands Needles Entrance	1 in 1
Grand Canyon South Rim Entrance	1 in 9
Grand Canyon Desert View Entrance	1 in 2
Grand Canyon North Rim Entrance	1 in 2
Mesa Verde Main Entrance	1 in 3
Petrified Forest North Entrance	1 in 1
Petrified Forest South Entrance	1 in 1
Zion East Entrance	1 in 3
Zion South Entrance	1 in 6 or 1 in 7[c]

[a] These represent *target* sampling rates; the *actual* sampling rates may have varied from these target rates for a variety of reasons. For example, at some entrances, the intercept location designated by NPS staff did not provide sufficient parking to intercept vehicles at the target sampling rate during high-visitation time periods. TRA personnel recorded information that will allow us to calculate the actual sampling rate at all entrances for the RUM analysis.
[b] The target sampling rate varied depending on the number of gates open to vehicles entering the park. The target rate was 1 in 5 when one gate was open, the target rate was 1 in 4 when two gates were open, and the target rate was 1 in 6 when three gates were open.
[c] The target sampling rate varied depending on the number of gates open to vehicles entering the park. The target rate was 1 in 7 when one gate was open and 1 in 6 when two gates were open.

Table C3: Visitation Data – Arches

Exhibit A-2 ARCHES NATIONAL PARK MAIN ENTRANCE: VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002									
	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data[a]									
Vehicles Eligible for Intercept	1,018	897	844	851	836	967	918	1,091	1,065
Vehicles Eligible for Intercept and Re-entering Park	106	81	56	78	78	83	66	114	126
Vehicles Ineligible for Intercept	7	10	4	5	6	7	5	8	12
Cash Register Total	1,131	988	904	934	920	1,057	989	1,213	1,203
Car Counter Total	1,406	1,253	1,175	1,212	1,223	1,223	1,221	1,507	1,487

[a] Arches National Park did not maintain detailed cash register data on June 22 and June 23, when NPS temporarily suspended park entry fees. We estimate vehicles for June 22 by assuming that the percentage change in vehicles between June 22 and June 15 (the previous Saturday) is equal to the percentage change in the car counter total between the two days. We estimate vehicles for June 23 in a similar manner.

Table C4: Visitation Data – Bryce Canyon

Exhibit A-3 BRYCE CANYON NATIONAL PARK MAIN ENTRANCE: VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002									
	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data									
Vehicles Eligible for Intercept	1,030	1,001	945	937	995	1,028	988	1,170	1,110
Vehicles Eligible for Intercept and Re-entering Park	318	334	290	348	311	327	303	389	407
Vehicles Ineligible for Intercept	170	214	208	209	231	236	179	175	187
Cash Register Total	1,518	1,549	1,443	1,494	1,537	1,591	1,470	1,734	1,704
Car Counter Total	1,682	1,787	1,537	1,784	1,835	1,750	1,802	1,912	1,970

Table C5: Visitation Data – Canyonlands Entrance 1

Exhibit A-4 CANYONLANDS NATIONAL PARK NEEDLES ENTRANCE: VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002									
	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data									
Vehicles Eligible for Intercept	75	60	76	82	40	57	58	87	82
Vehicles Eligible for Intercept and Re-entering Park	13	13	12	6	4	7	3	5	1
Vehicles Ineligible for Intercept	4	1	17	10	17	18	13	6	2
Cash Register Total	92	74	105	98	61	82	74	98	85
Car Counter Total	136	106	137	153	124	124	123	123	126

Table C6: Visitation Data – Canyonlands Entrance 2

Exhibit A-5
CANYONLANDS NATIONAL PARK ISLAND-IN-THE-SKY ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data									
Vehicles Eligible for Intercept	222	230	194	242	209	246	178	222	232
Vehicles Eligible for Intercept and Re-entering Park	10	6	9	6	5	3	7	8	3
Vehicles Ineligible for Intercept	0	4	11	10	12	12	6	9	8
Cash Register Total	232	240	214	258	226	261	191	239	243
Car Counter Total	299	314	283	315	302	316	296	322	255

Table C7: Visitation Data – Grand Canyon Entrance 1

Exhibit A-6
GRAND CANYON NATIONAL PARK DESERT VIEW ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data									
Vehicles Eligible for Intercept	477	505	485	523	547	564	484	541	491
Vehicles Eligible for Intercept and Re-entering Park	20	29	14	23	31	34	31	30	19
Vehicles Ineligible for Intercept	46	57	74	28	61	48	50	59	67
Cash Register Total	543	591	573	574	639	646	565	630	577
Car Counter Total	1,697	1,733	1,822	1,818	1,867	1,874	1,787	1,888	1,792

Table C8: Visitation Data – Grand Canyon Entrance 2

Exhibit A-7
GRAND CANYON NATIONAL PARK NORTH RIM ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data									
Vehicles Eligible for Intercept	487	462	362	415	496	491	392	349	543
Vehicles Eligible for Intercept and Re-entering Park	29	45	20	25	63	35	13	12	40
Vehicles Ineligible for Intercept	23	40	37	31	32	48	20	18	25
Cash Register Total	539	547	419	471	591	574	425	379	608
Car Counter Total	1,161	1,190	1,222	1,214	1,395	1,266	1,155	1,193	1,259

Table C9: Visitation Data – Grand Canyon Entrance 3

Exhibit A-8
GRAND CANYON NATIONAL PARK SOUTH RIM ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data									
Vehicles Eligible for Intercept	2,206	2,228	2,420	2,350	2,477	2,341	2,303	2,989	2,363
Vehicles Eligible for Intercept and Re-entering Park	486	564	556	681	704	644	596	517	546
Vehicles Ineligible for Intercept	547	539	792	709	764	735	576	573	482
Cash Register Total	3,239	3,331	3,768	3,740	3,945	3,720	3,475	4,079	3,391
Car Counter Total	13,264	11,895	13,280	14,172	13,964	13,554	13,500	14,566	12,873

Table C10: Visitation Data – Mesa Verde

Exhibit A-9
MESA VERDE NATIONAL PARK MAIN ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data									
Vehicles Eligible for Intercept	675	665	729	824	773	678	686	753	665
Vehicles Eligible for Intercept and Re-entering Park	17	1	15	25	23	18	28	21	8
Vehicles Ineligible for Intercept	21	27	69	44	58	52	89	44	27
Cash Register Total	713	693	813	893	854	748	803	818	700
Car Counter Total	1,908	1,909	2,119	2,362	2,387	2,179	2,058	1,916	2,248

Table C11: Visitation Data – Petrified Forest Entrance 1

Exhibit A-10
PETRIFIED FOREST NATIONAL PARK PAINTED DESERT ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data									
Vehicles Eligible for Intercept	326	384	449	463	400	292	320	241	401
Vehicles Eligible for Intercept and Re-entering Park	9	6	5	3	7	0	6	5	5
Vehicles Ineligible for Intercept	0	1	1	2	2	0	0	1	1
Cash Register Total	335	391	455	468	409	292	326	247	407
Car Counter Total	430	412	466	480	503	346	370	337	423

Table C12: Visitation Data – Petrified Forest Entrance 2

Exhibit A-11
PETRIFIED FOREST NATIONAL PARK RAINBOW FOREST ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data									
Vehicles Eligible for Intercept	219	235	268	295	269	218	172	85	176
Vehicles Eligible for Intercept and Re-entering Park	1	0	2	5	4	0	1	2	2
Vehicles Ineligible for Intercept	0	0	0	1	2	0	0	0	0
Cash Register Total	220	235	270	301	275	218	173	87	178
Car Counter Total	321	315	400	466	365	294	261	181	260

Table C13: Visitation Data – Zion Entrance 1

Exhibit A-12
ZION NATIONAL PARK SOUTH ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data[a,b]									
Vehicles Eligible for Intercept	1,119	1,224	1,030	998	1,103	1,331	1,395	1,236	1,367
Vehicles Eligible for Intercept and Re-entering Park	195	159	198	199	212	232	243	215	178
Vehicles Ineligible for Intercept	106	96	113	127	170	194	203	117	107
Cash Register Total	1,420	1,479	1,341	1,324	1,485	1,757	1,841	1,568	1,652
Car Counter Total	2,369	2,179	2,268	2,150	2,330	2,447	2,564	2,617	2,435

[a] The cash register data for June 18 and June 19 included totals for each of the three entrances to Zion National Park (South, East, and River). However, for June 15, 16, 17, and 20, the cash register data only included *overall* totals that combined the three entrances. For these four days, entrance-specific totals were calculated by assuming that the distribution across entrances was equivalent to the distribution across entrances for June 18 and 19.

[b] Zion National Park did not maintain detailed cash register data June 21 to June 23 (on June 22 and June 23, NPS temporarily suspended park entry fees). We estimate vehicles for June 21 by assuming that the percentage change in vehicles between June 21 and June 20 is equal to the percentage change in the car counter total between the two days. We estimate vehicles for June 22 and June 23 in a similar manner, but by using the percentage change from June 15 (the previous Saturday) and June 16 (the previous Sunday), respectively.

Table C14: Visitation Data – Zion Entrance 2

Exhibit A-13
ZION NATIONAL PARK RIVER ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data[a,b]									
Visitors Eligible for Intercept	240	263	221	215	237	285	286	268	323
Visitors Eligible for Intercept and Re-entering Park	75	61	76	70	88	89	89	84	75
Visitors Ineligible for Intercept	16	14	17	15	30	29	29	18	17
Cash Register Total	331	338	314	300	355	403	404	370	415
Car Counter Total[c]	--	--	--	--	--	--	--	--	--

[a] The cash register data for June 18 and June 19 included totals for each of the three entrances to Zion National Park (South, East, and River). However, for June 15, 16, 17, and 20, the cash register data only included *overall* totals that combined the three entrances. For these four days, entrance-specific totals were calculated by assuming that the distribution across entrances was equivalent to the distribution across entrances for June 18 and 19.
[b] Zion National Park did not maintain detailed cash register data June 21 to June 23 (on June 22 and June 23, NPS temporarily suspended park entry fees). We estimate visitors for June 21 by assuming that the percentage change in visitors between June 21 and June 20 is equal to the percentage change in the car counter total between the two days (using car counter data from the South and East entrances). We estimate vehicles for June 22 and June 23 in a similar manner, but by using the percentage change from June 15 (the previous Saturday) and June 16 (the previous Sunday), respectively.
[c] The River Entrance is for walk-in visitors only.

Table C15: Visitation Data – Zion Entrance 3

Exhibit A-14
ZION NATIONAL PARK EAST ENTRANCE:
VISITATION DATA FOR JUNE 15, 2002 THROUGH JUNE 23, 2002

	June 15	June 16	June 17	June 18	June 19	June 20	June 21	June 22	June 23
Cash Register Data[a,b]									
Vehicles Eligible for Intercept	606	663	558	517	621	721	653	690	1,008
Vehicles Eligible for Intercept and Re-entering Park	62	50	63	66	64	73	66	71	76
Vehicles Ineligible for Intercept	43	39	46	57	64	79	71	49	59
Cash Register Total	711	752	667	640	749	873	790	810	1,143
Car Counter Total	990	812	1,116	913	971	1,113	1,007	1,128	1,235

[a] The cash register data for June 18 and June 19 included totals for each of the three entrances to Zion National Park (South, East, and River). However, for June 15, 16, 17, and 20, the cash register data only included *overall* totals that combined the three entrances. For these four days, entrance-specific totals were calculated by assuming that the distribution across entrances was equivalent to the distribution across entrances for June 18 and 19.
[b] Zion National Park did not maintain detailed cash register data June 21 to June 23 (on June 22 and June 23, NPS temporarily suspended park entry fees). We estimate vehicles for June 21 by assuming that the percentage change in vehicles between June 21 and June 20 is equal to the percentage change in the car counter total between the two days. We estimate vehicles for June 22 and June 23 in a similar manner, but by using the percentage change from June 15 (the previous Saturday) and June 16 (the previous Sunday), respectively.

Table C16: Survey Response Rate by Park

RESPONSE RATE BY PARK						
	Intercept Survey[a]			Mail Survey		Overall response rate (intercept survey response rate times mail survey response rate)
	Total Eligible for Survey	Total Recruits	Response Rate[b]	Total Surveys Completed	Response Rate[c]	
Arches	682	654	96%	490	75%	72%
Bryce	625	585	94%	423	72%	67%
Canyonlands	413	407	99%	303	74%	73%
Grand Canyon	1,251	1,241	99%	819	66%	65%
Mesa Verde	760	697	92%	483	68%	63%
Petrified Forest	577	556	96%	349	63%	60%
Zion	735	696	95%	444	64%	61%
Total	**5,043**	**4,836**	**96%**	**3,311**	**68%**	**65%**

[a] The intercept survey consisted of a very brief set of questions designed to evaluate the respondent's eligibility for a longer mail survey (see Appendix C). Eligible respondents (U.S. citizens 18 or older) were recruited to participate in the longer mail survey.
[b] The response rate for the intercept survey was calculated as the number of recruits divided by the total number of eligible respondents.
[c] The response rate for the mail survey was calculated as the total completed surveys divided by the total number of recruits from the intercept survey.

Table C17: Entry and Exit Points in the Four States Region

Border	Access Point Name	IEc-code	Access Point Type	LAT	LONG
	Salt Lake City, UT	1	city (bus, train, airplane)	40.7883878	-111.9777731
	Denver, CO	2	city (bus, train, airplane)	39.8584081	-104.6670019
	Grand Jct, CO	3	city (bus, train, airplane)	39.1224125	-108.5267347
	Alburquerque, NM	4	city (bus, train, airplane)	35.0402222	-106.6091944
	Santa Fe,NM	5	city (bus, train, airplane)	35.6171086	-106.0894228
	Phoenix, AZ	6	city (bus, train, airplane)	33.4341667	-112.0080556
	Flagstaff, AZ	7	city (bus, train, airplane)	35.1384547	-111.6712183
	Tucson,AZ	8	city (bus, train, airplane)	32.1160833	-110.9410278
NV/UT	I 80	9	road crossing - Interstate	40.73	-114.03
	I 15	10	road crossing - Interstate	42.01	-112.21
WY/UT	I 80	11	road crossing - Interstate	41.28	-111.05
	I 25	12	road crossing - Interstate	41.01	-104.93
	I 76	13	road crossing - Interstate	41.01	-102.22
	I 70	14	road crossing - Interstate	39.33	-102.04
TX/NM	I 40	15	road crossing - Interstate	35.18	-103.03
	I 10	16	road crossing - Interstate	31.84	-106.58
	I 19	17	road crossing - Interstate	31.37	-110.96
	I 8	18	road crossing - Interstate	32.73	-114.59
	I 10	19	road crossing - Interstate	33.64	-114.52
CA/AZ	I 40	20	road crossing - Interstate	34.74	-114.48
AZ/NV	I 15	21	road crossing - Interstate	36.83	-114.04
	US 6/50	22	road crossing - secondary road	39.07	-114.06
UT/ID	I 84	23	road crossing - Interstate	42	-112.85
	US 191	24	road crossing - secondary road	41.02	-109.46
	SR-789(WY) / SR-13(CO)	25	road crossing - secondary road	41.01	-107.65
	US 6	26	road crossing - secondary road	40.58	-102.07

Table C17 Continued

	US 50	27	road crossing - secondary road	38.04	-102.05
	US 87/64	28	road crossing - secondary road	36.45	-103.04
	US 84	29	road crossing - secondary road	34.4	-103.04
	US 62/180	30	road crossing - secondary road	32.74	-103.07
	US 285	31	road crossing - secondary road	32.02	-104.06
	US 191	32	road crossing - secondary road	31.37	-109.58
	US 93	33	road crossing - secondary road	35.99	-114.86
	Cedar City, UT	34	city (bus, train, airplane)	37.7009664	-113.0988458
	Page, AZ	35	city (bus, train, airplane)	36.9261111	-111.4483611
	Gunnison, CO	36	city (bus, train, airplane)	38.5339444	-106.9330278
	Springs, CO	37	city (bus, train, airplane)	38.8058056	-104.700
	Durango, CO	38	city (bus, train, airplane)	37.1515167	-107.7537692
	Farmington, NM	39	city (bus, train, airplane)	36.74125	-108.230
	Gallup, NM	40	city (bus, train, airplane)	35.5110583	-108.7893094
	Grand Canyon (So. Rim)	41	city (bus, train, airplane)	35.9523539	-112.1469647

Note: For Coded entries 1-8 and 34-41, we used the lat/long for the city's main airport (see www.airnav.com). For Coded entries 9-33, we estimated the Lat/Long using "mouse rollover" on GIS.

Table C18: 111 Portfolio Specific Constant Model (Full set of PSC coefficients)

Variable	Coefficients	t-statistics
Travel Cost (in $1000)	-4.985	-13.15
Cost * Flextime	0.050	0.56
Cost * Car Renter	-0.227	-1.83
Cost * Income (in $1000)	0.010	7.52
Cost * Visited Other Sites	2.235	8.32
PSC2	-0.715	-3.95
PSC3	-2.330	-9.56
PSC4	2.850	23.85
PSC5	0.578	4.79
PSC6	-1.460	-8.43
PSC7	0.929	6.59
PSC8	-2.022	-5.38
PSC9	-0.050	-0.36
PSC10	0.078	0.3
PSC11	-1.092	-5.27
PSC12	-3.151	-6.13
PSC13	-0.694	-2.68
PSC14	-3.129	-5.29
PSC15	-0.033	-0.14
PSC16	-17.449	-131.21
PSC17	-3.090	-4.31
PSC18	1.284	8.98
PSC19	-1.212	-2
PSC20	-3.211	-7.58
PSC21	-3.622	-3.6
PSC22	-2.914	-4.06
PSC23	0.428	1.98
PSC24	1.439	9.46
PSC25	0.861	4.71
PSC26	-1.787	-6.55
PSC27	-1.596	-4
PSC28	-3.740	-3.71
PSC29	-1.692	-5.28
PSC30	-0.767	-1.97
PSC31	-1.848	-3.86
PSC32	-3.027	-2.98

Table C18 Continued

PSC33	-0.033	-0.14
PSC34	-1.111	-3.43
PSC35	-0.456	-2.46
PSC36	-2.654	-5.66
PSC37	-1.788	-4.99
PSC38	-0.404	-1.08
PSC39	-0.803	-2.45
PSC40	-1.320	-3.07
PSC41	-2.957	-4.09
PSC44	-2.579	-3.49
PSC47	-0.376	-0.83
PSC48	-1.117	-2.62
PSC49	-1.044	-3.39
PSC50	2.200	11.15
PSC52	-0.713	-1.97
PSC53	-0.478	-1.39
PSC54	-16.261	-72.86
PSC55	-2.270	-3.71
PSC56	-2.142	-2.9
PSC59	-16.877	-96.6
PSC60	0.681	2.84
PSC61	-0.361	-1.02
PSC62	-0.205	-0.67
PSC63	-2.269	-4.3
PSC64	-0.489	-1.17
PSC65	-1.225	-2.46
PSC66	-2.976	-4.02
PSC67	0.565	1.99
PSC68	-0.757	-1.75
PSC69	-0.792	-1.44
PSC70	1.016	3.58
PSC71	-1.901	-1.86
PSC72	-0.648	-1.5
PSC73	-0.460	-1.13
PSC74	-1.403	-2.5
PSC75	-0.834	-2.12
PSC76	-0.682	-1.6
PSC78	-3.096	-3.03
PSC79	-2.156	-3.51
PSC80	0.085	0.2

Table C18 Continued

PSC81	-0.688	-1.3
PSC82	-0.471	-0.91
PSC84	-1.658	-2.17
PSC85	-2.060	-2
PSC86	0.037	0.07
PSC88	-0.777	-1.41
PSC89	-1.484	-1.44
PSC90	-1.304	-2.37
PSC91	1.053	3.27
PSC92	1.396	5.1
PSC93	-1.526	-1.49
PSC94	-2.040	-2
PSC95	-1.252	-1.64
PSC96	0.199	0.36
PSC98	0.352	0.74
PSC99	-0.063	-0.13
PSC100	-0.347	-0.64
PSC101	2.090	6.31
PSC102	-3.173	-3.07
PSC103	0.502	1.45
PSC104	-0.258	-0.65
PSC105	-0.720	-1.36
PSC106	0.635	1.64
PSC107	1.571	4.37
PSC109	-0.598	-1.32
PSC111	-1.471	-1.92
PSC112	-2.449	-2.39
PSC113	-0.610	-1.02
PSC114	-0.454	-0.59
PSC115	-0.989	-1.27
PSC116	-1.688	-1.62
PSC118	0.792	2.12
PSC120	-0.713	-0.67
PSC121	2.481	6.02
PSC122	2.008	5.02
PSC123	-0.439	-0.67
PSC124	0.915	2.18
PSC125	-0.247	-0.41
PSC127	1.945	4.23

Table C18 Continued

Log-likelihood	-7152.5593
Sample size	2719