The Impact of Water Quality on Southern California Beach Recreation: A Finite Mixture Model Approach

Author
Hilger, James R.

Publication Date
2006

Peer reviewed

The Impact of Water Quality on Southern California Beach Recreation: A Finite Mixture Model Approach

by

James Robert Hilger

B.A.(Reed College) 1995
M.S. (University of Nevada at Reno) 1998

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Agricultural and Resource Economics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor W. Michael Hanemann, Chair
Professor Jeffrey T. LaFrance
Professor J. Miguel Villas-Boas

Spring 2006

The Impact of Water Quality on Southern California Beach Recreation:

A Finite Mixture Model Approach

by

James Robert Hilger

Abstract

The Impact of Water Quality on Southern California Beach Recreation:

A Finite Mixture Model Approach

by

James Robert Hilger

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor W. Michael Hanemann, Chair

This dissertation uses a finite mixture logit (FML) model to investigate the heterogeneity of preferences of beach users for water quality at beaches in Southern California. The results are compared with conventional approaches based conditional logit (CL) and random parameters logit (RPL). The FML approach captures variation in preferences by modeling individual recreator choices using a mixture of several distinct preference groups, where group membership is a function of individual characteristic and seasonal variables. The FML parameter estimates are used to calculate welfare measures for improvements in beach quality through a reduction of water pollution. The FML segment specific welfare measures bound the traditional CL and RPL mean welfare estimates, and have the advantage of highlighting the distribution of the population sample's preferences. Analysis of beach recreation site choice data indicates the existence of four representative prefer-

ence groups within the survey respondent sample. As a result, willingness to pay measures for improvements in water quality and other beach site attribute changes can be weighted across individuals to calculate the distribution of individual welfare measures.

One group of recreators is characterized as people who go to the beach and engage in water recreation with children. An interesting finding is that this group has a lower mean WTP for improving water quality than groups who go without children. This may well be an example of cognitive dissonance: parents find they go to the beach more often than others who don't have children, since that keeps the children occupied and happy, and they adapt their perception of the water quality to be consistent with their behavior.

Previous environmental and resource economic applications of the FML have been limited to applications with small choice sets (6) and group membership variables (4). This paper extends the FML model through the estimation of a large (51) choice set with 9 membership variables. This application is the first to incorporate seasonal variables into the group membership function to capture seasonal heterogeneity.

Estimated welfare changes are calculated using the compensating variation measure for several hypothetical beach closure and water quality degradation scenarios. Estimation results indicate that the FML welfare estimates differ from those calculated using the traditional logit or RPL models. The FML model sheds light onto which subsets of beach recreators are likely to be impacted by different scenarios of resource change.

Professor W. Michael Hanemann
Dissertation Committee Chair

To my family, friends, and the ocean people.

0.1 Preface

Coastal resources are increasingly strained by pollution from storm drain run off, sewage leaks, and other sources. Beach closures and advisories in Southern California have been increasing for the past nine years, reaching 1469 Los Angeles county postings in 2004 alone. And swimming in affected water is associated with illness. Any meaningful plan for coastal management that addresses these issues must not only take into account the complex and diverse nature of coastal resources, but also preference heterogeneity of potential users. Models are needed that generate accurate forecasts of an individual's choice of beaches among its alternatives and estimates of the value of improving water quality and other beach attributes. These models could serve as a useful tool to managers charged with the equitable management of the resource, and are crucial in litigation settings for the enumeration of damages caused by resource degradation. A useful choice forecast model in this context can be characterized by three main features: first it should provide an unbiased estimate of resource use and of welfare measures for changes to the resource base. Second, it should capture how the resource use forecasts and welfare estimates vary both within the population and seasonally. And finally, it should provide guidance regarding possible resource management policies.

Exploring the preference heterogeneity with respect to beach recreation choice as a function of water quality and additional beach characteristics allows me to incorporate these features in an econometric model approach. I use a panel diary dataset on beach recreation choice that was collected by the Southern California Beach Valuation Project research team, in which I participated. The economics choice modeling literature often assumes, for computational reasons, a single representative agent framework to estimate the impact of site attributes (e.g. water quality) on the decision between several alternatives such as recreational beach choice. However, these models often fail to address an important practical issue: different user groups typically value different characteristics of recreation sites and demand different services. One problematic trait of the single representative agent model (such as logit or random parameters logit (RPL)) is that it handles the variation in preferences for a given attribute by averaging over the individuals in the population. This misspecification of modeling multiple preference types with a single representative consumer can lead to biased estimates. Additionally, the failure to explore the variation in preferences, and its resulting distribution of welfare measures, leads to a loss of information reflecting the impact which changes in attributes, such as water quality, have on different segments of society.

To address these shortcomings, I utilize the finite mixture logit model (FML), which represents the population as a mixture of several distinct preference types. This model simultaneously estimates the marginal benefits associated with beach attributes for different groups, and the likelihood that an individual is a member of a specific group characterized by individual and seasonal attributes.

Using the FML model, I overcome three limitations of the standard logit model. First, the FML model produces less bias in parameter estimates which can be used to improve accuracy in forecasting resource use and generating welfare measures. My simulation results indicate that the FML outperforms the logit and RPL in estimating preference parameters and welfare measures for datasets characterized by systematic heterogeneity.

Second, the model allows for the estimation of unique preference groups. Estimation results indicate the existence of four unique representative preference groups which differ qualitatively and quantitatively with regard to the weights they place on water pollution levels and other beach characteristics that affect their beach choice. And finally, the model enables me to estimate the probability that an individual is a member of these preference groups. For instance, trips to the beach characterized by gender differences (male), seasonal differences (winter), and water use are associated with a higher relative valuation of an improvement in water quality; while the trips characterized by water contact and being in the presence of children are associated with lower welfare measures.

FML model estimates for the welfare impact of water quality changes for beach recreators is several times larger that the measure calculated from traditional logit or RPL model estimates. The marginal willingness to pay (mWTP) for a marginal improvement in the water pollution that is on average 4.6 times higher when using the FML model relative to the standard logit and RPL models. This valuation estimate spans from 3 times lower to 14 times higher the traditional mean mWTP estimate for individual beach recreators, depending on their individual characteristics. Calculation of the compensating variation consumer surplus welfare measure show a similar difference between those estimates calculated with FML versus CL model parameter estimates. Overall, the powerful combination of being able to simultaneously estimate the marginal benefits associated with different attributes for different groups and assigning group membership to individuals makes the FML an important tool for resource managers, analysts, and researcher alike.

Contents

List of Figures

List of Tables

Acknowledgments

Acknowledgments are one of those things that are fun to read if you're in them or curious about how a project came about. But they are also unfortunately prone to leave out people that helped out and in that special way were important to the project's completion but not, well, hmmm, you know who you are.

Caveats aside, I wish to first thank my parents, Christine Lee and Thomas Joseph Hilger; without their patience and support of my endeavors this paper and all of its background work may have never happened. For thoughtful comments, clear insights, dizzying questions, pats on the back, character building and late night projects, respect, support, patience, encouragement, and belly aching jokes I thank Michael Hanemann. Questions about utility theory, choice theory, and groupings of three; Jeffrey LaFrance has taught me so much about perseverance and the inherent value of getting things right and loads of RAM. I thank Miguel Villas-Boas for his thoughtful and clear comments that helped me proceed to the project's completion. What I appreciate most from the above is the unqualified nature of their support in helping me accomplish what I choose to do.

I am grateful to Cindy Thomson for recognizing the importance of this project and supporting it through the National Marine Fisheries Service and California Sea Grant. This support benefited the project not only in making it financially feasible, but also through the opportunity it gave me to interact with and learn from my fellow NMFS Fellows.

I am grateful to Guido Imbens and Sofia Villas-Boas; their comments and suggestions both before and after orals proved to be extremely useful. I want to thank other members of the faculty who were extremely generous with their input support, and time. I also acknowledge and thank Chris Busch, Caitlin Dyckman, Koichi Kurisawa, David Layton, Linwood Pendleton, and Felipe Vasquez for their involvement with the Southern California Beach Valuation Project. I thank Craig Mohn, not only for pulling together the dataset and answering my coding questions, but for his support throughout the entire period I worked on the project.

Heartfelt thanks go to my Portland people, both those that have found a place on past acknowledgments and those that are new members of my village. Moving forward through the years, I recognize the McDaniel's you both served as a useful ballast to this endeavor. Mark Edson, thank you for playing at recess since the 1^{st} grade. Thanks to my brothers, Jeffrey and Stephen; your encouragement has meant a great deal.

Ah, fellow ARE classmates. There are so many of you that have been instrumental. Karina Schogold and Kathy Baylis, both of you have welcomed my questions for since at least day one. Emma Aisbett for helpful support. Susan Stratton for her friendship and a steady stream of proofreads and LaTex tutorials. Gregory Rafert for his intellectual and his culinary curiosity. I also wish to single out Ricardo Cavazos-Cepeda, Kasandra Griffin, Kurt Huffman, Chris Lassell, Robert Seamans, Tom Sayre, and Harry Bacca.

Lastly, I recognize my Grandmother, Louise Dickson. If it was not for her I would not know the beach and ocean so well.

Thank you to all that helped, all remaining errors are mine.

Ike 'ia no ka loea i ke kuahu.

Chapter 1

Introduction

Environmental resource protection and management requires the ability to assign values to non-market goods. While the literature has generally focused on the *average* valuation and preferences for these goods, the importance of the *distribution* of preferences for environmental amenities for populations with diverse preferences has often been neglected.

With over 150 million visits a year, the beaches of Southern California provide environmental and recreational amenities to a diverse user population (USFS , 2000). However this coastal resource is strained by pollution that has led to increases in beach closures and pollution advisories in Southern California for the past nine years, with 1,469 postings in Los Angeles County in 2004 alone. The closure of beaches and the poor water quality even when beaches are open impair the public's use and enjoyment of the beaches of Southern California; and create loss of business and a significant reduction in the public welfare. However the costs associated with the widespread water management plan is estimated to be in the billions (Gordon, 2002). The magnitude of the cost estimates raises serious policy questions about the economic benefits of storm water pollution control.[1]

Not only is there a loss of welfare due to impaired use and enjoyment of the public beaches , but there can also be health impacts from swimming in polluted ocean waters, including upper respiratory infection and other illnesses.[2] Public concern regarding this degradation has prompted the approval of several Legislative and Assembly bills,[3] promoted cleanup and monitoring efforts, and increased the need for careful estimation of the welfare impact of water pollution at Southern California beaches for cost benefit analyses used in litigation, remediation, and general management. This paper investigates the willingness to pay for a reduction in beach water pollution and illustrates how these values vary by recreator characteristics and season.

Varying preferences of recreational users and the multiple use nature of beach sites complicate the estimation of willingness to pay measures for improvements in water quality and other beach attributes. Systematic preference heterogeneity can lead to bias in parameter estimates if left unaccounted for. This paper addresses systematic preference heterogeneity by utilizing a finite mixture logit (FML) random utility model. A panel trip diary data set documenting 4,462 Southern California winter and summer beach trips for 595 recreators from December 2000 to November 2001 is analyzed using the FML approach.

[1] Another arena where the valuation of beach recreation has arisen is the prevention of oil spills and the measurement of damages caused by oil spills and benefits from oil spill prevention. In 1969 there was a major oil spill near Santa Barbara which attracted serious attention from economists interested in measuring the economic value of the damages caused by the spill (Sorensen. 1975). In February 1990 there was an oil spill off the coast at Huntington Beach which triggered a law suit by the State of California against the vessel owner that actually went to trial in the fall of 1997 and led to the award of damages amounting to $12.75 million as the value of the public's lost use and enjoyment due to the closure of beaches following the spill (Chapman and Hanemann 2001).

[2] A large epidemiological study, The Santa Monica Bay Restoration Project study, found an increase in the risk of contracting an illness when swimming near storm drains. Recreators that swam near storm drains were 57% more likely to suffer symptoms of a fever than other swimmers (Haile et. al., 1996). For a recent review of health risks associated with beach water pollution see NRDC (2005).

[3] California Assembly Bill 411, the "Right to Know" bill, was passed in 1999.

Estimation results indicate that beach recreators can be characterized by one of four distinct representative groups by beach recreational decisions, and individual and seasonal attributes. This information is then used in calculating welfare estimates for each individual in the sample and the weighted average measure for the population. I find that the welfare estimates associated with an environmental improvement vary significantly both within the population and across seasons. One interesting result is that the presence of children on a beach trip that involves water contact is associated with lower mean mWTP estimates for improvements in water quality.

1.1 Contributions of this Research

While FML models have been estimated previously in the environmental and resource economics literature this research makes three main contributions. First, this dissertation is the first FML application focused on modeling the welfare and behavioral impacts of an environmental good associated with health outcomes. Water pollution is a widespread problem in many coastal and fresh water areas, and recreational swimming is the second most popular recreation activity in the United States with over 90 million participants (NRDC, 2004). Unfortunately, countless recreators swim in water that does not meet the EPA health standards (NRDC, 2004). This drives an increasing interest in determining what draws recreators to specific beaches (Hanemann et. al., 2004, and Lew and Larson, 2005) and what influences where they choose to recreate once at a particular beach (Pendleton, 2001). This research furthers the understanding of the impact that water pollution has on beach recreation through the estimation of preferences coefficients for a diverse group of beach recreators. Preference estimates can be used to forecast and explain beach choice behavior conditional on beach attributes, such as water pollution. The ability to investigate the composition of these preference groups conditional of individual and season data is a useful tool for managers and policy makers in both the resource and public health arenas.

Second, this application contributes to the modeling of heterogeneity with the FML through the incorporation of a seasonal variable in the beach choice occasion preference membership function. This enables the analyst to capture seasonal variation in preferences for beach attributes. Other studies have not utilized the model to account for seasonal changes in site attribute preferences.

Lastly, this research represents a substantial step forward in the technique's empirical application. Previous applications have been applied to fairly restrictive choice sets, primarily modeling binary participation choice or multinomial choice for up to 6 options (Boxall and Adamowicz, 2002). In contrast, this application models recreator decisions among a choice set of 51 beaches using a revealed choice data set. This model specification utilizes 9 individual trip membership function variables consisting of seasonal, activity participation, and demographic variables. This marks a substantial increase in the number of parameters estimated relative to other applications in the literature (Provencher et. al. use three (2002), Boxall and Adamowicz use 6 (2002), and Shonkwiler and Shaw use 3 (2003)).

I generate statistical estimates of the welfare impact to beach users due to changes in water quality for beaches in Southern California. Accounting and controlling for preference heterogeneity is the key objective of this research. To this end, I utilize the finite mixture logit random utility model which allows for variation in preferences across individuals and seasons.

Comparison of the welfare estimation results from the competing models indicates that the FML model provides an important insight into the heterogeneity of individual's willingness to pay (mWTP) for improvements in water quality. The FML model's average estimate of mean mWTP is roughly 4.64 times that of the standard logit model's mean mWTP , while the RPL's mWTP point estimate of the mean mWTP is roughly 0.13 of the standard logit model's mean mWTP . However, the estimated mean mWTP for individual beach recreators ranges from zero or negative to 14 times the logit's mean mWTP , depending on the type of recreators and the type of quality change. For example, trips involving water-contact recreation during the winter by male college graduates who are working full-time are associated with a high value for an improvement in water quality. Trips taken during the summer by male college graduates who are not working full time and who do not enter the water are associated with a low value for an improvement in water quality.

The remainder of this dissertation is organized as follows. Section 1.2 and 1.3 further motivate the application and the model, respectively. In Section 2.1, I describe the base model framework. In Section 2.2, I describe the problem of heterogeneity in a discrete choice setting. I will then review the standard conditional logit (CL) framework and describe several econometric techniques that have been developed in order to account for heterogeneity. In Section 2.3, the finite mixture logit model and estimation strategies are discussed. Chapter 3 discuses the extension of the marginal value and compensating variation measure to the Finite Mixture Model. Chapter 4 describes the trip, beach site and recreator data. Details of the Southern California Beach Choice analysis are reported in Chapter 5. Section 5.3 reports on the estimation results for the CL, RPL, and FML models. Section 5.4 reports estimated welfare impacts for several beach closure and water quality degradation scenarios. Chapter 6 concludes.

1.2 Application Background

Coastal and marine health play an important role not only in the prosperity of the fisheries industry but also in the welfare of the communities which border the California coast and rely on the coastal environment for recreation and tourism.[4] Beach trips serve as a primary recreational activity for some and as a source of income for others.

However, coastal environmental resources are increasingly strained and affected by pollution and overuse. In California there were 6,568 beach closures and advisories in 2001. This was a 14% statewide increase from 2000 and marked the fifth consecutive year that beach closures and advisories have increased (NRDC, 2002). The public awareness of poor water quality is so widespread in the Los Angeles area that, in a focus group a few years ago, eight out of ten participants said that they do not go into the water when they go to local beaches (Hanemann, 2005).

The main cause of the beach closures in the Los Angeles area is storm water runoff. Although rainfall events are infrequent, when they do occur they generate a large volume of runoff from the streets, parking lots and other paved surfaces containing high pollution loads that bypasses sewage treatment plants. This runoff is discharged directly

[4] Lew and Larson estimate the mean value of a recreational beach day to be $28.28 (2004).

to the ocean through storm drains at or the near the local beaches. Storm water pollution is now coming under increasing regulatory pressure, but is extremely costly to manage. An engineering study conducted for the California Department of Transportation in 1998 estimated that to divert and treat flows from about 90% of the annual expected storm events in Los Angeles County would cost almost $54 billion, and the Los Angeles County Sanitation District revised this cost estimate to $65 billon. A 2002 report by engineers at the University of Southern California estimated the cost at $156 billion to cover 97% of the expected storm events (Gordon et. al., 2002). Other pollution sources include discharges from point sources, and sewage spills, such as the January 15th, 2006, 2 million gallon Manhattan Beach raw sewage spill caused by a sewage pumping station electrical outage. The resulting degradation of the coastal environment damages marine and coastal flora and fauna, and adversely affects the welfare and possible health of beach recreators.

The risk of becoming ill while swimming at the beach reduces the welfare of those who venture in the water and contract an illness,[5] and diminishes the welfare for those who forgo swimming because of the risk.[6] Public concern regarding this coastal degradation has prompted several studies focusing on the adverse health effects of coastal pollution and has generated the approval of several Legislative and Assembly bills (NRDC, 2002).

The differences in the values placed on beach recreation by different user groups can have important practical implications for beach management. For example, shoreline anglers care about different aspects of the beach recreation experience than surfers or mothers taking young children to the beach. What is considered an amenity to one may be an unwanted

[5]Rabinovici et. al. (2004) review the valuation of health status literature and report that Mauskopf and French (1991) estimate the WTP for government programs to aid in the avoidance of gastrointestinal symptoms at $280 for a 2-4 day case and $1,125 for a 5-7 day case.

[6]Walsh et. al. (1992) report a mean value per visitor day of recreational swimming at $35.60 ($2001)

nuisance to others. Moreover, the resource manager may be forced to make trade-offs in meeting the needs of different groups. The information developed in this research can lead to improved management of coastal and beach resources.

Coastal and fishery resource managers should find it useful to have welfare measures of the values associated with the alternative uses of harbors, piers, and docks since beach recreation values may swing the direction of the overall coastal management plan. The model developed through this research will facilitate the implementation of balanced and equitable resource management through the increased understanding of taste differences across users.

Society has to decide what resources to provide on its shorelines. Accurate welfare and usage estimates can serve as a useful tool to resource managers concerned with understanding the equity implications of specific policies.

An important practical issue is that different user groups value different characteristics of recreation sites and demand different services from them. It is highly useful to be able to account for the variation in preferences among different user groups. Furthermore, robust welfare and usage estimates are increasingly called upon in a litigation setting for the enumeration of damages caused by resource degradation. The development and implementation of methods and techniques used to capture and control for heterogeneity is the key objective of this research.

1.3 Modeling Background

A rich diversity of preferences among decision makers creates difficulties in terms of accurately modeling recreational site choice and estimating the economic value associated with a change in resource attributes. Diverse user groups often value different attributes of recreation sites and demand different services from them. If preference heterogeneity can be easily controlled by segmenting the sample population by a variable known to the analyst, a standard logit random utility model (RUM) can be used to estimate coefficients and welfare measures for each group separately. For example, beach recreators who swim in the ocean are likely to have different preferences for water quality and other beach attributes than those lying on the sand. However it is often unclear where to draw the line in defining sub-samples of the population.[7] This may lead to bias in welfare measures for changes in site attributes and hinder proper aggregation of welfare measures across individuals or time periods and adversely affect policy and management decisions.

The logit model handles variation in preferences by averaging over the individuals. In cases where the population is fairly homogenous in their preferences this may not cause a major problem; however, if the population is characterized by considerable systematic preference heterogeneity, the model's results may be misleading due to an averaging out effect over preferences from distinct groups. Additionally, the distribution of preferences over individuals or time is commonly lost due to the restrictive single point or even modal distribution which the model imposes on the data. The preservation of the preference distri-

[7] If the analyst differentiated between beach and water users there would still be heterogeneity within users. For example, among water users, surfers may care about different aspects of the beach recreation experience than mothers taking young children to the beach to swim. Although both of these groups likely view clean water as desirable, they may differ in the level of importance they place on water quality.

bution may aid analysts focused on the welfare changes between user groups, or to a specific type of user, due to changes in the attributes of the choice set. Similarly, an understanding of the temporal fluctuations in preferences could have important policy implications.

Suppose there are two different groups of beach users: novice and expert surfers, who prefer small and large waves respectively. Membership in either group is unobservable to the analyst, but may be statically correlated with observable demographic and seasonal data. Imagine further that there are several means of undertaking a coastal project which can have the secondary effect of increasing or decreasing the size of waves. Estimation with the standard logit model causes the opposing preferences for waves of the two user groups to be averaged out, resulting in model estimates that call for the medium sized waves, which are not preferred by either group. In contrast, a model that could statistically distinguish the two different types of users and estimate their separate preferences could lead to a policy whereby a variety of waves are maintained at specific beaches, resulting in a welfare improvements for both groups.[8]

The finite mixture logit (FML) model used in this paper accounts for systematic heterogeneity by sorting the agents into separate behavioral groups or latent segments, with different attribute preferences.[9] Within each latent segment, individuals are assumed to have homogeneous preferences. The segments are termed "latent" since individual membership in a particular segment, as well as the segments themselves, are not observable. The FML model simultaneously assigns an agent a probability of membership to each latent segment, and estimates the discrete choice probabilities for the random utility model.

[8]Note, the emphasis on the word statistically. In practice, preference groups are often not clearly defined into easily identifiable groups.

[9]This model was first proposed by McFadden (1986) and implemented by Swait (1994).

This approach captures the variation in preferences across the population through a discrete distribution with multiple probability masses. The model is distinctive in that it not only accounts for heterogeneity, but is able to explain the sources of that heterogeneity.

Chapter 2

Random Utility Models

2.1 Basic Framework

Random utility models have a long history as a powerful tool for resource managers. The random utility model is the standard statistical framework used to estimate the value of the change in consumer welfare due to an incremental change in the level of resource attributes in a setting characterized by consumer choice between several alternative recreation sites with varying attributes.[1]

Consider the utility maximization problem that an individual solves in relation to a recreation choice occasion between a set of J alternatives ($j = 1, ..., J$):

$$\underset{j}{Max} : u_i = v_i(M_i - C_{ij}, Q_j, Z_i) + \epsilon_{ij}. \tag{2.1}$$

Where $u(.)$ is a function of individual income, M_i, the travel cost of individual i visiting

[1]The Conditional Logit Random Utility Model (CL RUM) is a widely used research tool. An early application of this model to recreational choice application is Hanemann (1978). For technical discussions refer to Greene (2000) and Wooldridge (2002). For a discussion of the application of RUMs to environmental economics refer to Haab and McConnell (2003).

site j, C_{ij}, the quality and attribute mix of the chosen site, j, in the recreational choice set, Q_j, and individual socioeconomic characteristics, Z_i. The unobservable portion of utility is denoted by ϵ_{ij} and is assumed to be a random variable. The decision to recreate at a particular area is viewed as the decision to consume, or incorporate into one's utility function, the specific attributes that uniquely identify the chosen recreational site from others in the choice set. When individual i chooses to consume bundle j out of her choice set J, $j \subset J$, it is assumed that u_{ij} is the maximum of the J possible utilities in the choice set. The conditional probabilities of individual i choosing choice j can be derived as

$$\Pr_{ij} = \Pr(u_{ij} > u_{ik}) \; \forall \; k \neq j. \tag{2.2}$$

Maximum likelihood estimation can then be used to estimate the parameters of the indirect utility function (McFadden, 1973; and Bockstael, Hanemann, and Strand, 1986).

The outcome of an individual choice occasion, designated by Y_i, is a random variable. If and only if the disturbances associated with j, $\forall j \subset J$, are independently and identically distributed with the generalized extreme value distribution,

$$F(\epsilon_{ij}) = \exp(-e^{-\epsilon_{ij}}). \tag{2.3}$$

The choice probabilities are

$$\Pr_{ij} = \Pr(Y_i = j) = \frac{e^{\beta' \mathbf{\Gamma}_{ij}}}{\sum_{j=1}^{J} e^{\beta' \mathbf{\Gamma}_{ij}}} \tag{2.4}$$

where Γ_{ij} is a vector of individual and alternative specific variables (McFadden, 1973). This model is known as the conditional logit model (CL). This formulation of the CL model causes individual variables, Z_i, that do not vary over the choice set to drop out of the choice probability. The choice probability is then determined by choice specific variables.[2]

Due to the formulation of the CL model, it can be shown that individual variables, Z_i, that do not vary over the choice set drop out of the choice probability.

$$\Pr_{ij} = \frac{e^{\kappa'\mathbf{Q}_j+\beta'\mathbf{C}_{ij}+\gamma'\mathbf{Z}_i}}{\sum_{j=1}^{J} e^{\kappa'\mathbf{Q}_j+\beta'\mathbf{C}_{ij}+\gamma'\mathbf{Z}_i}} = \frac{e^{\kappa'\mathbf{Q}_j} e^{\beta'\mathbf{C}_{ij}} e^{\gamma'\mathbf{Z}_i}}{\sum_{j=1}^{J} e^{\kappa'\mathbf{Q}_j} e^{\beta'\mathbf{C}_{ij}} e^{\gamma'\mathbf{Z}_i}} = \frac{e^{\kappa'\mathbf{Q}_j} e^{\beta'\mathbf{C}_{ij}}}{\sum_{j=1}^{J} e^{\kappa'\mathbf{Q}_j} e^{\gamma\beta'\mathbf{C}_{ij}}} \tag{2.5}$$

The choice probability is then determined by choice specific variables. However, through careful construction of interaction variables that vary over both individuals and choice attributes individual specific information can be retained as an argument in the choice probabilities.

The parameters of the indirect utility function, $v_i(\cdot)$,can be estimated using maximum likelihood techniques.

$$\Pr_{ij} = \frac{e^{v_i(M_i - Cji, Q_j, Z_i)}}{\sum_{j=1}^{J} e^{v_i(M_i - Cji, Q_j, Z_i)}} \tag{2.6}$$

Which can be simplified as,

$$\Pr_{ij} = \frac{e^{v_i(Cji, Q_j)}}{\sum_{j=1}^{J} e^{v_i(Cji, Q_j)}} \tag{2.7}$$

Several econometric and modeling issues commonly arise with the Random Utility Model. Econometric consideration should be given to the independence of irrelevant alternatives property[3] and to identification issues surrounding the scaling parameter. In terms

[2]However, through construction of interaction variables that vary over both individuals and choice attributes, individual specific information can be retained as an argument in the choice probabilities.

[3]The analyst must take note that in the standard multinomial or conditional logit models the odds ratios for a specific pair of choices, $\Pr_j / \Pr_k$, is independent of the remaining alternatives. This property is known as independence of irrelevant alternatives (IIA). This property is fairly restrictive because in implies that

of modeling, the construction of the travel cost variable[4] and the formation of the choice set are major issues that have been the focus of considerable research.[5]

2.2 Econometric accounting of Heterogeneity

The economic value associated with a change in resource characteristics can vary over individuals due to the rich diversity, or heterogeneity, among individual decision makers. Heterogeneous preferences are difficult to account for in behavioral choice models due to the formulation of the conditional logit (CL) model which has historically been the base tool for random utility models . Within demand system models, the analyst can directly incorporate demographic, temporal, or other individual characteristic data directly into the individual's utility function to address preference heterogeneity. However under the specification of the CL, these characteristics drop out of the probability of an individual selecting a specific choice, thus preventing the direct identification of these characteristics in the model.

If heterogeneity is not accounted for, RUM estimates are characterized by bias

the relative probability of choosing between alternatives remains constant after the introduction of a perfect substitute of one of the alternatives to the choice set. Several models such as the nested logit and random parameters logit models have been developed, in part, as a solution to IIA (Haab and McConnell (2003)).

[4]The assumption that travel cost prices are exogenously determined deserves comment, as the endogeneity in prices assumption is one of the primary issues critiqued in the discrete choice literature (Berry, Levinsohn, and Pakes, 1995; and Nevo, 2000). However as discussed in Train (2003), this issue is not of great importance outside of market-level demand models. Within customer-level demand models it is assumed that individual demand does not affect price. Moreover, within the recreational demand literature the price associated with choosing a specific good is determined by the cost of travel to that location. One alternative is that the consumption of the good is of large enough proportion in the individual's utility function that the individual incorporates the location of the recreational site as an important argument in the residential location decision making process. Secondly, site characteristics to some degree all relate to visitation. For most site attributes individual trips do not affect the attribute level. However, some attributes, such as solitude, offered by the site are highly sensitive to small changes in the number of trips taken to the site. Assuming that individual residence location and travel infrastructure is determined exogenously, the travel cost price is exogenous. See Parsons (1991) for a discussion on housing location.

[5]For a thorough review on the optimal size of the choice set see Kurisawa (2003).

and lead to inaccurate forecasts pertaining to changes in resource attribute levels and management policies (Chamberlain, 1978, 1980; and Jones and Landwehr, 1988). This bias adversely affects welfare estimates for simulated changes in resource attributes and/or management decisions.

To address heterogeneity, researchers have primarily focused on structural approaches requiring the *a priori* selection of typically demographic or choice variables. In "cluster models" individuals are segmented into demographically homogenous/similar groups. An alternative method incorporates into the indirect utility function and interaction variable composed of individual demographic variables, such as income, race, and family composition and various choice attributes (Adamowicz et al., 1997). These methods are limited by the assumption that preference groups can be accurately determined *a priori* by demographic variables, and theoretical issues pertaining to the choice of an interaction variable (Boxall and Adamowicz, 2003). Other related solutions to this problem include the fixed effects and random effects specification of the conditional logit model (McFadden, 1986). However, these methods are difficult to employ when the heterogeneity structure is complex and the sample consists of a large number of decision makers.

An additional structural method, the Generalize Extreme Value (GEV) Logit (or nested logit) disaggregates the decision between alternatives into subsets of similar alternatives, relaxing the IIA restriction (McFadden, 1978). In the context of beach recreation, the GEV framework has been used to model recreational beach choice conditional on the type of activity engaged in during the beach visit (Hanemann et. al., 2004). The primary benefits of this approach are that the model may be useful in highlighting the differences in choice behavior and welfare estimates for different user groups, and that it is not restricted by the IIA property. However, the approach requires that the "nesting" rules are defined *a priori*.

Another approach, the random parameter model, controls for heterogeneity across preferences by allowing estimated coefficients to randomly vary across individuals according to a continuous probability distribution, typically the normal or log-normal. By allowing for variation in coefficients over people, the unobserved portion of the respondent's utility is correlated over sites and time (Train, 1997). To set up the single choice occasion RPL model, we begin with the standard logit choice probability for individual i and relax the standard assumption that preferences for all individuals are equal. Assuming that individual tastes vary in the population we can write the probability density function of the preference parameters as

$$f(\beta|\Theta^*),$$

where Θ^* are the true parameters of the distribution of the preference parameters. The actual probability that an individual i chooses choice j is the integral of the standard logit probability for all possible values of the preference parameter weighted by the density of the preference parameters:

$$\Pr_{ij} = \int \left(\frac{e^{v_i(Cji,Q_j)}}{\sum_{j=1}^{J} e^{v_i(Cji,Q_j)}} \right) f(\beta|\Theta^*)\mathrm{d}\beta.$$

The parameters are estimated by using simulation to evaluate the integral in the choice probabilities. The RPL can easily be extended to a multiple choice occasion panel data setting (Train, 1997). Additionally, the RPL model is not restricted by the IIA property due to interactions within the choice probabilities of the attributes of all elements in the choice set (Train, 2003).

The RPL approach has two weaknesses. First, it assumes that preferences vary continuously across economic agents. Second, it does not offer a behavioral explanation for the source of the heterogeneity across people. Although the continuous distribution assumption is likely to be valid in many applications, for example the spiciness that one likes their food, there are many situations where actual preferences may be more accurately captured by multiple discrete probability masses. Moreover, from a management perspective a coarse grouping of preferences may sometimes be more useful. For instance, the presence of motorized watercraft likely enters either positively or negatively into the majority of individual beach recreator's utility functions. Resource managers are often concerned with obtaining the best possible estimates for specific individuals or user groups relevant to policy and equity concerns.

2.3 Finite Mixture Logit Approach

An alternative solution is the finite mixture logit (FML), or latent segmentation (LS) approach which simultaneously account for heterogeneity and helps explain its sources. This approach was suggested in a RUM setting by McFadden (1986), and was implemented by Swait (1994). There has been a recent increase in the application of this approach, including several recreational choice models applications (Provencher et. al (2002), Boxall and Adamowicz (2002), and Shonkwiler and Shaw (2003). The FML approach is based on two important assumptions. First, individual preferences are neither homogeneous nor continuously distributed, but can vary between population segments which can be represented by discretely distributed multiple probability mass points. Second, individual preferences are not purely a function of demographic variables, but can also be formed by perceptions, attitudes, behavior, past experiences, and unobserved variables. A primary benefit of this approach is being able to explain the preference variation across individuals conditional on the probability of membership to a latent segment. The gained explanatory power should be of benefit to resource managers in terms of welfare analysis and policy decisions.

Each "latent segment" is composed of like-minded individuals with homogeneous preferences. The segments are termed latent because individual membership to a particular segment is not observable, nor are the segments themselves. The FML model simultaneously assigns the economic agent the probability of membership to each latent segment and estimates the discrete choice probability for the random utility model. This approach captures the variation in preferences across the population through a discrete distribution with multiple probability masses. The model is unique in that it not only accounts for heterogeneity, but is able to explain the sources of that heterogeneity. This is of particular importance in regards to management decisions where user groups may either be demographically homogenous or where there is little correlation between user group preferences and the standard demographic variables. The FML model can estimate the coefficients associated with the choice occasion for each latent segment's utility function. The FML model additionally estimates the composition of the latent segments and can be used to help researchers and managers understand the processes involved in the formation of behavioral groups. The ability to segment the sample population through the estimation of the latent segment type may aid resource managers with welfare analysis and management policy.

The FML RUM is an extension of the CL model, and follows the assumption that individual i's indirect utility is maximized on a choice occasion by selecting alternative $j \subset J$. The probability that alternative j is chosen is the probability that the utility gained from choice j is greater than or equal to the utility forgone by not picking one of the other alternatives in the choice set, J.

Under the assumption that there exists some degree of heterogeneity in preferences across the sample, let S be the number of segments that the population is to be grouped into.[6] Individuals are assumed to belong to a segment $s(s = 1, .., S)$ within the sample population. Individuals within a segment are assumed to be characterized by homogeneous preferences. Additionally, in all but the trivial case, $S = 1$, the probability ratio between any two alternatives includes arguments from all other alternatives in the complete choice set, J. It has been shown that in these cases the FML model is not constrained by the IIA property.(Shonkwiler and Shaw, 1997).

2.3.1 Single Choice Occasion

In a cross sectional data setting, the optimal solution to the recreational choice decision for individuals within a given segment s, is to maximize

$$u_{i|s} = v(\beta_s X_{ij}), \tag{2.8}$$

where the β_s vector is the coefficients representing individual preferences conditional on individual i's membership in segment s.

[6]The optimal choice of S is discussed below.

The parameter coefficients for a specific segment of the population are estimated using the following probabilities.

$$\Pr_{ij|s} = \frac{e^{v_{i|s}(-Cji, Q_j)}}{\sum_{j=1}^{J} e^{v_{i|s}(-Cji, Q_j)}}. \tag{2.9}$$

Consider a latent membership likelihood function M^* that assigns individuals to segment $s \subset S$ (Swait, 1994). Arguments to M^* can include variables associated with the unobservable tastes, attitudes, and preferences of the members of the group as well as socioeconomic variables represented by the vector $Z_{i.}$ Segments can be identified using standard demographic variables, behavioral and preference data. Assume the following equation:

$$M_{is}^* = \gamma_s' \mathbf{Z}_i + \zeta_{is},\ s = 1, ..S, \tag{2.10}$$

where γ_s is a vector of segment specific parameters and ζ_{is} represents the error terms.

The membership likelihood function, M^*, is a random variable. To use the function in an econometric model requires assumptions about the distribution of its error terms. Following Kamakura and Russell (1989), Swait (1994) and Boxall and Adamowicz (2003) the error terms are assumed to be independently distributed across individuals with Type I extreme value distribution. The probability of individual i belonging to segment s can then be calculated as

$$\pi_{is} = \Pr(M_i = s) = \frac{e^{\gamma_s' \mathbf{Z}_i}}{\sum_{s=1}^{S} e^{\gamma_s' \mathbf{Z}_i}}. \tag{2.11}$$

This probability is modeled as multinomial logit framework where the independent variables in this function vary over individuals, unlike the conditional logit where the variation is in the choice specific variables. Addressing an indeterminacy in the model caused by the lack of normalization the following restriction must be imposed:

$$\pi_{is} = \frac{e^{\gamma_s' \mathbf{z}_i}}{1 + \sum_{s=2}^{S} e^{\gamma_s' \mathbf{z}_i}} \text{ for } s = 2, ...S, \tag{2.12}$$

$$\pi_{i1} = \frac{1}{1 + \sum_{s=2}^{S} e^{\gamma_s' \mathbf{z}_i}}, \text{ and} \tag{2.13}$$

$$0 \leqq \pi_{is} \leqq 1, \text{ such that } \sum_{s=1}^{S} \pi_{is} = 1.$$

To model choice behavior under the assumption that the sample population can be grouped into finite segments, the researcher estimates individual i's utility maximizing choice between J alternatives conditional on membership to a specific segment, s. The joint probability $\Pr_{ins}$ that an individual i is a member of segment s, and chooses alternative j is defined as

$$\Pr_{ijs} = \pi_{is} \Pr_{ij|s}. \tag{2.14}$$

It follows that for a single choice occasion the probability of individual i choosing alternative j unconditional on segment membership can be written as

$$\Pr_{ij} = \sum_{s=1}^{S} \pi_{is} \Pr_{ij|s}. \tag{2.15}$$

Defining d_{ij} as and indicator variable that takes the value of 1 if an individual i chooses site j and 0 if not, allows the writing of the individual likelihood function as

$$L = \sum_{s=1}^{S} \left[\pi_{is} \left(\prod_{j=1}^{J} \Pr_{ij|s} \right)^{d_{ij}} \right]. \tag{2.16}$$

The individual likelihood function can be rewritten as

$$L = \prod_{j=1}^{J} \left[\sum_{s=1}^{S} \left(\pi_{is} \Pr_{ij|s} \right) \right]^{d_{ij}}, \tag{2.17}$$

and the cross section likelihood function as

$$L = \prod_{i=1}^{I} \prod_{j=1}^{J} \left[\sum_{s=1}^{S} \left(\pi_{is} \Pr_{ij|s} \right) \right]^{d_{ij}}, \tag{2.18}$$

which yields the log likelihood function

$$\ln L = \sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij} \ln \left[\sum_{s=1}^{S} \pi_{is} \Pr_{ij|s} \right]. \tag{2.19}$$

2.3.2 Segment Membership Time Consistency

The extension of the single choice occasion likelihood function to incorporate a time dimension utilizing panel data introduces a few complications in terms of the assumptions of segment membership independence across choice occasions. One assumption is that preferences are constant over time, although there is preference heterogeneity across individuals (Figure 2.1). A second modeling assumption is that preferences can be allowed to vary both over individuals and time (Figure 2.2).

Constant over Time FM Membership

In the constant over time framework individuals agents are modeled to be characterized by the same preference segment for all choice occasions. The constant over time assumption is most appropriate when the set of choice occasions are temporally close (such as multiple decision choice occasions), or when preferences and choice attributes are stable over time. This specification has been applied in both the marketing and transportation

Figure 2.1: Constant over Time FML Membership

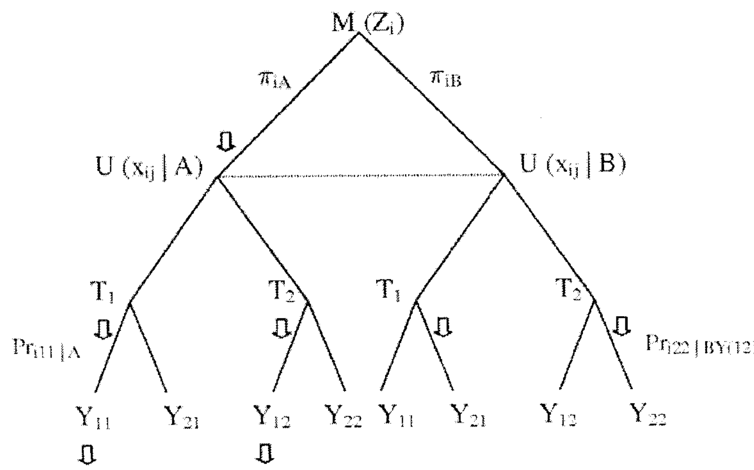
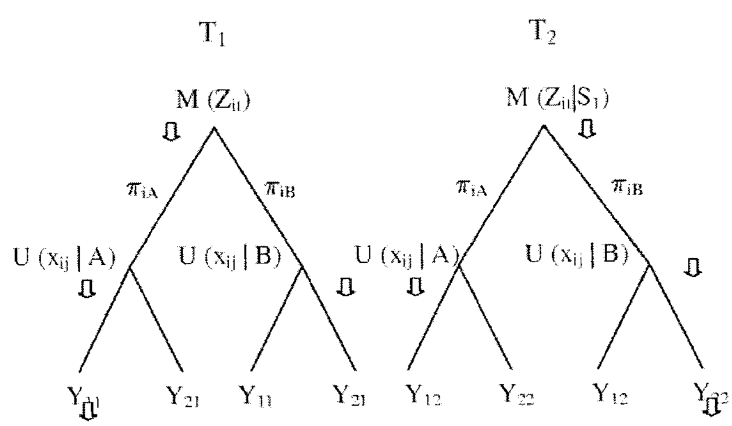


Figure 2.2: Variation over Time FML Membership



Note:

Indexes: i= individual, j=site choice, t=time

Segment membership types are denoted by S={A,B} where A & B are segment types

$Pr_{i11|A}$ indicates the probability of individual *i* choosing site 1 in time period 1, conditional on membership in segment A.

Choice outcomes, Y_{jt}, are indexed by choice then by time.

literature (Ramaswamy et al., 1999, Greene and Hensher, 2003). Following this assumption the probability of individual i choosing the set of alternatives j at each time t over the set T choice occasions is

$$\Pr_{ijt} = \sum_{s=1}^{S} \pi_{is} \left(\prod_{t=1}^{T} \Pr_{ijt|s} \right).^{7} \tag{2.20}$$

This assumption appears plausible in cases where the set of choice occasions are short in time duration, where the population segment's characteristics are constant over time, when the arguments of the segment membership function do not vary over time, and when there may be one choice occasion that is made up of several individual decisions.

The above probability gives rise to the likelihood function

$$L = \prod_{i=1}^{I} \left[\sum_{s=1}^{S} \pi_{is} \left(\prod_{t=1}^{T} \prod_{j=1}^{J} \left[\left(\Pr_{ijt|s} \right) \right]^{d_{ijt}} \right) \right]. \tag{2.21}$$

Yielding the log likelihood function

$$\ln L = \sum_{i=1}^{I} \ln \left[\sum_{s=1}^{S} \pi_{is} \left(\prod_{t=1}^{T} \prod_{j=1}^{J} \left[\left(\Pr_{ijt|s} \right)^{d_{ijt}} \right] \right) \right], \tag{2.22}$$

which can be written as

$$\ln L = \sum_{i=1}^{I} \ln \left[\sum_{s=1}^{S} \left(\frac{e^{\gamma_s' \mathbf{Z}_i}}{\sum_{s=1}^{S} e^{\gamma_s' \mathbf{Z}_i}} \right) \left(\prod_{t=1}^{T} \prod_{j=1}^{J} \left(\frac{e^{\beta_s' \mathbf{X}_{ijt}}}{\sum_{j=1}^{J} e^{\beta_s' \mathbf{X}_{ijt}}} \right)^{d_{ijt}} \right) \right]. \tag{2.23}$$

Note that in the above specification neither the segment type membership probability, π_{is}, nor the individual characteristic matrix include a time index, t. Therefore an individual i

[7] Note this can also be written as $\Pr_{i\varphi}$, where φ is a vector of length that represents the sequence of site choices over time T.

remains in the same segment regardless of exogenous temporal or individual characteristics. However this is not to say that individual choices are independent over time, as the $\Pr_{ijt|s}$ can also be conditional on the previous period's choice decision, $Y_{i(t-1)}$.

Variation over Time FM Membership

An alternative modeling specification, variation over time, can be useful as preferences often tend to vary with seasonal tastes as the underlying choice decision changes. This assumption is implemented in this paper and assumes that preferences can be allowed to vary both over individuals and time. Allowing for variation over time in preference membership relaxes the correlation between individual segment membership.

Seasonal variation in unobserved or unmeasured attributes necessitates the need to allow for seasonal variation in the segment membership function, allowing individual segment membership to change over time. For example, the surf in generally better in the winter and the weather is warmer in the summer. This may result in a winter preference set that gives high weight to water quality and surf variables, and a summer preference set that base choice on attributes that are important to sun bathers.

On a shorter time scale, allowing for variation between individual segment membership helps the model capture the correlation between segment membership between time periods. If a respondent was in a very active preference group during one period (swim), and they go to the beach the next day they are more likely in the second period to be in a more sedate group (lie on sand). In this case, segment membership is a function of both the previous segment classification and the time elapsed since the last choice occasion. Serial correlation of this type has been investigated in the marketing (Haaijer and Wedel, 2000) and recreational fishing literature (Provencher et al, 2002).

Write the probability of individual i choosing alternative j at time t as

$$\Pr_{ijt} = \sum_{s=1}^{S} \pi_{is} \Pr_{ijt|s}. \tag{2.24}$$

This leads to the likelihood function

$$L = \prod_{i=1}^{I}\prod_{t=1}^{T}\left[\sum_{s=1}^{S}\pi_{ist}\left(\prod_{j=1}^{J}\left(\Pr_{ijt|s}\right)^{d_{ijt}}\right)\right] \tag{2.25}$$

which simplifies as

$$L = \prod_{i=1}^{I}\prod_{t=1}^{T}\prod_{j=1}^{J}\left[\sum_{s=1}^{S}\pi_{ist} \Pr_{ijt|s}\right]^{d_{ijt}}, \tag{2.26}$$

and leads to the log likelihood function[8]

$$\ln L = \sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=1}^{J} d_{ijt} \ln\left[\sum_{s=1}^{S}\left(\frac{e^{\alpha\gamma_s'\mathbf{Z}_{it}}}{\sum_{s=1}^{S} e^{\alpha\gamma_s'\mathbf{Z}_{it}}}\right)\left(\frac{e^{\beta_s'\mathbf{X}_{ijt}}}{\sum_{j=1}^{J} e^{\beta_s'\mathbf{X}_{ijt}}}\right)\right]. \tag{2.27}$$

The above likelihood function has been utilized in both the marketing (Swait, 1994) and recreation (Boxall and Adamowicz, 2004) literature. Both applications utilized stated preference data where each respondent made a series of sequential choices from a structured choice experiment where all choice decisions are made at the same time, weakening the basis for the preference variation over time assumption. The basis of the FML is that decisions made by different members of the same preference segment will be more correlated than decisions made by members of different segments.[9] This holds true unless there is a

[8] Note the individual demographic variables, $\mathbf{Z}_{it}$, have a time index.

[9] This assumes that the information set and individual characteristics are constant across choice decisions.

mechanism for an individual's segment membership to change between choice decisions (Morey, 2003).

The correct time specification choice is dependent on the goals of the analysis and what data is used. As a general rule, the constant over time specification is appropriate for models over short time durations which do not utilize membership covariates that vary over time and where preferences are assumed to be constant. The varying over time specification better suits applications that seek to model FM membership as a function of seasonality, the effect of previous choices, or individual characteristics that vary over time (the decision to get into the water on a specific beach trip). This dissertation utilizes the varying over time specification, as individual preferences are expected to vary over time due to both seasonal effects and variety seeking throughout the survey year. It is noted that the constant over time model specification can be implemented by restricting the time varying individual characteristic variable parameters to zero. The consistency over time of estimated individual recreator segment membership is discussed in chapter 5 with the application estimation results

2.3.3 Additional Econometric Issues

Scale Parameter

In addition to attribute preference parameters, the variance of the disturbance terms may also differ across segments of the population. In the standard CL framework the analyst assumes that the unobserved factors have constant variance, hence utility is of the same scale across respondents. However, this restriction is not implicitly held in the

FML specification. Therefore FML model parameter estimates cannot be compared across segments directly. Researchers that do not take the differences in scaling parameters into account may incorrectly infer that the members of the segment with a larger coefficient estimate care about the attribute more than those individuals in the other segment. To properly interpret parameter results across segments analysts can compare the signs or ratios of parameter point estimates.[10]

Determining the Number of Segments

The appropriate number of segments is not identifiable in the FM class of models and is treated as exogenous. However, one can statistically test for improvements in the appropriate number of segments by estimating a series of models that iteratively increase the size of S. Improvements in model specification in terms of the number of latent segments in the population can be tested for through the use of McFadden's ρ^2, Bayesian Information Criterion, and Akaike Information Criterion test statistics. The use of traditional Likelihood Ratio tests in determining the number of segments should be used with caution as the regularity conditions are violated (Ben-Akiva and Swait, 1986, Jedidi, 1997, and Boxall and Adamowicz, 2003). In addition to the statistical tests, the analyst's judgment in regards to which model specification in terms of the number segments best describes the respondent population and addresses the relative policy questions should be applied.

Upon inspection of the FML model it is clear that through the selection of the appropriate number of segments the above model can mimic both the traditional CL and

[10] Alternatively, the scaling parameter can be normalized for one segment so that the variance of the disturbance term is the same across both segments. This leads to the identification of the scaling parameter (Train, 2003).

the RP models.[11] For instance, when $\gamma_s = 0, \beta_s = \beta, u_s = u, \forall s$, the FML reduces to the CL.

[11]In the present form FML is theoretically similar to the RPL where each respondent undertakes one choice occasion.

Chapter 3

Welfare Estimation Theory

The generation of welfare measures associated with a change in the attributes of the choice alternatives is a primary use of the RUM. The economic marginal value of site attributes and the compensating variation measure of consumer surplus associated with changes in site choice characteristics, such as water quality grades and other beach attributes can be calculated for each segment membership groups using model parameter estimates. The marginal value measure offers a readily assessable rule thumb welfare measure for changes in quality attributes. Whereas the compensating variation measure of consumer surplus takes into account the substitution patterns associated with a change in the choice set.[1]

The FML model provides a framework for the calculation of willingness to pay measures associated with changes in the choice set attributes using parameter estimates for each membership segment. The resulting willingness to pay calculations provide a detailed

[1]Note that an individual's income does not vary over the alternatives in the choice set, so this term drops out of the probability. However the relevant measure of income in regards to the choice occasion is the individual's total income less the cost of the utility maximizing choice.

estimate of the willingness to pay distribution. The CL with interaction terms model allows the researcher to more closely assign welfare measures to specific groups determined *a priori*. The probability of membership into a latent segment is a function of individual demographic variables. Welfare measures can be calculated for each individual by properly weighting the welfare measures of the representative consumer of each latent segment by the membership probability to each latent segment.

$$\Pr_{j^*}(i) = \frac{e^{v_i(-Cji,Q_j,Z_i)}}{\sum_{j=1}^{J} e^{v_i(-Cji,Q_j,Z_i)}} \tag{3.1}$$

3.1 Marginal Value Measure

Changes in welfare due to a marginal change in a given attribute can be calculated using the marginal mean willingness to pay measure (mWTP). This measure is defined as the maximum amount of income a person will pay in exchange for an improvement in the level of a given attribute provided and can be calculated as:

$$mWTP_i^* = \frac{\beta}{\gamma} \tag{3.2}$$

where β is the parameter on the attribute of interest and γ is the travel cost parameter. Both parameters measure the marginal utility of the object in question. This result can easily be applied using FML parameter estimates:

$$mWTP_{i|s}^* = \frac{\beta_s}{\gamma_s} \tag{3.3}$$

Because the degree of heterogeneity in preferences is assumed to be considerable in many recreational choice optimization problems, the ability to segment the changes in welfare over latent user types is important. However if the resource managers are interested in aggregate welfare measures over the sample, these can be calculated by adding up the welfare measures weighted by the latent segment probability (Boxall and Adamowicz, 2003).

$$mWTP^*_{ji} = \sum_{s=1}^{S} \pi_s \left[\frac{\beta_s}{\gamma_s}\right] \tag{3.4}$$

3.2 Compensating Variation Measure of Consumer Surplus

Changes in welfare due to the attribute/quality mix of the chosen bundles on one choice occasion can be calculated using the compensating variation measure and the estimated parameters of the indirect utility function (Small and Rosen, 1981; Hanemann 1982). This results in the per trip marginal change in welfare due to a decrease in some site attributes.

$$v_i(M_i - Cji - CV^*_{ji}, Q^1_j, Z_i) + \epsilon_{ji} = v_i(M_i - Cji, Q^0_j, Z_i) + \epsilon_{ji} \tag{3.5}$$

$$E[v^*(M_i - Cji - CV^*_{ji}, Q^1_j, Z_i) = E[v^*(M_i - Cji, Q^0_j, Z_i)] \tag{3.6}$$

$$CV^*_{ji} = \frac{\ln\left[\sum_{j=1}^{J} e^{v(\beta Q^1_j)}\right] - \ln\left[\sum_{j=1}^{J} e^{v(\beta Q^2_j)}\right]}{\gamma} \tag{3.7}$$

This result can easily be applied to using FML parameter estimated. Analysts interested in the welfare effect to specific groups can generate welfare measurements for an arbitrary change in choice set attributes for each latent segment.

$$CV^*_{ji|s} = \frac{\ln\left[\sum_{j=1}^{J} e^{v(\beta_s Q_j^1)}\right] - \ln\left[\sum_{j=1}^{J} e^{v(\beta_s Q_j^2)}\right]}{\gamma_s} \tag{3.8}$$

Because the degree of heterogeneity in preferences is assumed to be considerable in many recreational choice optimization problems, the ability of segmenting the changes in welfare over latent user types is important. However if the resource managers are interested in aggregate welfare measures over the sample, these can be calculated by adding up the welfare measures weighted by the latent segment probability (Boxall and Adamowicz, 2003).

$$CV^*_{ji} = \sum_{s=1}^{S} \pi_s \left[\frac{\ln\left[\sum_{j=1}^{J} e^{v(\beta_s Q_j^1)}\right] - \ln\left[\sum_{j=1}^{J} e^{v(\beta_s Q_j^2)}\right]}{\gamma_s}\right] \tag{3.9}$$

Utilization of the FML for welfare analysis provides an improvement over the traditional welfare calculation using the logit and RPL models. Choice attribute and membership variable coefficients can be used to estimate the appropriate welfare measures for each observation. The ability to construct distributions of the mWTP and CV measures as a function of individual trip characteristics is a useful research and policy tool.

Chapter 4

Data

The empirical choice model application utilizes an extensive recreational panel data set for recreational beach trips to 51 Southern California beaches (Table 4.1). The data come from a survey of households in Southern California. Respondents were asked to keep a diary of all their trips to beaches in Southern California from December 2000 through November 2001. The data consists of observation over a 12 month period for 4,642 beach recreation choice occasions of 595 beach recreators living in Southern California (Figure 4.1). Recreators include fishers, boaters, divers, surfers, sunbathers, runners, cyclists and other beach users. Beach recreator data contains demographic and behavioral data. An attribute data set contains individual beach attributes including water quality data and the travel times and distances between each beach and respondent residence.[1] The CL and RPL

[1]The complete data set consists of a screener and recruitment survey, 6 bi-monthly diary surveys, and 7 supplementary modules that focus on a variety of topics. The original data set comes from a random telephone sample of 1,848 respondents. Of these, 824 respondents were classified as non-beach users and 202 declined to take part in the survey The remaining 822 respondents agreed to be included in a large panel data set. Analysis shows that the demographics of the final sample is similar to those who declined to participate and therefore it is assumed that there is not a substantial amount of systematic self-selection bias. For a thorough discussion of the data see Hanemann et. al. (2003).

Table 4.1: Beach Sites in Study

1	San Onofre South	18	Bolsa Chica	35	Mother's
2	San Onofre North	19	Sunset	36	Venice
3	San Clemente State	20	Surfside	37	Santa Monica
4	San Clemente City	21	Seal	38	Will Rogers
5	Poche	22	Alamitos Bay	39	Topanga
6	Capistrano	23	Belmont Shores	40	Las Tunas
7	Doheny	24	Long Beach	41	Malibu (Surfrider)
8	Salt Creek	25	Cabrillo	42	Dan Blocker (Corral)
9	Aliso Creek	26	Point Fermin	43	Point Dume
10	Laguna	27	Royal Palms	44	Free Zuma
11	Crystal Cove	28	Abalone Cove	45	Zuma
12	Corona Del Mar	29	Torrance	46	El Matador
13	Balboa	30	Redondo	47	La Piedra
14	Newport	31	Hermosa	48	El Pescador
15	Santa Ana River	32	Manhattan	49	Nicholas Canyon
16	Huntington State	33	El Segundo	50	Leo Carrillo
17	Huntington City	34	Dockweiler	51	County Line

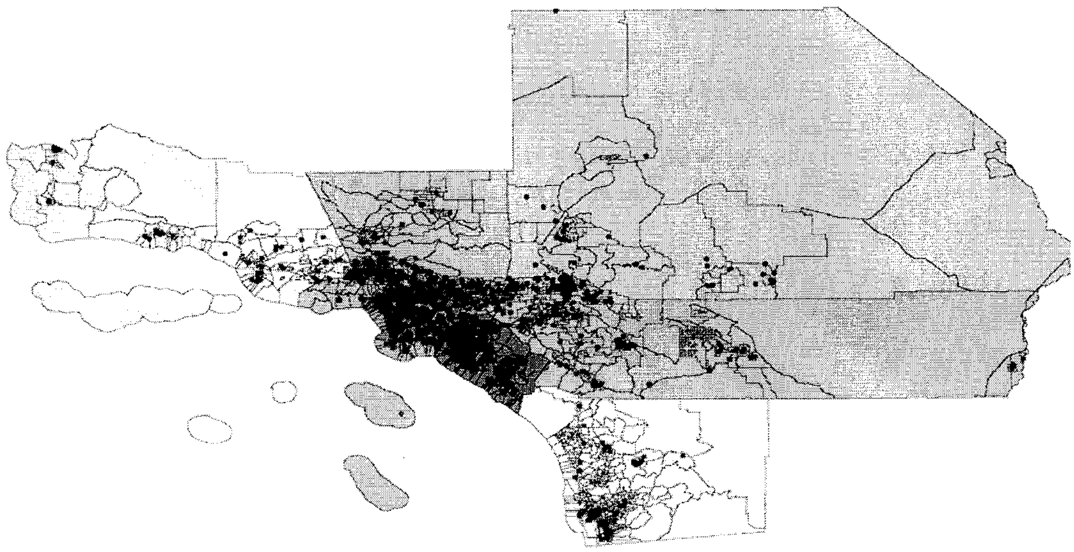
models are estimated using the same choice probability specification as the FML model.

Modeling individual site choices for beach recreation requires explanatory variables in terms of how the beaches in the choice set differ from one another. Beach attributes incorporated into the model specifications include beach location, water quality, presence of children's playgrounds, restaurants, tide pools, rest rooms, and foot or bike paths. The choice set for the complete panel consisted of 304 beaches, which were then aggregated into a set of 53 beaches. Properly defining the choice set is of great importance in model estimation. This is increasingly important when dealing with large choice sets. To help address these issues, respondents were asked questions to determine their familiarity and subjective quality opinions of the beaches included in the complete choice set. Summary statistics are included in Table 4.2.

Table 4.2: Respondent Beach Site Familiarity

	Beach	Obs.	Familiar? Number saying yes	Familiar? Percentage saying yes
1	San Onofre South	473	73	15%
2	San Onofre North	482	128	27%
3	San Clemente State	481	174	36%
4	San Clemente City	475	99	21%
6	Capistrano	484	185	38%
7	Doheny	483	162	34%
8	Salt Creek	484	76	16%
9	Aliso Creek	483	91	19%
10	Laguna	484	371	77%
11	Crystal Cove	483	133	28%
12	Corona Del Mar	484	235	49%
14	Newport	484	399	82%
16	Huntington State	484	339	70%
17	Huntington City	469	280	60%
18	Bolsa Chica	482	194	40%
19	Sunset	482	175	36%
21	Seal Beach	484	294	61%
22	Alamitos Bay	484	54	11%
23	Belmont Shores	484	191	39%
24	Long Beach	483	308	64%
25	Cabrillo	480	192	40%
27	Royal Palms	473	49	10%
28	Abalone Cove	473	69	15%
29	Torrance	482	117	24%
30	Redondo	484	350	72%
31	Hermosa	484	281	58%
32	Manhattan	484	262	54%
33	El Segundo	483	128	27%
34	Dockweiler	484	77	16%
35	Mother's	484	265	55%
36	Venice	484	352	73%
37	Santa Monica	482	382	79%
38	Will Rogers	482	147	30%
39	Topanga	482	92	19%
40	Las Tunas	479	51	11%
41	Malibu (Surfrider)	479	207	43%
42	Dan Blocker (Corral)	479	30	6%
43	Point Dume	479	74	15%
45	Zuma	479	130	27%
49	Nicholas Canyon	479	22	5%
50	Leo Carrillo	482	143	30%
51	County Line	482	85	18%

Figure 4.1: Respondent Residence Locations



Beach sites that had zero trips, and low name recognition are kept in the choice set. This decision is made based on the observation that these beaches are within close proximity to other beaches in the choice set that were visited. Actual beach choice often includes a degree of search; one may know the general area that they wish to visit, but their final choice is not made until a degree of "window shopping" is undertaken.

Each respondent is assigned a unique numeric identifier in order to link survey responses from all segments of the project and thus create a large panel data set. The screener and recruitment surveys collect standard socioeconomic household data, as well as beach and non-beach recreation data. Respondents were asked to keep a record of every Southern California beach trip in a bi-monthly diary throughout the survey period.[2] For

[2]Individual recreators frequented several beaches. 73% of all beach trips were to the recreator's most frequently visited beach.

Table 4.3: Probablility of Water Recreation by Season

	Trips	Trip %	Recreators	Recreator %	Avg Individual Seasonal %
Total	4,642	27%	595	23%	22%
Winter	987	14%	222	6%	9%
Summer	1,749	38%	378	62%	30%
Shoulder Season	1,906	23%	377	58%	22%

each trip, respondents were asked a series of trip details including the date of the trip, the specific beach they went to, the number of minors in their group, and information about up to four beach activities. Beach recreational activities are expected to be affected by seasonal variables. To control for this effect the data set is split into three time periods: winter (December and January), summer (June through September) and the remaining shoulder season months. Summary statistics on the seasonal distribution of trips, the probability of the average beach recreator's immersion rate, the percentage of trips that involves water immersion, and the proportion of recreators that enter the water are listed in Table 4.3.[3] Summary statistics on beach site trip counts are displayed in Table 4.4.

The implicit price of visiting each beach used in modeling is the travel cost construct. This construct is a function of the respondents reported income, and the estimated vehicle operational cost ($0.145/mile), travel time and the distance between the respondent's residence and each beach in the choice set.[4] One way travel distance and travel time between a respondent's address and the beach address are calculated using the computer

[3]Due to multiple site trips or inconsistencies among the screener, recruitment, and diary surveys 14.2% of the trip observations have been dropped from the dataset. Multiple site trips make up 3.9% of the dataset and have been excluded from this analysis due to complications in capturing the percentage cost of travel from one beach to another for the price matrix, and the proper weighting of beach attributes. Multiple site trips are commonly handled in the literature by assuming that they are independent trips.

[4]This cost is calculated as

Table 4.4: Seasonal and Water Recreation Beach Trip Counts

	Dry Trips by Season			
	Total	Winter	Shoulder	Summer
Min	0	0	0	0
Avg	67	17	29	21
Max	492	214	154	124
Std Dev	267	119	82	66
Total	3,409	850	1,474	1,085

	Wet Trips by Season			
	Total	Winter	Shoulder	Summer
Min	0	0	0	0
Avg	24	3	8	13
Max	208	32	66	110
Std Dev	114	18	36	60
Total	1,233	137	432	664

program PC-Miler. The time and distance data is transformed into the round trip travel cost of each trip, and is one of the model's primary explanatory variables.[5] See Table 4.5 and for average round trip costs to each beach recreation site.

Beach water pollution data is obtained from Heal the Bay, a Southern California non-profit group. This data contains weekly ratings on a scale of A+ to F for beach water quality for dry days at many monitoring stations throughout Southern California between June 1998 and April 2001. The A+ to F ratings are based on three biological pollutants measures: total coliform, fecal coliform, and enterococcus. The presence of these pollutant

$$Cost_{ij} = 2 * [one\ way\ travel\ dist * 0.145 + (one\ way\ travel\ time) * (0.5\ *\ hourly\ wage)] \quad (4.1)$$

[5]For a discussion on the percentage choice of wage rate in a travel cost model in a beach recreation application see Lew and Larson (2004).

are indicators of several illnesses such as stomach flue, ear infection, upper respiratory infection, and skin rashes. The calibration of the A+ to F scores are set at levels where a D score caused by a high fecal coliform ratio is associated with a water recreators having a 1 in 85 chance of becoming ill; and D water caused by enterococcus is associated with a 1 in 77 chance of becoming ill. Beaches that are rated as "Failing" with an F score caused by a high fecal coliform ratio are associated with a 1 in 20 chance of becoming ill (Heal the Bay, 2005).

Table 4.5: Beach Site Details: Cost, Water Quality, and Trips

	Beach	Avg. Cost	Avg. Water Grade	Observed Trips
1	San Onofre South	\$6.91	4.0	34
2	San Onofre North	\$8.83	3.8	40
3	San Clemente State	\$6.30	4.2	33
4	San Clemente City	\$6.09	3.0	36
5	Poche	\$5.69	2.0	1
6	Capistrano	\$5.49	1.4	17
7	Doheny	\$5.46	1.5	38
8	Salt Creek	\$5.43	4.1	70
9	Aliso Creek	\$4.98	3.8	17
10	Laguna	\$4.71	3.9	268
11	Crystal Cove	\$4.21	4.2	57
12	Corona Del Mar	\$4.07	4.0	116
13	Balboa	\$3.57	4.3	49
14	Newport	\$3.67	4.1	659
15	Santa Ana River	\$3.50	3.5	1
16	Huntington State	\$3.47	2.5	213
17	Huntington City	\$3.38	3.9	301
18	Bolsa Chica	\$3.26	4.0	206
19	Sunset	\$3.21	4.3	33
20	Surfside	\$3.21	4.2	2
21	Seal	\$3.18	3.3	240
22	Alamitos Bay	\$3.39	4.0	45
23	Belmont Shores	\$3.27	3.6	31
24	Long Beach	\$3.54	2.9	310
25	Cabrillo	\$4.04	3.0	52
26	Point Fermin	\$4.01	4.2	7

Table 4.5: Beach Site Details (continued)

	Beach	Avg. Cost	Avg. Water Grade	Observed Trips
27	Royal Palms	$4.04	4.1	13
28	Abalone Cove	$4.35	4.2	3
29	Torrance	$3.91	4.2	65
30	Redondo	$3.84	3.6	191
31	Hermosa	$3.74	4.1	249
32	Manhattan	$3.71	4.2	302
33	El Segundo	$3.85	3.8	6
34	Dockweiler	$3.78	3.7	16
35	Mother's	$4.11	2.5	76
36	Venice	$4.13	3.9	199
37	Santa Monica	$5.06	3.3	400
38	Will Rogers	$5.08	3.1	39
39	Topanga	$5.10	3.0	4
40	Las Tunas	$6.78	2.1	0
41	Malibu (Surfrider)	$8.43	2.1	58
42	Dan Blocker (Corral)	$9.81	4.0	10
43	Point Dume	$9.35	3.2	22
44	Free Zuma	$9.97	4.1	0
45	Zuma	$10.01	4.2	79
46	El Matador	$9.22	4.1	4
47	La Piedra	$9.22	4.1	0
48	El Pescador	$9.22	4.1	2
49	Nicholas Canyon	$9.65	4.1	2
50	Leo Carrillo	$9.62	4.1	20
51	County Line	$10.03	4.0	6

Three water quality variables are constructed utilizing this data: yearly average grade, bimonthly average for all years, and the bimonthly worst grade reported during the survey year (Mohn et. al., 2003). See Table 4.5 for average yearly water quality grades. In addition to these three measures a set of discrete water quality variables, indicating an F or D grade, were constructed. Table 4.6 reports summary statistics on the bi-monthly occurrence of water quality grade ratings, the bimonthly within beach variance for water

Table 4.6: Water Quality Grade and Variance

	Grade				
	F	D	C	B	A
Occurrence	7	8	22	62	207
% Occurrence	3%	3%	7%	17%	58%
Trips	53	329	483	1,542	2,232
% Trips	1%	7%	10%	33%	48%

	Variance						
	0	0.25	0.5	0.75	1	1.25	1.5
Occurrence	34	6	2	0	2	0	4
% Occurrence	71%	13%	4%	0%	4%	0%	8%
Trips	3,883	369	5	0	289	0	96
% Trips	84%	8%	0%	0%	6%	0%	2%

grades, and the number of trips taken by water quality grade and variance category.

To be included in the final data set a trip requires a valid destination and the respondent who took the trip must have supplied all of the demographic variables included in the model. This data source not only contains the necessary variables to implement the standard models, but also is rich enough in preference, choice set awareness, and past activity data to be able to implement the latent segmentation assignment of individuals.

Chapter 5

Application

5.1 Recreational Beach Choice Model

Following the literature, recreational site choice decision occasions are modeled using the discrete choice RUM as a function of site attributes, individual characteristics, and seasonal data holding the number of trips taken as exogenously determined. The CL, RPL, and FML variants of the RUM are estimated using an identical specification for the site choice probability. The FML model uses additional variables as arguments to the group membership function.

To capture the seasonal variation in preferences, a seasonal dummy is included into the segment membership function. Previous recreational modeling studies which have focused on trip temporal characteristics, such as season or part of the week, have operationalized the temporal data as an interaction variable or used it to segment the data set *a priori*. The use of the time variable in the FML enables the analyst to capture the probabilistic nature of seasonal influences on beach recreation in Southern California where there

Table 5.1: Composite Beach Variables and Their Components

Composite Variables		Component Variables
Developed Beach Very Developed	3 or more 8 or more	Street Access Public Transit Restaurants Stores Concessions Rentals Beach Clubs Houses Condos/Hotels Pier Concerts Volley Ball Tournaments
Wild Beach	1 or more	Pedestrian Access Only Rocky Tide pools Dogs Allowed
Ugly Beach	1 or more	Oilpumps Oilrigs PowerSewer Stormdrains

are often unseasonably warm and cold days during the winter and summer respectively.

Explanatory variables used in the RUM specifications can be categorized into beach choice and group membership variables. Modeling individual site choices for beach recreation requires explanatory variables in terms of how the beaches in the choice set differ from one another. Binary composite variables for development, very developed, wild, and ugly beaches serve to collapse twenty component attributes into four composite indicator variables (Table 5.1).[1]

Beach choice variables incorporated into the CL, RPL, and FML model specifica-

[1] The data set includes a large number of beach attribute variables (42) relative to the number of beaches in the choice set (51). Therefore, a composite choice variable strategy for the appropriate right hand side variables was developed in part to handle correlation within the beach attribute data set (Mohn et. al., 2003). The variables that are used to construct the composites are 0/1 indicator variables for the absence/presence of the relevant attributes. For a detailed discussion on the formation of the composite choice set, see Hanemann (2004).

Table 5.2: Choice Variable Summary Statistics

Choice Variables	Min	Mean	Max	Std Dev
Cost	3.183	5.546	10.027	2.375
Water Quality	1.373	3.602	4.333	0.757
Beach Length (ln)	-2.207	0.352	2.088	0.940
Developed	0.000	0.549	1.000	0.503
Very Developed	0.000	0.196	1.000	0.401
Wild	0.000	0.314	1.000	0.469
Ugly	0.000	0.275	1.000	0.451

Table 5.3: Correlation of Choice Variables

	Cost	Water Quality	Beach Length	Developed	Very Developed	Wild	Ugly
Cost	1	0.034	-0.347	-0.168	-0.215	-0.013	-0.329
Water	0.034	1	-0.099	-0.293	-0.009	0.094	0.038
Length	-0.347	-0.099	1	0.302	0.385	-0.137	0.108
Developed	-0.168	-0.293	0.302	1	0.448	-0.236	0.116
Very Dev	-0.215	-0.009	0.385	0.448	1	-0.227	-0.082
Wild	-0.013	0.094	-0.137	-0.236	-0.227	1	-0.132
Ugly	-0.329	0.038	0.108	0.116	-0.082	-0.132	1

tions include beach travel cost, water quality, the length of the beach, and a set of binary composite variables for capturing the developed, very developed, wild, or ugly nature of the beaches. Beach attribute summary statistics and correlation matrices are displayed in Table 5.2 and Table 5.3 respectively.

Group membership dummy variables used in the FML specifications indicate whether the trip occurred during winter, the recreator got in the water, the recreator is male, kids are present on the trip, the recreator is a student, the recreator works full time, and the recreator is a college graduate (Table 5.4 and Table 5.5).

The model specification reported upon in this paper is a preliminary specification designed to illustrate the level of heterogeneity which characterizes preferences for attributes that describe beach recreation site choices. The objective of this paper is to illustrate the

Table 5.4: Membership Variable Summary Statistics

	Min	Mean	Max	Std Dev
Constant	0.0	1.000	1.0	0.000
Winter	0.0	0.213	1.0	0.409
Summer	0.0	0.377	1.0	0.485
In Water	0.0	0.266	1.0	0.442
Male	0.0	0.561	1.0	0.496
Kids	0.0	0.266	1.0	0.442
Student	0.0	0.175	1.0	0.380
Work Fulltime	0.0	0.649	1.0	0.477
College Grad	0.0	0.534	1.0	0.499

Table 5.5: Correlation of Membership Variables

	Winter	Summer	In Water	Male	Kids	Student	Work Fulltime	College Grad
Winter	1	-0.404	-0.149	0.061	-0.079	-0.032	0.068	0.089
Summer	-0.404	1	0.201	-0.056	0.098	-0.017	-0.064	-0.011
Water	-0.149	0.201	1	0.082	0.064	0.010	0.003	-0.010
Male	0.061	-0.056	0.082	1	-0.210	-0.075	0.238	0.037
Kids	-0.079	0.098	0.064	-0.210	1	-0.014	-0.058	-0.125
Student	-0.032	-0.017	0.010	-0.075	-0.014	1	-0.148	-0.116
Fulltime	0.068	-0.064	0.003	0.238	-0.058	-0.148	1	0.134
College	0.089	-0.011	-0.010	0.037	-0.125	-0.116	0.134	1

importance of handling systematic preference heterogeneity in a discrete choice setting characterized by diverse user groups. Estimation results indicate that the FML model is a useful tool in analyzing Southern California beach choice recreational decisions. The choice model specification reported in this paper focuses on broad composite beach attribute variables and excludes several activity specific variables. Inclusion of these omitted variables is expected to impact the parameter and welfare estimates reported in this paper. Additionally, inclusion is expected to strengthen the preference group separation of the FML model due to an increase in the dimensionality of preference space.[2]

5.2 Estimation

The log likelihood functions for the three FML model specifications discussed above each have two major components: the segment membership probability, π_{is}, which is specified as a multinomial logit with individual attributes, Z_{it}, arguments; and the site choice probability,$\mathrm{Pr}_{ijt|s}$, which is specified as a conditional logit with site attribute, X_{ijt}, arguments.

Estimation of the preceding log Likelihood function using traditional derivative based maximum likelihood search algorithms can be troublesome. The non-linear nature of the likelihood function, and the exogenously determined number of segments, S, cause instability because the likelihood function is maximized on a ridge in parameter space if S is misspecified (Wedel, 1993). This is a common issue in the finite mixture model literature and a common solution is to implement the Expectation Maximization (EM) algorithm (see

[2]Whereas use of composite categorical data variables as a data reduction tool leads to a loss of information in the pattern of data over the attributes and respondents; as it is the pattern of data which allows the identification of latent segments (Ramaswamy, 1999).

Ruud, 1991 for a thorough discussion of the algorithm and Arcidicon and Jones (2003) for a recent application to finite mixture models).

Observation Weighting

Due to the unbalanced panel nature of the data, observation weighting can affect the estimation results A common approach in the literature is to weight each observation equally, however problems can arise due to the overweighting of the segments of the respondent population which have the most observations. An alternative approach would be to weight the observation of each individual by the inverse of the number of observations for that individual. Both of these approaches can be estimated and the results tested for robustness. Alternative weighting strategies can be researched in the choice avidity literature. This research will use the standard equal weighting approach.

5.3 Choice Model Estimation Results

Estimation of the CL, RPL, and FML models is implemented using numerical solutions with the GAUSS programming language and the Maxlik maximum likelihood software.[3] The CL and RPL model estimation is performed using the Newton-Raphson (NR) search algorithm and the FML is estimated using the Broyden-Fletcher-Goldfarb-Shanno method (BFGS) followed by the NR method. The model specification for beach choice variables is the base model specification from the preliminary report by the Southern California Beach Valuation Project (Hanemann et. al., 2004).[4] White's standard errors are

[3] Gauss code for the RPL is available on-line from Kenneth Train (2001).

[4] Additional model specifications for the standard logit and nested logit are analyzed in the Southern California Beach Project reports.

calculated for all regressions to correct for violations of independence between observations from a respondent.

The CL model parameter estimates are of the expected and plausible sign, except for the 'ugly beach' dummy parameter estimate. Parameter estimates for travel cost, and very developed are negative. Parameter estimates for water quality rating, beach length, and developed beach dummy variables are positive. Counter intuitively the ugly beach dummy variable coefficient is positive.[5] The wild beach dummy coefficient is negative and not statistically different than zero. CL model parameter estimates are presented in Table 5.6.

The RPL model parameter estimates are of the same sign as those of the CL model. This result is expected. However the coefficient estimate for water quality is negative and not statistically significant, and the wild beach dummy's coefficient estimate is negative and statistically significant. RPL model parameter estimates are presented in Table 5.6. As expected, the RPL has greater explanatory power than the CL model indicated by high pseudo R^2 and other test statistics (Table 5.8).

5.3.1 Finite Mixture Logit Segment Testing and Results

Model estimation using the FML specification allows for an increased focus regarding the heterogeneous nature of the sample population's preferences. The FML is estimated iteratively with an increasing number of preference segment groups per specification. For specification of the FML model, a complete set of beach attribute coefficients is estimated for each latent segment. Additionally, a set of probabilities for each segment is estimated

[5] This is likely due to an ommited variable.

Table 5.6: Parameters on Choice and Membership Variables

	Logit	RPL		FML-4			
Choice Variables		Mean	SD	Seg 1	Seg 2	Seg 3	Seg 4
Cost	-0.085	-0.182	0.109	-0.653	-0.021	-0.408	-0.366
	(-50.887)	(-34.016)	(23.921)	(-6.074)	(-11.919)	(-15.980)	(-10.415)
Water Quality	0.105	0.028^a	-0.007^a	-7.950	0.047^a	10.382	-0.637
	(4.316)	(1.008)	(-0.055)	(-4.395)	(0.852)	(10.673)	(-5.606)
Beach Length	0.470	0.567	-0.006^a	-0.871	0.259	2.160	0.814
	(18.627)	(19.184)	(-0.114)	(-2.320)	(5.166)	(9.73)	(7.508)
Developed	0.789	1.192	-1.885	1.422	0.527	1.998	-0.448
	(17.456)	(5.770)	(-4.317)	(3.541)	(5.200)	(11.693)	(-2.226)
Very Devlp	-0.097	-2.271	9.546	8.857	0.637	-6.347	1.836
	(-2.458)	(-2.728)	(3.252)	(4.482)	(5.746)	(-14.289)	(8.261)
Wild	-0.008^a	-0.662	2.200	-2.291	0.206	-6.253	1.706
	(-0.192)	(-4.040)	(7.537)	(-3.995)	(2.271)	(-7.288)	(8.754)
Ugly	0.073	0.100	-0.364^a	10.343	0.537	-8.461	0.748
	(2.122)	(2.186)	(-0.889)	(4.156)	(6.220)	(-13.663)	(6.369)

Segment Variables	Seg 1	Seg 2	Seg 3	Seg 4
Constant	-2.674	-0.788	-1.322	0
	(-8.781)	(-4.943)	(-8.871)	
Winter	0.299^a	-0.204^a	0.704	0
	(1.392)	(-1.295)	(5.162)	
Summer	1.195	0.282	0.516	0
	(6.006)	(2.058)	(3.994)	
In Water	-6.814	0.294	0.028^a	0
	(-2.119)	(1.962)	(0.205)	
Male	1.907	-0.129^a	0.855	0
	(7.321)	(-0.929)	(6.334)	
Kids	0.204^a	0.446	0.120^a	0
	(0.977)	(3.219)	(0.888)	
Student	0.083^a	0.313^a	-0.786	0
	(0.287)	(1.775)	(-3.823)	
Work Fulltime	-0.980	0.480	-0.309	0
	(-4.972)	(3.757)	(-2.597)	
College Grad	1.235	0.209^a	1.006	0
	(5.764)	(1.616)	(8.206)	

^{a} Indicates that the parameter is not significantly different than 0 at the 5% level.
(T-statistics)are calculated using White's standard errors.

Table 5.7: FML-4 Membership Probabilities

Individual Seg Probabilities	Seg 1	Seg 2	Seg 3	Seg 4
Min	0.0%	5.7%	3.2%	10.2%
Mean	10.6%	29.8%	27.2%	32.4%
Max	64.2%	70.4%	69.8%	55.9%
Seg. Membership By Max Probability	6.4%	25.1%	33.8%	34.9%
Water Quality mWTP	-$12.18	$2.19	$25.46	-$1.74

Table 5.8: Model Selection Statistics

Estimation Results Conditional Logit, Random Parameters Logit, and Finite Mixture Models[a]					
Model	Logit	RPLb	FML[c]		
Segments	1		2	3	4
LL at Convergence	-14014.08	-13380.74	-12863.50	-12317.03	-12066.10
Convergence LL at 0	-18251.55	-18251.55	-18251.55	-18251.55	-18251.55
Parameters	7	14	23	39	45
AIC[d]	28042.16	26789.48	25773.01	24712.06	24222.20
AIC-3[e]	42063.25	40184.23	38659.51	37068.08	36333.31
BIC[f]	14043.63	13439.84	12960.60	12481.66	12256.07
ρ^{2g}	0.232	0.267	0.295	0.325	0.339
mWTP			-$6.87 $18.40	-$7.66 $7.37 $21.03	-$12.18 $2.19 $25.46 -$1.74
Avg mWTP[h]	$1.23	$0.16	$5.64	$5.89	$5.71

[a]Sample size is 4642 choices from 595 individuals (N).
[b]RPL represents the random parameters logit model.
[c]FML represents the finite mixed logit model.
[d]AIC (Akaike Information Criterion) is calculated using -2(LL-P).
[e]AIC-3 (Akaike Information Criterion-3) is calculated using -3(LL-P).
[f]BIC (Bayesian Information Criterion) is calculated using -LL+[(P/2)*ln(N)].
[g]ρ^2 is calculated as 1-(LL)/LL(0).
[h]Average Willingness to Pay is a weighted average of the willingness to pay by segment, using estimates of segment membership. Weighted WTP ranges from -$4.06 to $17.66.

assigning segment membership as a function of the individual characteristics incorporated into the model.[6] The FML model is estimated for specifications with 2, 3, and 4 segments. Following the statistical segment testing methodology from the literature, the 4 segment model is chosen as having the greatest explanatory power. The 4 segment model (FML-4) has the highest R^2 compared to the CL, RPL, and 2 and 3 segment FML specifications. The 4 segment model also shows statistical significant improvements over the 3 segment model for several other test statistics: AIC, AIC-3, and BIC (Table 5.8). A 5 segment model is programmed in Gauss, but did not converge despite using a variety of parameter starting values and search techniques.[7] The lack of convergence with the 5 segment model signals that 5 segments is too many, as parameter estimates are known to tend towards negative and positive infinity when an $N + 1$ segment FM model is implemented on data which actually has N preference segments (Beard et. at., 1991).

The literature cautions against absolute reliance on statistical tests to determine the number of segments in a finite mixture and suggests the use of common sense (Beard et. al., 1991, McKachlan and Barford, 1988, and Boxall and Adamowicz, 2003). It is suggested that in most cases no more than 5 segments are needed in the FM framework (Heckman and Singer, 1984). The maximum number of feasible segments for a 7 dimensional preference space is 8 segments.

In terms of within sample forecast accuracy the 4 segment model outperforms the CL, and 2 and 3 segment models. Table 5.9 displays the percent of correct beach recreation site choice predictions for weighted segment membership and maximum probability single

[6]The number of segments minus one set(s) of segment membership function coefficients are estimated in order to account for the indeterminacy in the model.

[7]FML model for 1 to 6 segments are programmed and estimated with simulated data consisting of 1 to 6 preference segments.

Table 5.9: Within Sample Forecast Accuracy

Segments	Weighted Model % Correct	Point Estimate % Correct
1	15.2%	15.2%
2	20.6%	18.9%
3	21.7%	20.2%
4	23.6%	24.3%

segment membership. In both cases the 4 segment model predicts a larger number of trips correctly. Taking all of the above factors into account, I conclude that a 4 segment FML model is the best model.

The ability to construct the distribution of welfare estimates for the sample population is one of the primary benefits of the FML model. In the beach choice application each trip occasion is characterized by a constant and 8 individual and trip specific binary variables. This simple characterization of each trip by agent and seasonal characteristics results in 256 different probability assignments which are used to assign beach choice preference group membership to each choice occasion triple.

The 4 segment FML model estimates the probability that an individual is a member of each preference group conditional on the season of trip and individual recreator characteristics. Each individual choice occasion in the sample thus has a probability of being in each segment.[8] For some choice occasions the probability is high (up to 70%), while for others it approaches zero (Table 5.6). Segments 4, 3, and 2 are the most likely preference groups to characterize the largest number of beach choice occasions at 34.8%, 33.8%, and 25.1% of the total number of trip. Segment 1 is least likely preference group to characterize a choice occasion (6.4%) with the lowest mean percentage of group membership, 10.6%.

[8]Choice occasions are the individual recreator, water use, season triples that characterize each trip.

However it has a 64.2% probability of characterizing some choice occasions.

Summary statistics for the composition of the estimated segment membership in terms of trip and individual characteristics are displayed in Table 5.10. Beach trips that are estimated to be characterized by segment 1 preferences are 77% likely to occur during the summer and 45% likely to be taken by male beach recreators. However membership in segment 1 is the lowest out of all groups Estimated segment 2 preference type trips are 91% taken by beach recreators that are employed full time. Just under half of these trips are taken during the summer months and include water recreation. Segment 2 is characterized by summer trips, water use, kids on trips, female recreators, and full time employment. Segment 3 trips are characterized by male beach recreators that work full time. Winter trips, trips taken by male recreators, and those involving water recreation are most likely characterized by segment 3. Segment 4 trips are likely to occur during the shoulder season and have 66% male beach recreators. Trips taken by student recreators are likely to be characterized by segment 4 preferences.

Water Quality and Membership Consistency

Three specifications of the water quality variable are investigated: average yearly grade, monthly grade, and a dirty water dummy variable. The results of all estimated models are qualitatively robust, however the continuous yearly grade water quality variable specification is chosen over the competing specifications based on improved measure of fit, improved coefficient robust standard errors, and ease of convergence. Table 5.11 reports the log likelihood score at convergence of the CL, RPL, and FML models for the yearly and monthly grade specification. While it is noted that the competing specifications are

Table 5.10: Segment Membership Composition

Estimated Segment Composition by Membership Variable

		Segment 1	2	3	4	Total
Winter Trip	0	521	1217	1055	862	3,655
	1	3	38	578	368	987
Summer Trip	0	123	662	1109	999	2,893
	1	401	593	524	231	1749
In Water	0	524	696	1067	1,122	3,409
	1	0	559	566	108	1233
Male	0	287	1004	331	415	2,037
	1	237	251	1,302	815	2605
Kids	0	437	799	1253	919	3,408
	1	87	456	380	311	1234
Student	0	363	980	1629	857	3,829
	1	161	275	4	373	813
Work Fulltime	0	485	117	421	606	1,629
	1	39	1138	1,212	624	3013
College Graduate	0	232	781	33	1,118	2,164
	1	292	474	1,600	112	2478
	Total	524	1,255	1,633	1,230	4,642

Table 5.11: LL scores for Yearly and Month Water Quality

	Logit	FML-2	FML-3	FML-4	RPL
Yearly	-14014	-12864	-12317	-12066	-13381
Monthly	-14017	-13209	-12915	-12356	-13392

Table 5.12: Segment Membership Time Consistency

Number of Different Segments Per Individual					
	1	2	3	4	Total
Individuals	427	140	28	0	595
Trips Taken	2650	1394	598	0	4642

not nested, the log likelihood scores indicate that the yearly water quality grade variable provides an improved fit. This result indicates that beach recreators may base their recreational decisions based on impressions about water quality that are formed over many years as opposed to current information. Hanemann et. al. report a similar finding regarding GEV beach choice model estimation (2004).

Estimation results indicate that segment membership consistency is characterized by variation over time preferences. 28% of individuals accounting for 43% of the trips took trips that are characterized by more than one preference segment (Table 5.12). Individuals that are characterized by one segment type take an average of 6.2 trips, those that are characterized by two or three segment types take an average of 10 or 21.4 trips respectively. No beach recreator in the sample took trips characterized by all four segment types. Additionally, statistically significant parameter estimates on time varying attributes in the segment membership function indicate membership variation over time.

5.4 Welfare Estimates for Water Quality Changes

5.4.1 Marginal Value of Estimates

The average beach recreator in the sample has an estimated marginal value or willingness to pay (mWTP) of \$5.71 for a water pollution rating increase of one letter grade when estimated using the 4 segment FML specification. This FML estimate is 4.64 times greater relative to the CL specification estimate of \$1.23 (the mWTP measure for the RPL is \$0.16[9]). However, this valuation estimate ranges from negative to \$17.66 for individual beach recreators (roughly 14.35 times the CL mWTP measure).[10] See Table 5.6.

Latent groups 3 and 1, respectively, have the highest and lowest mean mWTP estimates for a one letter grade increase in water quality. With a mean mWTP point estimate of roughly 20 times the CL mWTP estimate, Group 3 membership is particularly likely for winter trips taken by male college graduates that work full time and do not have children accompanying them to the beach. Individuals with Group 3 preferences are likely to choose beaches that have long beach length, development, but are not very developed, wild, or ugly (Table 5.6 and Table 5.8).

On the bottom half of the mWTP distribution, Group 1 has a mean mWTP point estimate of roughly negative 10 times the CL mWTP point estimate. Trips that occur during the winter, where the respondent went into the water by recreators that work full time are less likely to be characterized by Group 1 preferences. Additionally, recreators that

[9]Note the parameter estimate on water quality is not statistically different than zero for the RPL model.
[10]Theoretically I expect that WTP is greater or equal to 0. However, a non-negativity constraint is not imposed during the process of estimation. In the case of RPL, although the RPL mWTP is positive, a portion of the distribution of the mWTP takes on negative values. In the case of FML, I believe the negative estimates of mWTP for Group 1 and 4 are likely due to an omitted variables bias, because the model fitted here does not include certain activity-specific beach characteristic variables that are expected to impact the parameter estimates.

are male, students, do not work full time, and are not college graduates are more likely to be characterized by Group 1 than Group 3. Those with Group 1 preferences are likely to choose beaches that are very developed, ugly, and have poor water quality.

The existence of multiple preference groups allows the construction of a multi-modal welfare distribution. A major strength of the FML approach is that the location within the distribution of specific welfare measures is recoverable conditional on individual and trip specific characteristics. The mWTP distribution for an improvement in water quality of one letter grade illustrates the heterogeneity in preferences for coastal water quality (Figure 5.1). Trips that occur during the winter, involve getting in the water, and are taken by male college graduates are associated with the representative groups that have a high valuation for an improvement in water quality. Conversely, trips taken during by students are strongly associated with representative groups with low mWTP for water quality.

5.4.2 Second Stage Estimated Marginal Value Regression

To analyze the relationship between the estimated mWTP for individual trip occasions and the group membership variables. The weighted estimated mWTP for each beach trip are regressed on individual and seasonal characteristics of the trip with ordinary least squares (OLS), and both cross-section and panel specifications of generalized least squares (GLS)(Table 5.13). Coefficient estimates for the winter trip, in the water, and college graduate variables are positive for all three estimators. The coefficient estimate for the student variable is negative for all three estimators.

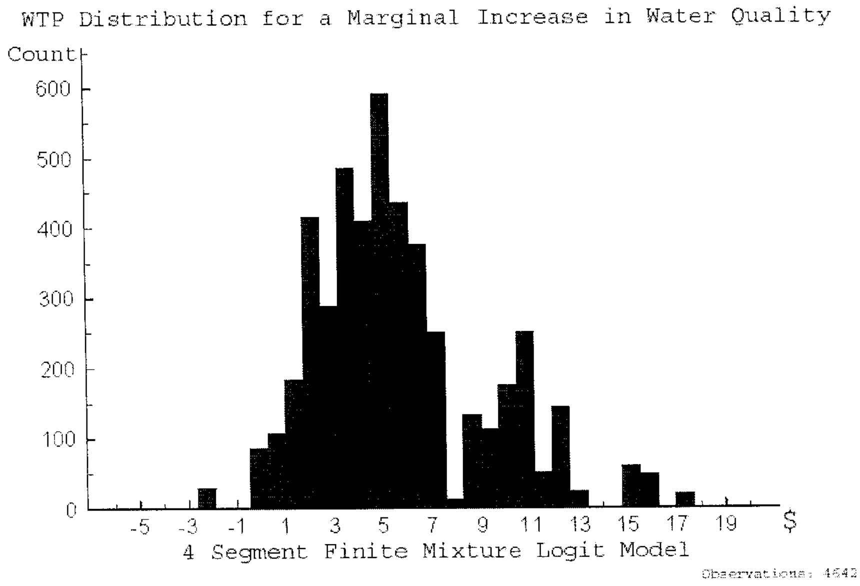
The coefficient estimate on the children present on trip variable is of particular

Table 5.13: Estimated mWTP Regression

Regressors from Group Membership Function	OLS		GLS		GLS panel	
Intercept	**2.451**	**2.366**	**2.680**	**2.691**	**2.294**	**2.234**
	(0.058)	(0.058)	(0.041)	(0.039)	(0.149)	(0.148)
Winter Trip	**3.803**	**3.780**	**3.640**	**3.746**	**4.120**	**4.043**
	(0.058)	(0.057)	(0.400)	(0.035)	(0.213)	(0.211)
Summer Trip	**0.138**	**0.129**	**0.464**	**0.324**	0.247	0.228
	(0.049)	(0.048)	(0.040)	(0.038)	(0.182)	(0.180)
Water	**3.626**	**4.016**	**2.241**	**3.023**	**3.941**	**4.273**
	(0.050)	(0.058)	(0.049)	(0.063)	(0.141)	(0.160)
Male	**1.085**	**1.057**	**1.478**	**1.365**	**1.050**	**1.028**
	(0.046)	(0.045)	(0.032)	(0.031)	(0.100)	(0.098)
Kids	**-0.145**	**0.237**	0.033	0.023	-0.165	0.218
	(0.050)	(0.058)	(0.038)	(0.029)	(0.110)	(0.142)
Student	**-3.656**	**-3.660**	**-3.355**	**-3.516**	**-3.654**	**-3.657**
	(0.057)	(0.056)	(0.045)	(0.045)	(0.125)	(0.123)
Work Fulltime	**-0.272**	**-0.260**	**-0.369**	**-0.263**	**-0.273**	**-0.261**
	(0.047)	(0.046)	(0.039)	(0.036)	(0.102)	(0.101)
College Graduate	**3.154**	**3.165**	**2.769**	**2.671**	**3.134**	**3.149**
	(0.044)	(0.043)	(0.033)	(0.031)	(0.096)	(0.095)
Kids Water		**-1.318**		**-1.119**		**-1.312**
		(0.105)		(0.088)		(0.312)
Regression Statistics						
R Sqr	0.828	0.833	0.855	0.874		
R Sqr-all					0.826	0.832
Adj R Sqr	0.827	0.833	0.855	0.873		
Obs	4642	4642	4177	4264	4642	4642

All coefficient estimates are significant at the 1% level.
Standard Errors are in parenthesis.
Bold indicates significantly different from 0, at the 1% level.

Figure 5.1: Marginal Value of Water Quality



interest. The OLS coefficient estimate for this variable is negative, whereas both GLS models produce coefficient estimates that are not significantly different than zero. However, the introduction of an interaction term for trips characterized by both the presence of children and getting into water produces negative and significant coefficient estimates for all three estimators. One would expect that beach trips that are taken with children and involve water recreation would have a higher probability of occurring at beaches with higher levels of water quality and would be associated with higher mWTP for water quality. One explanation for this result may be that the polluted beaches are characterized by features that are perceived by parents to provide safer environments for their children to swim, such as a lack of surf, but at the same time perpetuate water pollution. This result may be an example of cognitive dissonance and suggests the need for further research.

As illustrated by the paradoxical above result above the ability to construct the distributions of the relative importance which site attributes have on site choice is an important tool for resource and health officials charged with the management of resources used by diverse user groups.

5.4.3 Compensating Variation Simulation Estimates

The compensating variation (CV) measure can be used to estimate the welfare change or consumer surplus (CS) resulting in from a change in the composition of site quality attributes. The CV measure captures the substitution effects due to a change in the choice set; where as the marginal value measure (mWTP) illustrates welfare changes for a marginal change in one attribute holding all others constant. Consumer surplus measures are calculated for four hypothetical attribute scenarios to illustrate the difference in consumer surplus measures calculated based on logit and FML choice model estimates. The four scenarios are: A) the closure of Santa Monica and Venice beaches to all beach use; B) the closure of 13 popular beaches;[11] C) degrading the water quality at all beaches to a D score; and D) dropping the water quality at Newport, Bolsa Chica, and Manhattan beaches by one letter grade to roughly a B score.

The estimated change in consumer surplus for each of the four scenarios is displayed in Table 5.14 for the logit and FML models. The simulated closure of beaches result in an estimated loss in consumer surplus for all beach recreators, regardless of model choice or segment membership. As expected the simulated closure of additional beaches result in a

[11]These are: Laguna, Corona Del Mar, Newport, Huntington State, Huntington City, Bolsa Chica, Seal, Long Beach, Redondo, Hermosa, Manhattan, Venice, and Santa Monica Beach.

greater welfare loss.

The change in consumer surplus for degradation in water quality is negative on average for both scenarios C and D. However CS estimates for segments 1 and 4 for the 4 segment model are positive for both scenarios. The large CS gain for segments 1 and 4 are mathematically expected in scenario C, "D" grade water quality at all beaches. The cause of the negative welfare measure for segments 1 and 4 is likely due to omitted variable bias and not consumer preferences for poor water quality. In terms of water based recreation, segment 1 and 4 account for 0% and 9% respectively of trips. It follows that beach recreators characterized by segment 1 and 4 preferences will be less adversely impacted by a degradation in water quality.

Scenario D narrowly focuses and a degradation of one water grade, from roughly A to B, for three popular swimming beaches. Preference segments that are characterized by engaging in water based recreation have proportionally greater welfare changes than the preference segments that do not engage in water based recreation.

For the two beach closure scenarios the two competing models provide CV welfare measures of -\$0.95 and -\$7.91 (CL model) versus -\$1.16 and -\$11.96 (FML-4 model) for the simulated closures of 2 and 13 popular beaches respectively. For the two water quality degradation scenarios the two competing models provide CV welfare measures of -\$3.27 and -\$0.17 (CL model) versus -\$17.41 and -\$1.31 (FML-4 model) for the degradation of water quality to a 'D' grade for all area beaches and the dropping of one water quality grade for 3 popular swimming beaches respectively.

While the magnitude of the welfare loss generally increases with the number of

Table 5.14: Welfare Senarios: Beach Closure and Water Degradation

	A: Close Santa Monica and Venice Beaches				B: Close 13 Popular Beaches			
	Logit	FML-2	FML-3	FML-4	Logit	FML-2	FML-3	FML-4
Min	-$9.86	-$11.67	-$13.60	-$13.60	-$18.21	-$23.16	-$24.10	-$27.79
Mean	-$0.95	-$0.70	-$1.16	-$1.16	-$7.91	-$9.40	-$10.58	-$11.96
Max	$0.00	$0.00	-$0.01	-$0.01	-$0.40	-$0.44	-$2.40	-$3.03
	Segment 1 WTP				Segment 1 WTP			
Min		-$8.08	-$6.26	-$1.60		-$26.09	-$24.69	-$24.05
Mean		-$0.46	-$0.28	-$0.03		-$5.42	-$4.25	-$4.14
Max		$0.00	$0.00	$0.00		$0.00	$0.00	$0.00
	Segment 2 WTP				Segment 2 WTP			
Min		-$14.02	-$20.31	-$20.15		-$29.10	-$35.38	-$42.42
Mean		-$0.95	-$2.81	-$3.54		-$13.27	-$21.45	-$29.04
Max		$0.00	-$0.02	-$0.08		-$1.21	-$10.78	-$17.86
	Segment 3 WTP				Segment 3 WTP			
Min			-$21.72	-$23.11			-$25.49	-$27.64
Mean			-$0.43	-$0.39			-$6.27	-$6.85
Max			$0.00	$0.00			$0.00	$0.00
	Segment 4 WTP				Segment 4 WTP			
Min				-$11.50				-$18.38
Mean				-$0.48				-$3.15
Max				$0.00				$0.00

	C: Water Quality is 'D' at All Beaches				D: 3 Swimming Beaches Fall 1 Grade			
	Logit	FML-2	FML-3	FML-4	Logit	FML-2	FML-3	FML-4
Min	-$3.79	-$45.85	-$42.95	-$51.51	-$0.60	-$9.15	-$6.49	-$6.65
Mean	-$3.27	-$15.84	-$17.63	-$17.41	-$0.17	-$2.15	-$1.40	-$1.31
Max	-$1.57	$7.20	$2.62	$8.95	$0.00	$0.95	$1.63	$5.07
	Segment 1 WTP				Segment 1 WTP			
Min		$5.97	$4.71	$7.74		$0.00	$0.00	$0.00
Mean		$18.02	$19.96	$30.79		$0.36	$0.43	$0.86
Max		$21.41	$24.20	$38.83		$5.68	$7.31	$12.08
	Segment 2 WTP				Segment 2 WTP			
Min		-$56.70	-$20.98	-$6.08		-$11.86	-$1.22	-$0.35
Mean		-$50.20	-$19.62	-$5.73		-$4.57	-$0.55	-$0.16
Max		-$31.47	-$16.46	-$4.99		-$0.27	-$0.19	-$0.07
	Segment 3 WTP				Segment 3 WTP			
Min			-$66.76	-$80.86			-$12.60	-$13.26
Mean			-$61.73	-$75.28			-$4.60	-$5.14
Max			-$41.51	-$50.27			$0.00	$0.00
	Segment 4 WTP				Segment 4 WTP			
Min				$1.24				$0.00
Mean				$4.68				$0.11
Max				$5.57				$1.61

segment groups estimated in the model, it is noted that this does not always hold. For instance, for scenario A, the estimated welfare loss calculated with the logit model estimates is greater than that estimated using the FML-2 model, and the welfare loss for the FML-3 and FML-4 are the same. Likewise, for scenario D, the ranking of the models with the largest estimated welfare loss is {2, 3, 4, 1} segments, with the 2 segment model resulting in the largest estimated change in consumer surplus. Interestingly, the 'unordered' welfare estimates are observed in the two scenarios that model a small change in site attributes and not the two scenarios with greater attribute changes.

While the FML model has much strength, care must be taken to properly specify the utility model to avoid single preference segment estimates with omitted variable bias that can lead to biased welfare measures.

Chapter 6

Conclusion

Coastal water quality impacts recreation and tourism. Southern California beach recreators cite pollution as a primary reason for abstaining from swimming, a belief supported by studies linking swimming in polluted water with illness. While there is interest in understanding the impact of water quality on beach recreation to improve resource and public health management, this task is complicated by the diversity of user preferences and the multiuse nature of the beach.

This paper implements the FML RUM to highlight the importance of capturing preference heterogeneity. Exploiting an extensive beach recreational panel data set, this paper furthers the literature by applying the FML approach to model preference heterogeneity regarding the impact of an environmental variable related to health and seasonality on recreational choice. The application also increases the number of choice alternatives and the number of variables included in the segment membership function estimated with the FML model in the literature.

Application of the FML model to the Southern California beach recreation data set recovers 4 preference groups, highlighting the variation in the importance of water pollution on beach choice for a diverse sample of beach users. For these groups, the impact that water quality has on recreational site choice, as measured by the mean mWTP, ranges from negative to \$17.66, with an average of \$5.71, for an improvement in water quality of one letter grade. The mWTP estimate calculated with the CL model is \$1.23. The RPL coefficient estimate on water quality is not significantly different from 0, and yields a mWTP estimate of \$0.16. Compensating variation measures for consumer surplus associated with changes in beach attributes tell as similar story.

The FML approach facilitates the estimation of the distribution of water quality preferences and welfare measures across a diverse user-base, and enables researchers to identify user preference groups characterized by several variables. This increases the ability of resource managers to forecast the impact that changes in site characteristics will have on the beach choice and welfare across segments of society.

Estimation results indicate that recreators who enter the water have a higher estimated mWTP for water quality. Gender, employment, education, and seasonal variables are also important in estimating one's preferences. One troubling result of the model is the finding that the presence of children on beach trips which include water activity is not associated with a higher mWTP for improvements in water quality. This result highlights the model's ability to identify groups that resource managers and public health officials may desire to concentrate their educational outreach efforts.

The FML approach is likely to become increasingly important as diversity continues to grow, and the identification of user groups by a small number of variables becomes less feasible. The application of the model to a unique beach recreation data set is of major significance from the environmental management perspective. The powerful combination of being able to specify a model which simultaneously estimates the marginal benefits associated with different attributes for different groups and assigns group membership makes FML a particularly attractive model for policy analysis.

Bibliography

[1] Adamowicz, W.L., J. Swait, P.C. Boxall, J. Louviere, and M. Williams. 1997. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. Journal of Environmental Economics and Management 32:65-84.

[2] Andrews, R. and I. Currim. 2003. A Comparison of Segment Retention Criteria for Finite Mixture Logit Models. Journal of Marketing Research. Vol. 40. pp. 235-243.

[3] Arcidiacono, P. and J.B. Jones. 2003. Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm. Econometrica. Vol 71, no. 3, pp 933-946.

[4] Ben-Akiva, M. and J. Swait. 1986. The Akaike Likelihood Ratio Index. Transportation Science. 20: 133-136.

[5] Bockstael, N.E., W.M. Hanemann, and C.L. Kling. 1987. Estimating the Value of Water Quality Improvements in a Recreational Demand Framework. Water Resources Research. Vol. 23, no. 5, pp. 951-960.

[6] Bockstael, N.E., W.M. Hanemann, and I.E. Strand, Jr. 1986. Measuring the benefits

of water quality improvements in a recreational demand framework. Water Resources Research. Vol. 23, no. 5, pp. 951-960

[7] Boxall, P.C. and W.L. Adamowicz. 2002. Understanding heterogeneous preferences in random utility models: A latent class approach. Environmental and Resource Economics. 23: 421-446.

[8] Chapman, D.J. and W. Michael Hanemann. "Environmental Damages In Court: The American Trader Case," In Anthony Heyes' (ed) The Law and Economics of the Environment, pp. 319-367, 2001.

[9] Deacon, R. T., C.D. Kolstad. 2000. Valuing Beach Recreation Lost in Environmental Accidents. Journal of Water Resources Planning and Management, Vol. 126, No. 6, November/December 2000, pp. 374-381

[10] Dempster, A.P., N. M. Laird, and D. B. Rubin. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1. (1977), pp. 1-38.

[11] Ebert, U. 1998. Evaluation of Nonmarket Goods: Recovering Unconditional Preferences. American Journal of Agricultural Economics. Vol 80. pp. 241-254.

[12] Genius, M. and E Strazzera. December, 2000. Evaluation of Likelihood Based Tests for Non-Nest Dichotomous Choice Contingent Valuation Models. Working Paper. Centro Ricerche Economiche Crenos. www.crenos.it/ working/ working.html.

[13] Gordon, Peter, John Kuprenas, Jiin-Jen Lee, James E. Moore, Harry W. Richardson, Christopher Williamson. An Economic Impact Evaluation of Proposed Storm Water

Treatment for Los Angeles County. Los Angeles: School of Engineering and School of Policy, Planning, and Development, University of Southern California. November 2002.

[14] Greene, W.H. 2000. Econometric Analysis. Fourth Edition. Prentice-Hall, New Jersey.

[15] Greene, W, and D. Hensher. 2003. A Latent Class Model for Discrete Choice Analysis:Contrasts with Mixed Logit. Transportation Research Part B. Vol 37. pp. 681-698.

[16] Haab, T,C,, and K.E. McConnell. 2003. Valuing Environmental and Natural Resources: The Econometrics of Nonmarket Valuation. Edward Elgar Publishing, Massachusetts.

[17] Haaijer, R., and M. Wedel. 2000. Habit Persistence in Time Series Models of Discrete Choice. Marketing Letters. Vol. 12, pp 25-35.

[18] Haile RW, Alamillo J, Barrett K, et al. An Epidemiological Study of Possible Adverse Health Effects of Swimming in Santa Monica Bay: Final Report. Santa Monica, Calif: Santa Monica Bay Restoration Project; 1996.

[19] Hanemann, W.M. 1982. Applied welfare analysis with qualitative response models, Working Paper. No. 241. University of California, Berkeley, 26 pp.

[20] Hanemann, W.M., J.R. Hilger, C. Mohn, and L. Pendleton. Novermber, 2003. Data Reduction for Los Angeles Area Beach Site Attributes. Southern California Beach Project.

[21] Hanemann, W.M., L. Pendleton, C. Mohn, J. Hilger, K. Kurisawa, D. Layton, C. Busch, and F. Vasquez. August, 2004. Using revealed preference models to estimate

the effect of coastal water quality on beach choice in Southern California. Southern California Beach Project.

[22] Hanemann, W.M. Personal communication. November, 2005. Berkeley, California.

[23] Hausman J., D. McFadden. 1984. Specification Tests for the Multinomial Logit Model. Econometrica. Vol. 52. pp. 1219-1240.

[24] Heal the Bay. 2005. Beach Report Card: Grading System. www.healththebay.org/ brc/ gradingsystem.asp

[25] Heckman, J., and B. Singer. 1984. A Method for Minimizing the Impact of Distibutional Assumptions in Econometric Models for Duration Data. Econometrica, Vol. 52, No. 2., pp 271-320.

[26] Helmbold, D, R. Schapire, Y. Singer, and M. Warmuth. 1995. A comparison of new and old algorithims for a mixture estimation problem. Annual Workshop on Computational Learning Theory. Santa Cruz, California, United States. pp. 69 - 78

[27] Jedidi, K., H.S. Jagpal, and W.S. DeSarbo. 1997. Finite_mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity. Marketing Science. Vol. 16, pp 39-59.

[28] Kamakura, W., and G. Russell. 1989. A probabilistic Choice Model for Market Segmentation and Elasticity Structure. Journal of Marketing Research. Vo. 26, pp 379-390.

[29] Kamakura, W. and M. Wedel. 2004. An Empirical Bayes Procedure for Improving

Individual-Level Estimates and Predictions from Finite Mixtures of Multinomial Logit Models. Journal of Business and Economic Statistics. Vol. 22. pp. 121-125.

[30] Lew, D.K., Larson, D.M. 2005. Valuing Recreation and Amenities at San Diego County Beaches. Coastal Management, 33:71-86.

[31] McFadden, D. 1973. Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka, ed., Frontier in Econometrics. New York: Academic Press.

[32] McFadden, D. 1986. The choice theory approach to market research. Marketing Science. Vol. 5, pp. 275-297.

[33] Mohn, C., M.H. Hanemann, L. Pendleton, and D. Layton. 2003. Shifting Sands: A Preliminary Exploration of Seasonal Variation in Preferences for Beaches. Working Paper.

[34] Morey, E.R., R.D. Rowq, and M. Watson. 1993. A repeated nested-logit model of Atlantic salmon fishing. American Journal of Agricultural Economics. 75:578-592.

[35] Morey, E.R. 2003. 'Latent Segmentation Likelihood Functions.' Personal Communication.

[36] Naik, P.A, M.R. Hagerty, and C. Tsai. 2000. A New Dimension Reduction Approach for Data_Rich Marketing Environments: Sliced Inverse Regression. Journal of Marketing Research. Vol 37, pp 88-101.

[37] Natural Resources Defense Council. 2002. Testing the waters: A guide to water quality at vacation beaches, 2002. Volume 12, pp 24-57.

[38] Natural Resources Defense Council. 2005. Testing the Waters: A Guide to Water Quality at Vacation Beaches.

[39] Office of Environmental Health Hazard Assessment. June, 2001. California: Sport Fish Consumption Advisories. www.oehaa.ca.gov

[40] Parsons, G. and D. Massey. 2003. "A Random Utility Model of Beach Recreation" in The New Economics of Outdoor Recreation, N. D. Hanley, ed. Edward Elgar. pp 241267.

[41] Pendleton, L. 2001. Managing Beach Amenities to Reduce Exposure to Coastal Hazards: Storm Water Pollution. Coastal Management, 29:239-252.

[42] Provencher, B., K.A. Baerenklau, and R.C. Bishop. 2002. A Finite Mixture Logit Model of Recreational Angling with Serially Correlated Random Utility. American Journal of Agricultural Economics. Vol 84, pp 1066-1075.

[43] Provencher, B. and R. Bishop. 2004. Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study. Journal of Environmental Economics and Management. Vol. 48. pp. 793-810.

[44] Rabinovici, S.J.M., Bernknopf, R.L., Wein, A.M., Coursey, D.L., and Whitman, R.L. 2004. Economic and Health Risk Trade-offs of Swim Closures at a Lake Michigan Beach. Environmental Science and Technology. Vol. 38, No. 10, pp 2737-2745.

[45] Ramaswamy, V., R. Chatterjee, and S. Cohen. 1999. Reply to "A Note on Ramaswamy, Chatterjee, and Cohen's Latent Joint Segmentation Models." Journal of Marketing Research. Vol 36, pp 115-119.

[46] Richardson, S. and P. Green. 1997. On Bayesian Analysis of Mixtures with an Unknown Number of Components. Journal of the Royal Statistical Society. Series B. Vol 59. pp. 731-792.

[47] Ruud, P.A. Extension of estimation methods using the EM algorithm. Journal of Econometrics. Vol 49, pp 305-341.

[48] Scarpa, R., K. Willsi, ad M. Acutt. 2004. Comparing Individual-Specific Benefit Estimates for Public Goods: Finite Versus Continuous Mixing in Logit Models. Social Science Research Network Electronic Paper Collection. www.feem.it/ Feem/ Pub/ Publications/ WPapers/ default.htm

[49] Shonkwiler, J.S. and W.D. Shaw. 2003 "A Finite Mixture Approach to Analyzing Income Effects in Random Utility Models" in The New Economics of Outdoor Recreation, N. D. Hanley, ed. Edward Elgar, pp. 268-278.

[50] Small, K.A. and H.S. Rosen. 1981. Applied Welfare Economics with Discrete Choice Models. Econometrica, Vol. 49, No. 1. pp. 105-130.

[51] Sorensen, B. 1975. Energy and resources. Science. Vol. 189 : 255-60

[52] Swait, J. and J. Louviere. 1993. The role of the scale parameter in the estimation and use of multinormal logit models. Journal of Marketing Research. Vol. 30. pp 305-314.

[53] Swait, J.R. (1994) A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. Journal of Retailing and Consumer Services Vol. 1, no 2, pp. 77-89.

[54] Train, K. (1997) Recreation Demand Models with Taste Differences Over People. Land Economics. Vol. 74, No. 2, May 1998.

[55] Train, K. 2001. Discrete Choice with Simulation Gauss Code..www.elsa.berkeley.edu/ Software/ abstracts/ train0401.html

[56] Train, K. 2003. Discrete Choice Methods with Simulation, Cambridge. Cambridge University Press.

[57] United States Forest Service. National Survey on Recreation and the Environment: American's Participation in Outdoor Recreation. Southern Research Station.

[58] Vermunt, J.K. 2004. An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. Statistica Neerlandica. Vol. 58, nr. 2, pp. 220-233.

[59] Wedel, M., W.S. DeSarbo, J.R. Bult, and V. Ramaswamy. 1993. A Latent Class Poisson Regression Model for Heterogeneous Count Data. Journal of Applied Econometrics. Vol. 8, No. 4, pp 397-411.

[60] Wooldridge, J.M. 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Massachusetts.

[61] Yatchew, A. and Griliches, Z. 1984. Specification error in probit models. Review of Economics and Statistics. Vol. 66, pp. 134-139.

Appendix A

Data Reduction

A.1 Introduction

The Southern California Beach Project panel dataset is a unique dataset providing a wealth of information. In fact there are over 42 beach attribute variables in the dataset which can uniquely distinguish 51 beaches in the choice set of 53 beaches and estimate choice models. This abundance of data which is often seen as a blessing can also have its disadvantages. Whereas studies that lack this wealth of explanatory variables will often estimate models using all of the available variables, the variables used in the analysis for beach choice must be carefully chosen.[1]

Determining the appropriate set of explanatory variables for the discrete choice modeling becomes an enormous challenge. The validity of the estimation can be compromised by missing variables problems, on the one hand, but, on the other, including many variables that are highly correlated with one another can make the estimated coefficients

[1]This section consists primarily of the author's contribution to jointly written report prepared for the Southern California Beach Project (Hanemann et al. 2003).

unreliable and unstable. Lastly, the researcher needs to understand what it is that each particular site variable represents in order to use the fitted model in a meaningful and intelligent manner for the purpose of simulation and policy analyses.

These effects of these issues are further magnified in the context of FML estimation. Missing variables, both in the choice and segment probabilities, can hinder both the identification and estimation of latent segment members, and their choice and membership parameters. For example, the transformation of two binary variables into one discrete four class variable results in a loss of information pertaining to the relationship between the variables (see Ramaswamy, 1999, for a discussion on Identification in FM models). Likewise, issues associated with too many explanatory variables can cause further problems in the FML framework due to the expansion of the set of coefficients estimated per latent segment.

However, if the attributes are highly correlated or there exist combinations of attributes that are nearly linear; the coefficients the standard errors of the coefficients may be spuriously large and could be biased if one or more of the correlated variables is omitted from the model.

In order to address these issues, several methods of reducing the set of site variables to a manageable size for the purpose of model estimation. These techniques include: (1) the ad hoc selection of variables; (2) Principal Components Analysis (PCA); (3) a Composite Variable Approach. These approaches range from those that rely heavily on the researchers' expert opinion to those which rely almost solely on a statistical foundation. Two additional methods which may be useful include Cluster Analysis and Sliced Inverse Regression. The Beach Research Team is utilizing the Cluster Analysis data reduction methodology for estimation of nested logit specifications (Hanemann et al, 11-2003). For a discussion on Sliced Inverse Regression see Naik (2000).

The next section gives an overview of the data available and discusses some relevant characteristics of it and the transformations performed to the variables for different purpose of analysis. This is followed by a presentation of the data reduction approaches, considering a brief description of the method, the results and problem associated.

A.2 Project Data

A.2.1 Overview

The Southern California models beach going behavior to fifty-one mostly contiguous beaches in San Diego, Orange, Los Angeles, and Ventura County. In addition, we consider beach visits to beaches to the north and south of the study area. We use several sources of data to model beach choice behavior. The largest set of data on beach attributes includes 42 attributes of the beaches in the fifty one beaches of the study area (herein referred to as the beach attribute data). (We create site specific alternatives for beaches to the south of the study area and also for beaches to the north.) These data are supplemented with water quality information collected by Heal the Bay, and beach lengths compiled by the Berkeley research team.

The beach attribute data consist largely of binary variables indicating the presence or absence of a specific non-seasonal beach characteristic; whereas count variables measure the quantity or abundance of a resource present. The data are summarized in Table A.1.

In addition to data on beach attributes, the research team also compiled data on

Attribute Name	Description	Attribute Name	Description
Access_lot	1/0, has access from parking lot	Parking	1/0, presence of public parking
Access_ped	1/0, pedestrian access	Pier	1/0, presence
Access_street	1/0, presence of access to street	Playground	1/0, presence
Beachclubs	0-3, number of beach clubs	Powersewer	1/0, power or sewage plant visible from beach
Bikepath	1/0, presence of bike path adjacent to beach	Pubfac	1/0 presence of public facilities
Camping	1/0, campgrounds or RV parking	public transit	1/0, presence of public transit stop
Concerts	1/0, concerts sometimes held at beach	Rentals	1/0, bike or skate rentals available
Concessions	0-50, # of concession stands at beach	Restaurant	0-18, # of beachside restaurants
Condoshotels	1/0, condo or hotel visible from beach	Restrooms	0-20,# of restrooms
Diving	1/0, diving allowed	Rivers	1/0, river or creeks flows through or abuts beach
Dogsok	1/0 dogs allowed	Rocky	1/0, beach is rocky
Firepits	0-261, # of firepits	Sandy	1/0, beach is sandy
Fishing	1/0, fishing allowed	Showers	1/0, presence
Harbor	1/0, presence	Sidewalk	1/0 presence of sidewalk adjacent to beach
Houses	1/0, presense of beachside houses	Stores	0-3, # of beachside stores
Isleview	1/0, view of Catalina or Channel Islands	Stormdrains	0-77, # of storm drains
Lifeguards	0-24, # of lifeguard towers	streetparking	1/0, parking along street near beach
Marina	1/0, presence	Surfing	1/0, surfing at beach
Nature	1/0, abuts natural area	Tidepools	1/0, presence
oil rigs	1/0, off-shore oilrigs visible from beach	Vballnets	0-107, # of permanent volleyball nets
Oilpumps	1/0, on-shore oil pumps visible from beach	Vballtourney	1/0, volleyball tournaments held at beach

Table A.1: Beach Attribute Variables

beach water quality, based on data provided to the public by the not for profit Heal the Bay. The Heal the Bay (HTB) water quality data consist of site-specific letter-grade ranking of bacteriological water quality, measured at numerous data collection points in the study area. These collection points were mapped to the beaches used to define the sites in the beach attribute data set. Water quality grades were collected over several seasons, however the number of observations available varies both by beach and over time due to irregularities in sampling frequencies. Further, HTB data were collected for both wet periods (immediately after a rain) and dry periods; separate wet and dry HTB grades are made available to the public. >From the HTB data, we constructed composite dry-weather grades for each beach in the study area based on averages across all corresponding HTB observations – we constructed an overall average as well as averages and minimums across all years for all observations in the months corresponding to each wave. This gave 3 measures of quality for each beach – one which is constant across waves and 2 which vary. There were not sufficient wet-weather grades to construct averages for all beaches in all waves.

The research team also used GIS techniques to estimate data on the length of each beach. Because the size of a site influences the probability of randomly choosing a site (even when a visitor does not have preferences for any site attributes), it is generally necessary to include the natural log of length in a correctly-specified model. This is necessary to allow quality measures to behave correctly under aggregation.

The beach study examines 53 beaches. However, the extreme northern and southern beaches are composites not represented in the USC dataset, so their quality is represented only by an alternative-specific constant.

A.2.2 Candidate Reduction Variables

The first step in the data reduction process involved determining the scope of the data reduction. The principals involved in this stage of the analysis are continuously applied throughout the entire data reduction process. The first step of the data reduction process is to categorize the attribute data into groups. In this section we describe our approach to categorizing variables and even the qualitative transformation of variables. Not all of these categories or transformations will be used in our final models. Nevertheless, this processes is discussed because it illuminates the evolution of our thinking concerning the final choice of explanatory variables that are used in our model of beach choice.

Policy Variables

Variables that are thought to be of primary importance for policy analysis would be excluded from the candidate list for inclusion in the data reduction processes. The primary policy variables were considered to be:

Travel Cost   Water Quality

Other variables with possible policy implications were included as candidates for data reduction but with a careful eye balancing the trade-offs between losing the variable as a policy variable and gaining a more manageable data set. These secondary policy variables included:

Bike path	Playgrounds	Near Nature Area
Tide pools	Surfing	Diving
Fishing	Volleyball Nets	Volleyball Tournaments

It will turn out that many of these variables will be incorporated into attribute composites during latter stages of the research.

Interaction Variables which affect activity choice

In selecting these variables consideration was given to how the beach site choice of individuals was related to their choice of activity and the presence of the appropriate amenities at that site. However, the researcher must keep in mind that an individual bases their decisions on attributes that are directly utilized for their activity, but also amenities that are indirectly utilized or that are pointedly avoided. For example, an individual not only cares about playgrounds or 'dogs being allowed" if they have children or dogs, but also if they don't want to be around either. Another example of this indirect interaction occurs with the surfing variable. One of the activities that people report taking part in is "watching surfers.' If this is true it would be incorrect to limit the use of the surfing variable as an interaction variable with solely actual surfers. Out of the 53 beaches studied 12 beaches not only have no surfing, but also no "surfer watching."

The variables that fall into this category are:

Bike path	Surfing	Diving
Fire Pits	Pier Fishing	Camping
Lifeguards	Marina/harbor	Dogs Allowed
Playground		

Parking

All beaches in the sample either have street or lot parking, this results in little variation in the parking variables. Virtually all lots are pay parking; 7 of the 52 beaches have free parking (not metered or fees). However there is no reliable way of capturing the variation in access costs for all beach trips. There are multiple parking options for the majority of the beaches in our data set. Sources of the variation stem from what day of the week it is, what season it is, what the weather is like, how far one is willing to walk, and how lucky one is in getting one of the few public street parking spaces.

Beach Length

Due to the large variation in the size of the beaches in the study area, a preliminary length variable for the usable portion of the beach was incorporated into the data reduction process. Where as the attribute level of specific presence/absence variables is invariant to beach length normalization, incorporating the length of the beach into several continuously measures attributes was suspected to be important. For example, it is likely more meaningful to use the variable "rest rooms per mile" compared to the count variable "rest rooms." Apart from its use in normalizing continuous variables, we also believe that length may be an important variable in explaining beach choice.

A.2.3 Scaling

Although a majority of the beach attribute data are binary data, several attributes are characterized by count or continuous variables. Some of the variables for which we have count data are: beach clubs, beachside restaurants, concession stands, fire pits, lifeguard stations, public rest rooms, storm drains, and volleyball nets.

Several scaling strategies are applied to the data to both better capture the way in which beach and water quality attributes are experienced by beach goers. These scaling strategies also reflect competing requirements of the statistical methods we employ for data reduction. These strategies include:1) keeping the data in its raw form (a mix of binary presence/absence variables and continuous count and ordinal qualitative ratings); 2) normalizing the count variables by beach length while maintaining the raw data for presence /absence and qualitative variables; 3) transforming all non-binary and non-policy (travel cost and water quality) variables into mean zero, standard deviation one variables; and 4) transforming the non-binary and non-policy variables into binary presence/absence variables for specific attribute levels.

In several cases the correct normalization strategy for variables was not clear from an intuitive viewpoint. For some variables the relevant question appeared to be whether or not the attribute was present or absent at the beach in question, whereas for other variables the relevant question is the level of density of a specific attribute. As an example, lifeguard towers and beach clubs are uniformly distributed over beaches where restaurants, concessions, and rest rooms are typically clustered into specific areas.

The transformation of variables from count variables to binary variables also requires substantial judgment in determining the threshold levels of importance. For example how many rest rooms, restaurants, or fire pits are enough in order for the attribute to be adequately measured by binary variables? To address this problem the Beach Research team asked two major questions: 1, intuitively how would the variable be interpreted; and 2, what is the distribution of the count variables.

A.3 Data Reduction Strategies

A.3.1 Ad Hoc Data Reduction

The simplest means of dealing with highly dependent variables is to eliminate variables. Unfortunately, when attribute lists are large, multi-collinearity may not be obvious; dependencies in the explanatory variables may not be pair wise, but may result from linear combinations of many variables. Deciding which variables to eliminate by judgment alone can be difficult if not impossible. The elimination of important variables can create omitted variable biases. Even worse, remaining variables may act as proxies for the omitted variables and the interpretation of the estimated coefficients may no longer be straightforward. If all attribute variables we increase the risk of collinearity problems. While ad hoc reduction of the number of attribute variables can lead to an omitted variables problem. Ad hoc data reduction can be used as an intuitive basis for the dataset creation.

A.3.2 Intuitive Approach

The first step of the data reduction process involved building a believable theoretical model of what was driving beach recreation behavior. Although it is expected that all of the attribute variables in the dataset enter the beach recreation decision process for some sub sample of beach goers, the research team is trying to build a model that captures the behavior of beach goers in the sample. With this in mind, the data reduction process starts with removing those attribute variables that are judged to be unimportant to the sub sample represented in the data set.

Noting that the analysis is limited to focusing on single day trips one possible candidate for data reduction is the "camping" attribute. Other variables that are candidates for removal include the "fishing" variable. This variable does not provide much information as there are only one or two places that prohibit fishing. The "diving" variable also is rather meaningless since that is not based on "hard evidence" that people actually dive at these sites, but rather that one is permitted to dive. Additionally, it should be noted that "diving" was only reported as an activity once.

The "oil pumps" variable is unlikely to be of significant value, as all oil pumps captured by the attribute variable are located on the shore side of the Pacific Coast Highway and are not clearly visible from the beach. Additionally, all beaches that are characterized by "oil pumps" also have other attributes that can be characterized by being eyesores.

A.3.3 Correlation Based Approach

The dataset contains 42 variables describing non-seasonal attributes of 51 beaches used in our analysis. This is obviously problematic, since if these variables are not linear combinations of each other, then any complex patterns of correlation between them will spuriously bias the coefficients of any models estimated, and possibly greatly misstate the impact of policy changes. The potential error arises from the nature of economic choice models, and is likely to be worse as the number of covariates approaches the number of choice alternatives.

As an example of this consider the two variables marina and harbor. For obvious reasons there is a great correlation between a beach being in a harbor and a beach being near a marina. If there were only a few beaches which were in harbors but not near a marina, then any idiosyncratic deviation in popularity of these beaches would be attributed to the harbor variable, and the effect of a harbor would be captured in the marina variables coefficient. As sites can be more closely identified by a linear combination of the variables, the site-specific (as opposed to attribute-determined) aspects of beach popularity are increasingly attributed to the attributes, rather than being averaged out in the error terms of the model. With 42 variables and 53 sites, the potential for erroneous modeling is large.

In order to address issues of variable correlation, we mechanically constructed logit models for the 0/1 variables using a step-wise-determined subset of the other variables as RHS. In many cases the models were inestimable when more than 5 or 6 RHS variables were added, because either 0 or 1 outcomes were completely determined by the RHS variables, which leads to an infinite coefficient. This indicates that there were strong relationships among the variables. When we have few potential RHS variables, correlations are useful because there is unlikely to be much redundancy in the data and so simple measures catch most of it. However, when we have too much information, complex relationships almost certainly exist, and simple tools can catch only the simplest of problems.

Out of 44 site attributes there are only two cases where the correlation coefficient exceeds 0.7. One is Marina & Harbor, where the correlation coefficient is 0.731. The other is Restaurants and Lifeguard, where the correlation coefficient is 0.713. There are no obvious dependencies among variables that can usefully serve as a basis for a preliminary reduction of dimensionality prior to the application of more formal or more systematic variable reduction/grouping.

However the issue of whether or not a variable can be dropped from the attribute list just because it is correlated with another must be addressed. In the case of 'marina' and "harbor" it would most likely be acceptable to drop one of these attributes since the two are similar in interpretation and are technically related. However for other variable combinations, such as 'rest rooms' and 'tide pools', it is not clear how to handle high degrees of correlation.

Another issue is that correlations only detect pair wise dependencies. Furthermore, they are not completely appropriate for 0/1 variables, which almost all of these are. Collinearity can cause estimation problems by creating alternative specific constants that completely identify specific choice alternatives, and through omitted variable problems that arise from trying to correct for the problem.

More sophisticated means of handling dependencies in data involve the use of indices that reduce the number of attributes yet capture most of the information present in large sets of attributes. Further, when properly constructed, these indices can reduce the covariance among the indexed variables -- even reducing the covariance to zero in some cases. Principal components analyses (Pearson 1901, and more recently Rabe-Hesketh and Everitt 2001) can be used to "re-organize" explanatory data to find the unit length linear combinations of variables that have the highest variance. The dependencies among the variables are embedded within the principal components which may themselves be constrained to be orthogonal. Unfortunately, the resulting principal components may be difficult to interpret economically; there is no guarantee that the best linear combinations of variables will reflect technical or economic relationships among variables. Thus while Principal Components offers a very parsimonious way of constructing variables which capture the differences between the beaches, these variables may not have any obvious interpretation in the real world.

Sample correlations:

As an example, 9 beaches had either a Marina or a Harbor present. 7 of these 9 beaches had both. So 49/53 (92.45%) of the time they are identical. Several other variables are highly correlated (camping and fire pits, and a large number of volleyball nets with volleyball tournaments).

Restaurants and Concessions are very weakly positively correlated. (.26 when Restaurants is normalized, but 0.01 if Restaurants is left as continuous- note that Concessions is Binary).

Storm drains and Rivers are very weakly negatively correlated. (-.25 if Normalized , -0.075 if Storm drains is left as continuous– note Rivers is Binary).

Beach Specific Dummy Variables: the case of Venice Beach and Concessions

In cases where a variety of selections from the covariate list does not work well in regression analysis we have looked into the possibility of using alternative-specific constants for beaches. Here "working well" is measured by the difference between the probability value for each beach calculated based on using only covariates which differ across individuals and activities plus an alternative-specific-constant for each beach, and a model fit using the beach attribute covariates instead of ASCs for each beach.

It is suspected that the most likely candidate for a site-specific variable would be Venice Beach. Venice Beach, California is a major tourist attraction and is known for its non-beach attributes as much as it is for its beach attributes. It has substantial non-beach related attractions (street performers) and infrastructure (muscle beach and the skating pit) and has parkland between the beach and these attractions.

Beyond the specific attributes that are uniquely found at Venice Beach, Venice Beach also has certain attributes is greater abundance than other beaches. A primary example would be "concession stands." While a majority of the beaches have no concession stands and a few beaches have up to three, Venice Beach has fifty. A site-specific variable for Venice, then this would handle the concessions issue.

Using an Alternative Specific Constant for Venice is not an ideal solution in that Venice Beach is a fairly popular beach, and this would mean that trips where Venice is the destination would tell us nothing about the covariates of interest, such as water quality. However, if the models show that Venice Beach is sufficiently different from what the covariates predict than it may be necessary to use an Alternative Specific Constant for Venice Beach and loose the ability to use that data for policy experiments.

A.3.4 Principal Component Analysis

Principal Component Analysis was applied to the full set of site attribute variables. Over 95% of the variation of the attribute variables can be explained with 2 or 3 principal components. However these principal components turn out to be hard to interpret; they lack any simple interpretation in terms of the underlying variables. The PCS was conducted both on the raw data and also on various transformations of the data, including transformations that standardize variables by the estimated length of the beach and transformations that normalize variables to have a standard deviation of 1 and a mean of zero. An interesting finding is that the normalized data require a significantly greater number of principal components to capture the variation; thus the normalization appears to have been relatively counterproductive for the reduction of data.

One approach would be to first reduce the dataset by eliminating highly correlated variables, followed by the further reduction of the dataset using PCA

To address the lack of economic interpretability of the principal components constructs, the beach attribute data was separated into categories that represent the possible dimensions of beach preferences. These categories included parking, availability of athletic recreational actives, coastline facilities development, commerce development, and nature. Principal components analysis was completed for several iterations of these categories in hopes of constructing a set of principal components that could be used as a set of interpretable indices and reduce the number of attribute variables necessary to characterize the choice set. The results of these analysis indicated that the reduction of attribute variables per category was insignificant while constraining the principal components to account for at least 95% of the variation between the attribute variables. This illustrated that it was unlikely that the number of beach attribute variables could be reduced in this manner.

Several normalization strategies were investigated through the means of creating correlation tables, running PCA, and estimating exploratory choice models. Note that the variables were first normalized for length (if needed) and then normalized (mean 0, std. 1). Continuous attribute variables are normalized by length, except for concessions which are transformed into a dummy (even for Venice Beach).

Additionally it is possible that normalizing by length is not appropriate for Principal Components Analysis since at many beaches, there are long stretches of unused beach. In fact, it is the opinion of several researches that most use at the beaches is clustered into small areas.

A pure principal components analysis may be the most statistically efficient way of both accounting for collinearity in the RHS variables and reducing the number of RHS variables. However there is no readily accessible interpretation of the principal components or their coefficients.

A.3.5 Category Based PCA

Principal Component Analysis also was applied to subgroups of site attribute variables. The idea behind this approach is that by grouping attributes into appropriate groups and then conducting the principal component analysis, the data reduction could take place while forming a set of indices that would be more easily interpretable. The subgroups of variables we explored included Parking, Recreation, Seaside Development, Development, and Nature. This approach produced somewhat mixed results. It was possible to form subgroups that yield principal components that explained a large amount of variation in the attribute matrix, but these groups do not fit neatly into any natural categories. There was also some difficulty in organizing the subgroups, if this was done properly we would have a fairly large number of categories that each require several PC's. The reduction in the number of variables afforded by this approach does not seem large enough to compensate for the loss in explanatory power using this approach.

Another approach would be to group the attributes into categories and complete a PCA on each category forming a type of index that may be more interpretable than a straight 'one shot' PCA including all variables. Proposed groupings are listed in Table A.2

.

	PCA Group	Alternative PCA	Format	Normalize by Length
Beachclubs	Dev	Development	0 to 3	
Concerts	Dev	Development		
Condoshotels	Dev	Development		
Houses	Dev	Development		
Stores	Dev	Development		
Concessions	Dining		0 to 50	
Restaurant	Dining		0 to 18	
Isleview	Nature			
Oilpumps	Nature			
Oilrigs	Nature			
Powersewer	Nature			
Rivers	Nature			
Rocky	Nature			
Sandy	Nature			
Stormdrains;	Nature		0 to 19	Yes
access_lot	Parking			
access_ped	Parking			
access_street	Parking			
meters	Parking			
Parking	Parking			
Parkingfee	Parking			
Publictransit	Parking			
Sidewalk	Parking			
Streetparking	Parking			
Dogsok	POLICY			
TRAVEL COST	POLICY			
WATER QUALITY	POLICY			
Camping	Seaside Dev	Development		
Firepits	Seaside Dev	Development	0 to 241	Yes
Harbor	Seaside Dev	Development		
Lifeguards	Seaside Dev	Development	0 to 24	Yes
Marina	Seaside Dev	Development		
Pier	Seaside Dev	Development		
Pubfac	Seaside Dev	Development		
Rentals	Seaside Dev	Development		
Restrooms	Seaside Dev	Development	0 to 15	Yes
Showers	Seaside Dev	Development		
Bikepath	SECONDARY			
Diving	SECONDARY			
Fishing	SECONDARY			
Nature	SECONDARY			
Playground	SECONDARY			
Surfing	SECONDARY			
Tidepools	SECONDARY			
Vballnets	SECONDARY		0 to 107	Yes
Vballtourney	SECONDARY			

Table A.2: Variable Subgrouping for Catergory Based and PCA Data Reduction

One approach is to use the "major" PC's as data and then do an LM test for the linear restrictions that the coefficients on the other minor PC's are zero – this would address the computational issue (not running the model with everything in it), and assuming we do not reject the null that the minor PC's are zero, then we have good support for the approach. This methodology does not address the interpretation issue, although this may not be of importance as long as major policy variables are not included in the PCA.

Several preliminary PCAs were estimated using all attribute variables. The main thing to note here is that one needs about 26 Principal Components to explain for 95% of the variation in the attribute data – approximately 20 more than we would like.

Lastly, there is no consistent recommendation in the literature regarding the criteria to be used in determining the appropriate number of principle components. For example, we find that for the General Development sub-grouping the PCA results are unsatisfactory. The variables were: houses; condos/hotels; concerts; stores; beach clubs. One principal component explains much of the variation (see immediately below), but adding more principal components provides addition explanatory power. The question remains, when does the marginal cost of adding additional components outweigh the marginal benefit increasing the degree to which the PCs capture variation in the data.

PC	1	2	3	4	5
Variation Explained	64.2%	18.1%	11.0%	5.1%	3.0%

A.3.6 Composite Variable Approach

Introduction

A more practical and easily-understood method of data simplification is to construct composite variables which group together closely-related attributes. A priori, it is considered that indicators of commercial activity, development, natural amenities, and scenic blight would represent good combinations of the variables. These variables were constructed and a high degree of interdependency was found between them; this dependency will likely result in model instability.

Variables

The first approach to creating composite variables was to look at groupings that the research team felt a priori should work. The key to constructing these variables is to identify attributes that capture the same, or very similar, information for the beach goer.

One example was to collapse "rocky" and "sandy" into a single dichotomous variable. It was noticed that there were zero cases where both the rocky and sandy variables were valued at zero, since a beach be either sand or rocks. Thus there were not the four possible cases one would assume given a pair of dichotomous variables. Since the absence of sand dramatically changes the recreation possibilities for a beach goer, we thought that it may be useful to combine these two variables into a single "very rocky" variable indicating rocky but not sandy.

The team initially explored many composite variables, and the rest of this section will highlight some of them.

Rockiness

The "rocky" and "sandy" variables are used to construct the following trichotomy:

- sandy = 1 and rocky = 0 (== Not at all rocky)
- sandy = 1 and rocky = 1 (== somewhat rocky)
- sandy = 0 and rocky = 1 (== very rocky)

The 'rockiness' variable captures how rocky the beach is and can be considered to range from 0 to 2 (which implies an undesirable cardinal relationship between somewhat rocky and very rocky), or as a dichotomous variable which captures either no rocks or no sand, depending on definition.

Beach Access

The original data set included several attribute variables pertaining to the type of access available to the sand (pedestrian, parking lot, street). In terms of these options the research team has proposed that the most important aspect of beach access is how far one has to walk after getting out of ones car. As an illustration, at most urban beaches one can walk to the sand from the street or parking lot in several minutes. Whereas at some of the rural beaches, for example Salt Creek, reaching the sand necessitates a fairly long hike from the parking lot.

- Pedestrian access: there is a path to the beach (only)
- Street access: one can walk off the street to the beach

- Lot access: one can walk directly from the lot to the beach.

The research team tested the composition of several aggregation strategies to capture the variation in this variable. The most promising alternative was to distinguish beaches that only had pedestrian access from all other beaches. This strategy would identify beaches that are only accessible from a path – this category of beaches is perhaps the most different from the others. The 'Pedestrian Only' variable is valued 1, if there is no Street or Lot Access, and there is Pedestrian Access.

Natural Indicators

A new 'Natural Area' variable was created using the "near nature area" and "tide pools" variables. The composite variable is set to 1 if either or both of these variables equals 1, and 0 otherwise. For example, using "Tide pool" and "Nature", 25 beaches have neither, 18 have one, and 9 have both. A composite variable was generated which was zero if both of these elementary variables was zero, and one if either of them was a one.

Nice View

- River with a nice view (this would be created as a subset of RIVER).

- Isle View

A 'Nice View' variable was created to equal 1 if either or both of these variables equals 1, and 0 otherwise. This was not implemented because of difficulty objectively classifying the quality of river views.

Ugly Beach

- Oil Pumps
- Oil Rigs
- Power Plant/ Sewer
- Storm Drains

An 'Ugly Beach' variable was created to equal 0 if none of these is 1, and 1 otherwise.

"Ugly" was constructed using 'Oil pumps', 'Oil Rigs', Power/Sewer Plants', and 'Storm Drains." 4 beaches have none, 34 have one, 12 have two, 2 have three, and 0 have all four.

It should be noted that there are no oil pumps on the ocean side of the Pacific Coast Highway, which probably mitigates the impact of the Oil Pumps. It should also be noted that the variable Oil Pumps seemed to cause instability in the model because it worked with other variables to generate what may have been effectively an alternative specific constant.

General Development

- Houses
- Condos Hotels
- Beach clubs

A 'General Development' variable was created to equal 1 if either or both of these variables equals 1, and 0 otherwise. "Development" was constructed using: Houses, Condos/Hotels, Beach Clubs: 17 have zero, 18 have one, 11 have two, and 6 have three.

Commerce /Dining

- Restaurants
- Concessions
- Stores
- Volleyball Tournaments
- Concerts
- Pier

A trichotomy was created to equal to 0 if all of these are 0 (corresponding to zero commercial activity), 1 if several of them are 1 (corresponding to some commercial activity) and 2 if many or all of them are 1 (corresponding to a lot of commercial activity). "Commerce" was constructed using: concerts, concessions, restaurants, stores, volleyball tournaments, and rental shops. 22 beaches have zero , 14 have one , 5 have two, 4 have three, 5 have four, 0 have five, and 2 have six. It seems natural that scores are grouped together as (0,1) and (2-6)

As with rockiness, the trichotomy itself will be used in one of two ways: 1) as a cardinal measure taking on the values 0, 1, or 2. and 2) as an indicator variable taking on the value 1 if some threshold is reached and 0 otherwise. We experimented with several variations.

Seaside Development

- Rentals
- Lifeguards
- Fire pits
- Showers
- Public Facilities

Again, a trichotomy was created that was equal to 0 if all of these are 0 (corresponding to zero commercial activity), 1 if several of them are 1 (corresponding to some commercial activity) and 2 if many or all of them are 1 (corresponding to a lot of commercial activity). We also can create one or more binary variables capturing the level of development.

Good for young children

- Playgrounds
- Sandy

The idea is to create a variable that equals 1 if the site is likely to be attractive to parents with small children, and 0 otherwise. Because this requires assumptions about household composition at a given time (which is tricky because of divorce/custody/visitation issues), this variable was not explored to the same extent as some of the others.

Parking

Two composite variables can be constructed:

- A) Abundant and easy parking: which is subjectively coded based on field observations.
- B) Parking is free: Equals 1 if there is parking at the site and all parking if free, and 0 otherwise.

These can be used to replace 6 original variables:

lot access	street access	parking lot
street parking	meters	parking fee

Since all beaches have available parking and parking at a variety of costs, these variables did not end up being well-defined and thus were not useful.