

RESIDENTIAL RELOCATION CHOICE AND CONSEQUENT BEHAVIORAL CHANGE

by

XIAOQIANG CHEN

A dissertation submitted to the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2009

© 2009

XIAOQIANG CHEN

All Rights Reserved

This manuscript has been read and accepted for the
Graduate Faculty in Engineering in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

_____	Cynthia Chen
Date	Chair of Examining Committee
_____	Mumtaz Kassir
Date	Executive Officer

Robert Paaswell

Claire McKnight

Hongmian Gong

Kaan Ozbay
Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

RESIDENTIAL RELOCATION CHOICE AND CONSEQUENT BEHAVIORAL CHANGE

by

XIAOQIANG CHEN

Adviser: Professor Cynthia Chen

Transit-oriented developments with the goal of reducing dependency on the automobile have been observed nationwide. The implicit hypothesis behind these developments is that where one lives would affect his or her consequent activity and travel behavior. Many questions remain despite decades of research in residential location choice and activity and travel behavior. In residential location choice literature, previous studies have modeled residential location choices as static ones, with no memory of the past. In addition, few studies have modeled the spatial correlation between alternative locations, although they may exist in reality. In activity and travel behavior literature, most studies implicitly assume that behavioral changes are instantaneous, despite the recognition that a behavioral change in response to an external trigger may take time to occur, due to various types of constraints one may face. This dissertation will introduce a residential relocation model that addresses spatial correlation and accounts for the influence of the prior residential location on one's current preference. Furthermore, I will examine how a change in the built environment (spurred by a change in residence) can affect the response lag generated by a significant change in the time people allocate to discretionary activities and travel.

I developed a generalized extreme value model for residential location choice to account for spatial correlation. I considered the influence of prior location in the model by assuming that the preferences toward various attributes in the current location choice (represented by coefficients in the utility function) are functions of the characteristics of the current location and the past location. Furthermore, the model also considers lifecycle interaction with the prior location influence. To study how a change in the residential location may trigger activity and travel behavioral changes, I constructed a survival model to investigate the response lag in changes to the time allocated to discretionary activities and travel. I addressed three issues in the survival model: data censoring, partial observations, and multi-event.

The empirical results confirmed that past housing experiences influence current location choices. In particular, I found that people have become more tolerant to long commute distances and place more value on attributes like retail, open space, and business opportunities after previous heavy exposure. I also found that lifecycle could interact with the influence of past housing experience. In addition, spatial correlation is confirmed. The results from the response lag study suggested that temporal constraints, as well as family and social obligations were significant contributing factors to the length of the response lag, while the built environment was not.

This dissertation contributes to the existing literature by recognizing the dynamics inherent in human behavior, developing methodologies to model real-world phenomena, uncovering underlying behavioral mechanisms, and lending insights into how these factors can influence policy decisions. In particular, it contributes to the residential location choice literature by

empirically validating the existence of past housing experience in current location choices and demonstrating the interactive effect between past housing experience and lifecycle. In travel behavior analysis, this study is one of the first, if not the first, to investigate the response lag of a significant change in time use in response to a residential relocation. Methodologically, the dissertation contributes to the literature by developing a generalized extreme value model to account for spatial correlation in location choices and a survival model handling left-censoring, partial observation, and multi-event in the response lag study.

ACKNOWLEDGMENTS

I could not have finished this research without the support of many people. I would like to express my gratitude to Dr. Cynthia Chen, my supervisor, for her invaluable advice, support, and guidance. She helped me to develop the necessary skills required for this study. Without her guidance and persistent help, this dissertation would not have been possible. It is a privilege and pleasure to work with her. My deepest gratitude is also due to the entire Dissertation Committee: Dr. Hongmian Gong, Dr. Claire McKnight, Dr. Kaan Ozbay, and Dr. Robert Paaswell. Each member provided insights that guided and challenged my thinking, substantially improving the finished product.

Special thanks go to my family and friends. My wife, Xiaole Qi, has shared with me the joy and frustration that accompanied my five years of PhD study. I wish to express my love and gratitude to my beloved families for their understanding and endless love, throughout these years. Finally, I wish to convey my special thanks to all my graduate friends, for sharing the literature and providing invaluable assistance.

Table of Contents

Chapter 1	Introduction.....	1
1.1	Introduction	1
1.2	Dissertation Outline.....	7
Chapter 2	Literature Review of Residential Location Choices.....	9
2.1	Introduction.....	9
2.2	Conceptual Framework.....	10
2.3	Impacts of Household Socioeconomics and Demographic and Location Attributes on Residential Location Choice	11
2.4	Gap and My Research Goal of Location Study	25
Chapter 3	Literature Review of Behavior Dynamics.....	27
3.1	Linking Residential Relocation and Behavior Dynamics.....	27
3.2	Past Studies About the Behavior Dynamics	30
3.3	Gap and My Research Goal of Behavior Dynamics.....	36
Chapter 4	Theoretical Framework and Methodology of Residential Location Model.....	38
4.1	Introduction	38
4.2	Random Utility Maximization Framework – A Theoretical Basis	38
4.3	Existing Model Structures	40
4.4	Model Structure for Residential Location Study.....	54
4.5	Neighborhood and Accessibility Attributes	65
Chapter 5	Theoretical Framework and Methodology of Response Lag.....	74
5.1	Introduction.....	74
5.2	Survival Analysis.....	75
5.3	Censored Data, Partial Observation and Multi-event	78
Chapter 6	Empirical Residential Location Choice Study	90
6.1	Empirical Dataset.....	90
6.2	Explanatory Variables.....	93
6.3	Model Estimation Results	98
Chapter 7	The Empirical Study of the Response Lag Question	106
7.1	Objective	106
7.2	Dataset.....	106
7.3	Variable Definition	109
7.4	Descriptive Analysis	111
7.5	Model Results	114
Chapter 8	Conclusions, Policy Implications, and Future Directions	119
8.1	Residential Location Choice Study.....	119
8.2	Response Lag Study.....	122
References		127

List of Tables

Table 2- 1 Attributes Considered in Residential Location Choice Research.....	12
Table 6- 1 PSTP Dataset Information	92
Table 6- 2 Descriptive Statistics of the Sample and the Study Region	98
Table 6- 3 Estimation Results of the Distance-based GEV Models	100
Table 7- 1: Descriptive Statistics of the Four Groups in the Sample.....	112
Table 7- 2: Estimation Results of the Weibull Based Survival Model	115

List of Figures

Figure 4. 1: Assumed Correlation Structure	55
Figure 4. 2: An Example of Four Alternatives in a Nested Logit Model Setting	58
Figure 5. 1: Censoring Types.....	79
Figure 5. 2: Left-Censored Data	80
Figure 5. 3: Left-Censored Data and Left & Right-Censored Data.....	83
Figure 5. 4: Illustration of Censoring Due to Partial Observation.....	85
Figure 6. 1 Puget Sound Region	91
Figure 6. 2: Retail Accessibility in Puget Sound (2000)	94
Figure 6. 3: School Quality in Puget Sound (2000).....	95
Figure 6. 4: Population Density in Puget Sound (2000)	96
Figure 6. 5: Distribution of Housing Value per Room in Puget Sound (2000) ...	96
Figure 6. 6 Parameter Estimates for Commute Distance.....	102
Figure 6. 7: Parameter Estimates for Retail Land Use.....	102
Figure 6. 8: Parameter Estimates for Open Space	103
Figure 6. 9: Parameter Estimates for Recreational Land Use	103
Figure 6. 10: Estimated Spatial Correlation vs. Distance between Zones (miles)	105
Figure 7. 1: Illustration of the Importance of Continuous Participation in Survey	108
Figure 7. 2: Illustration of the Importance of A Common Study Period	109

Chapter 1 Introduction

1.1 Introduction

Our 50-year history of transportation planning rests upon two essentials: supply or demand. The early days of transportation relied entirely on the supply principle by seeking to relieve congestion through additional lanes (Weiner, 1997). In the 1980s, people began to realize that supply-based strategy was a futile effort in the context of an ever-increasing demand. Eventually, society embraced a joint force of demand and supply by implementing Transportation Systems Management (TSM) and Transportation Demand Management (TDM) strategies (Weiner, 1997). More recently, as governments increasingly view transportation investments as a way to address a multitude of urban problems (e.g., air pollution, energy consumption), land-use planning on the built environment has re-emerged as a useful tool to achieve sustainability at all levels. To hypothesize, the makeup of one's residential location will affect one's travels. By re-designing our neighborhoods with a focus on population density and the diversity of their activities, we can eliminate or reduce the need to use automobiles. Potentially, such neighborhoods, known as traditional neighborhoods or transit-oriented developments, will not only reduce Vehicles Miles Traveled (VMT) but also lower energy consumption. Indeed, these new neighborhoods can be observed across the country. In 2001, there were 213 traditional neighborhood projects (15 acres or above) completed or in construction, which is 37 percent more than in 2000. If we count those projects that are still in the planning phase, the total number jumps to 375, an increase of 22.5 percent over those completed and under planning projects in 2000 (October/November 2001 issue of New Urban News).

The force that residential location exerts on one's travels is complex. First, it is elastic. Depending on the land-use distribution surrounding a residential location, it can either constrain or promote one's activity and travel sphere. A location abundant with one or more opportunities encourages people to engage in those opportunities, while a location scarce in resources hinders

them. It is possible that heterogeneity within in land-use distribution has varying impacts on one's activity and travel pattern.

Second, the force is not instantaneous, but gradual. New behaviors induced by a change in the built environment may not become evident until some years later. For example, after relocating to a place with ample outdoor space, the increase in time allocated to recreational activities may not become significant until after one has developed a taste for outdoor activities. In his seminal paper, Kitamura (1990) points out three unrealistic behavioral assumptions frequently adopted in literature: symmetric, instantaneous, and stationary. Instead, he argues that behavior is asymmetric, delayed, and dynamic.

Studies in residential location decisions are abundant. An important recognition from literature is that people are heterogeneous, and in addition to commute time (which is frequently used to represent home-work link) and rent (or housing cost), which are the only two factors considered in the earliest models (Alonso, 1964; Muth, 1969), many other factors play a role in deciding one's residential location (Rouwendaal and Nijkamp, 2004). Thus, since McFadden's development of discrete choice models, which allowed many factors to be considered (McFadden, 1978), we have seen a long list of variables incorporated in residential location models. These attributes range from those related to dwelling units, the built environment, the household, and the individual. For example, in terms of socio-economic characteristics, people under 35 (Dieleman, Clark, and Deurloo, 2000) and renters tend to change residential locations more often than other groups. Deng, Ross and Wachter (2003) found that cultural differences play a role in the selection of residential location. For neighborhood-level characteristics, Hwang, McGarrell, and Benson (2005) found that public satisfaction with the police correlates with residential location selection. Galilea and de Dios (2005) identified the role of noise level in the residential location context. As for attributes related to dwelling units, housing price has been consistently found to decrease a household's utility (So, Orazem, and Otto, 2001). Among the built environment characteristics, accessibility to employment is the most often considered, and this is usually represented by commute time (Humpel et al., 2004). In general, there is a negative relationship between commuting time and the probability of selecting a location (Srouf et al., 2002; Molin, et al., 1999; Waddell, 1996; Timmermans et al., 1996; Hunt et al., 1994).

The recent surge of interest in the built environment has also resulted in a new sprout of studies investigating its linkage to travel behavior. Past studies have examined different dimensions of travel, including travel distance (Næss et al., 1995; Curtis, 1995; Stead, 2001), trip frequency (Cervero and Gorham, 1995; Ewing et al., 1996; Friedman et al., 1994;), mode split (Crane and Crepeau, 1998; Dunphy and Fisher, 1996; Kitamura et al., 1997; Chen et al, 2007;), and travel time (Giuliano and Small, 1993). For instance, Stead (1999) showed that the proximity to local facilities is positively associated with the average distance traveled. In a study of trip frequency, Friedman et al (1997) suggested that daily trip-rates and auto driver trip-rates for households in suburbs are significantly higher than the corresponding rates for those in traditional neighborhoods. Dunphy and Fisher (1996) found that households in a dense neighborhood tend to have higher transit usage compared to those residing in a low-density area. Chen et al. (2007) concluded that density as well as cost-variables (travel time and travel cost), job accessibility, and transit accessibility together will influence the mode choice of home-based work tours. Giuliano and Small (1993) demonstrated that there is a small but significant relationship between job ratio (the ratio of the number of jobs in the area to the number of workers residing in that area) and commuting time, suggesting a possible spatial mismatch problem.

When it comes to impact of built environment on activity patterns, some studies (Wilcox et al. 2000; Parks et al. 2003; Rutten et al. 2001) investigated how the built environment could affect physical activity. Meloni et al. (2007) reported how individuals segregate time spent on discretionary and other activities. Using a sample of homemakers in New York City and the surrounding suburbs, Chen and McKnight (2007) found that the built environment does affect people's activity patterns, but their magnitude is comparable to that of socio-economic characteristics. The implication of their findings is that policies focusing on changing the built environment only are likely to achieve limited success.

The majority of the existing research linking the built environment and one's travels relies on cross-sectional datasets. This means that parameters are estimated via inter-person comparisons. A potential problem with the use of a cross-sectional dataset is its limited ability to distinguish between correlation and causality. Residential self-selection is a major hindrance in the

researchers' conclusion on the impact of the built environment on travel behavior; it refers to the phenomenon that people select a particular neighborhood for the unique lifestyle it offers. For example, a person who loves walking and transit will prefer a dense neighborhood in the city to a spread-out environment in the suburb. In this case, the decision to walk or use transit is determined prior to relocation. Cao et al. (2006) described two methods to mitigate the residential self-selection problem: direct questioning and the inclusion of attitudinal variables in the regression model. Researchers have attempted both methods in empirical applications. For the former method, Hammond (2005) investigated the inter-dependency between residential location choice and mode choice among respondents living in Century Wharf, Cardiff. Through discussions with respondents, Hammond found that people selectively locate in a residential neighborhood to realize their commute mode. For the latter, Chen et al. (2007) used a simultaneous double equation system to control for the residential self-selection problem. Their study verified the existence of a residential self-selection effect.

A more efficient way to handle this issue is via a panel dataset, since they record the process from the occurrences of an external stimuli and a behavioral change. If we can identify a behavior difference after the change of the built environment, we can at least reason that the built environment is one of the plausible causes of behavioral change. The confirmation of a sequentially observed linkage between a change in the built environment and a change in one's travels proves a stronger causality than cross-sectional studies that control for residential self-selection. Some researchers have also entertained the panel dataset approach. Using the Puget Sound Panel dataset, Krizek (2003) estimated a regression model with the change in vehicle miles traveled, person miles traveled, number of tours, and number of trips per tour as dependent variables. He then used socio-demographic and the built environment features as explanatory variables. Research indicates that increases in neighborhood accessibility (defined as urban forms providing attractive, multimodal options such as auto, transit, cycling, and walking for a variety of travel purposes, both within and between neighborhoods, thereby reducing auto travel, see Krizek, 2003 for a detailed discussion) would decrease vehicle miles traveled, person miles traveled, and the number of trips per tour. However, neighborhood accessibility increases the number of tours. The impact of the built environment on mode split, however, is minuscule.

People do not appear to change mode choice due to changes in built environment; instead, they choose to live in neighborhoods that match their mode preference.

The existing literature in residential location is deficient in several aspects. Households model residential choices in a static approach; the households are assumed to have no memory of past locations when making new decisions. If the past location does influence one's current location choice, the exclusion of previous locations in the study prevents us from answering some important questions relating to the impact of residential history. For example, how does one attribute of a previous location affect one's preference towards that attribute? Everything else being equal, will a household accustomed to a suburban environment tend to relocate to a neighborhood with large open spaces? Without considering previous locations, we are unable to answer these questions. In this paper, I will use the "historical deposition influence" to define the effects of previous locations on one's preference for future ones. 1. The other gap in previous residential location studies is about spatial correlation. Neighborhoods or Census tracts (which usually serve as alternatives to residential location studies) close to each other may share common attributes. If the common attributes were not captured in the residential choice model, it can create correlation errors. This type of correlation is called "spatial correlation" because it is generated from the spatial proximity of two alternatives. The majority of residential literature doesn't account for the spatial correlation of alternatives. Ignorance of the spatial correlation results in biased estimates (Greene, 2003).

Studies analyzing the impact of the built environment on behavioral changes have largely focused on understanding the direction and the magnitude of the change, given a change in the built environment. Missing from the large body of literature is a time related question. In his seminal paper on travel behavior and panel datasets, Kitamura (1990) noted that a behavioral change is often not instantaneous and there exists a "response lag" between the time of the occurrence of an external stimulus and a behavioral change. A response lag exists when there is a time lag between a trigger (e.g., a change in home location) and a response, for example, a change in travel behavior (Kitamura, 1990). Several reasons may explain the existence of a lag. The individual may not immediately acquire the knowledge of a trigger. Alternatively, the individual may ignore the initial trigger until it accumulates over a time, and to a significant and

perceivable threshold (Coombs, Dawes, and Tversky, 1970). In other cases, when the behavioral change involves large costs, the constraints of family resources may lead to a response lag. Finally, individuals may delay a behavioral response until the next trigger (Clarke et al., 1982).

Objectives

My effort in this dissertation research will strike in two directions. First, I will develop a residential location model accounting for the historical deposition influence and spatial correlation. Second, I will examine how a change in the built environment (brought by a residential relocation) will affect the response lag for a significant change in the time allocated to discretionary activities and travel. The specific focus on the time allocated to discretionary activities and travels is due to the fact that discretionary activities and travel are the least constrained by a person's societal responsibilities (e.g., work) or biological needs (e.g., sleeping).

In particular, the specific aims of this dissertation are:

- To develop a residential location model accounting for historical deposition influence and spatial correlation. With this model, I plan to answer the following questions:
 1. How does accessibility to an opportunity at a prior location modify one's preference in residential location choice?
 2. How does the influence of historical deposition vary between childbearing and non-childbearing households?
 3. How geographically located zones are correlated with each other?
- To develop a survival model of response lag for the change in time allocation to discretionary activities. With this model, I am going to answer these questions:
 4. After a residential relocation, how long will it take for a significant change in the time allocation to discretionary activity?
 5. What is the role of the built environment in affecting the response lag?
 6. In addition to the built environment, do socio-demographic, temporal, and spatial constraints also affect the response lag?

The study on residential location contributes to the geography and regional science literature by providing more accurate estimation results on a set of interacting factors in residential choices. I will do this by recognizing the influence of historical deposition on a household's residential decision, and modeling the residential choice in a spatially correlated setting. The new result will contribute to the understanding of the influence of contextual effects on people's preferences in residential location choice. People are adaptive to the local environment; it is likely that the environment they reside in is modifying their preferences. The longer they live in a location, the more likely this is the case. This research will shed light on how the local context modifies people's preferences.

The study on the response lag would lend us insights on the time process of a behavioral change, following a change in residence. Although many studies have already revealed how activity and travel behavior would change in magnitude and direction, these studies tend to focus on short-term changes. Studies using cross-sectional datasets make inter-person comparisons and they completely ignore the fact that behavioral changes do not have to be symmetric and instantaneous. Studies using the panel dataset, however, (e.g., Krizek, 2003) examined the behavioral changes in the year immediately following a residential relocation. These studies assumed that all changes would occur within one year's time. This selection criterion prevents us from understanding the long-term impacts of the built environment on behavior.

Studies on the response lag will also have implications in travel surveys designed to capture behavioral changes over time. If most of the behavioral changes take place within a certain period, a panel survey that exceeds that particular period may not be necessary. This would potentially have a significant cost-saving effect. The verification of the existence of a response lag also sends a message that cross-sectional datasets, as expected, are not sufficient to capture dynamic changes over time.

1.2 Dissertation Outline

The rest of the dissertation is organized as follows. I will review the existing studies in residential choice modeling in Chapter 2. First, I will provide a conceptual framework of

residential relocation as a guideline for the whole chapter. Following the conceptual framework, I will conduct an extensive review of existing studies on residential location decisions. After the literature review, I will discuss the limitations of the existing studies and the specific issues I am going to address in my dissertation.

In Chapter 3, I will review existing studies on the connection between the built environment and activity and travel. . Given the large volume in the area, the focus will be primarily on studies using a panel dataset. I will discuss what the past studies have answered and where gaps remain.

I will present the theoretical framework and methodology of the residential location study Chapter 4. In more detail, I will discuss the background of the discrete choice model and develop a residential location model accounting for the historical deposition effect and spatial correlation. I will address two problems in model development: the inclusion of the previous location attributes and the method to deal with spatial correlation.

In Chapter 5, I will discuss the theoretical framework and the methodology for the response lag study and provide a survival model, which will propose a new approach to deal with left-censoring data, partial observation, and multi-exit issues.

I will estimate the model developed in Chapter 4 with the Puget Sound Panel Dataset (1989-2002). The panel data will provide households residential locations for each survey wave, as well as their socioeconomic and demographic variables. Chapter 6 will detail the estimation results and the conclusion of relocation study.

In Chapter 7, I will discuss the regression results from the survival model. In particular, the focus is to answer the questions 4, 5, and 6 (see page 5). I will also discuss the policy implications regarding the effect of the built environment on response lag.

Chapter 8 will summarize the main findings from Chapters 6 and 7 as well as discuss study limitations and policy implications. I will also discuss future research directions in this chapter.

Chapter 2 Literature Review of Residential Location Choices

2.1 Introduction

While there has been voluminous and profound literature generated on residential location choice (e.g., Alonso, 1964; Muth, 1969; Lancaster, 1966; Quigley, 1976; Lerman, 1977; Williams, 1979; Segal, 1979; Onaka and Clark, 1983; Abraham and Hunt, 1997; Parkes et al., 2002; Cao et al., 2008), greater attention is given to the current location. These studies use cross-sectional data and address the question of how a household trades between various attributes associated with the current location. Historical deposition, or how a household settles into its current position, is typically not analyzed. Therefore, *ceteris paribus*: a person who has been living in an urban setting, would have the same preference as one who has only been exposed to the rural environment in the past. This perspective is contradictory to psychologists' theory of past actions predicting future behavior (Aarts et al., 1998), sociologists' life course perspective (Elder and Shanahan, 2006), or behavioral scientists' discovery of reference points (e.g., Kahneman and Tversky, 1979). Their work points to an influence of historical deposition.

The historical deposition influence suggests that one's preference toward an attribute depends on how much he/she has been exposed to that attribute. Previous exposure to a particular environment may help a person become accustomed to a particular attribute. A loss scenario, i.e. having a long commute distance before relocating, can make a person become more tolerant to a long commute in the future (Green, 1997). In a gain scenario, a person having grown up in an open environment may value open space more than someone who has not. The hypothesis of tolerance developing after historical exposure is supported by behavioral theories that state: "human beings are adaptive organisms constantly adjusting themselves" (Anderson and Milson, 1989; Anderson, 1990) and that assume "preferences are often formed over time (Camerer, 1998)". It is also possible that one will become more averse to an attribute in a loss scenario or acquire an adventurous aptitude for something completely different in a gain scenario. In this case, a long commute in the past may prompt one to avoid any location with a long commute,

whereas life in a rural area may position one to lean toward an urban environment. Such examples are not well documented in the existing literature, but are certainly heard of in our own observations as anecdotal evidence (Chen and Lin, 2008).

This dissertation focuses on residential location choices with respect to the historical deposition influence. The rest of this chapter is organized as follows. In Section 2.1, I will provide a conceptual framework for the residential location decision under the historical deposition influence. In Section 2.2, I will discuss past study results on the residential location model. Discussions on residential relocation studies will follow in Section 2.3. In Section 2.4, I will summarize the research gap and identify objectives of a residential choice component in my dissertation.

2.2. Conceptual Framework

In the residential location choice model, it is assumed that households try to maximize the utility from location attributes in the selection process. The attributes of a residential location could bring positive utility such as open space or retail opportunity, and on the other hand attribute such as commute distance would mean a disutility for households. Underlying the selection of a location is the trade-off process between these attributes to maximize total utility. The earliest residential location model is monocentric (Alonso, 1964), in which households are assumed to trade-off between housing space and commute distance. Such a simple two-factor model is no longer sufficient to depict today's reality, which is characterized by polycentric urban structure, females' labor participation, and heterogeneous population demographics. With the introduction of the discrete choice model, in which households are assumed to maximize the utility of a set of attributes, many more attributes could be incorporated in a residential location choice study. McFadden (1978, p.72) defined the utility for residential location as "a function of attributes of ... accessibility, quality of public services, neighborhood and dwelling characteristics, etc., as well as a function of the consumer's characteristics, such as age, family size, income, and so on". Jara-Diaz and Martinez (1999) suggested that variables in utility function should include neighborhood characteristics, income, housing price, and accessibility. In these studies, neighborhood characteristics refer to the quality of the neighborhood such as school quality, air

quality, average income, crime rate, etc. The accessibilities are the extent to which a location enables an individual to participate in different types of activities in different locations.

The quantity of utility for one attribute depends on household preference. Strong preference indicates that households realized high amount of utility per unit attribute and vice versa. The preference may be affected by household socioeconomic and demographic characteristics, such as income and race. As mentioned in the previous paragraph, location characteristics include two types: neighborhood and accessibility. Location attributes directly affect the utility obtained by households; they are explicitly included as explanatory variables in the utility function. Socioeconomic and demographic attributes do not directly affect one's utility, but they modify household preferences toward neighborhood characteristics and accessibilities, and thus indirectly affect one's utility.

As discussed in Section 2.1, residential history may play a role in people's preference toward one attribute. In other word, a household's preferences for residential locations may be influenced by the attributes of previous locations. If for some reason a household needs to relocate, the influence of previous locations could play a role in the selection of a new location. For instance, households that originally lived in the suburb with large open spaces could have stronger preferences for open spaces than more urban households. Hence, there will be three types of attributes included in the construction of utility function: household socio-demographic characteristics, previous location attributes, and new location attributes.

2.3. Impacts of Household Socioeconomics and Demographic and Location Attributes on Residential Location Choice

In the past several decades, we have observed a substantial number of residential location studies investigating how location attributes affect residential location decisions, and how household socioeconomic and demographic attributes could modify residential preferences. Table 2-1 shows three categories of variables that have been used in the past.

Table 2- 1 Attributes Considered in Residential Location Choice Research

Socioeconomics and Demographic Attributes	Location Attributes	
<i>Household Characteristics</i>	<i>Neighborhood Characteristics</i>	<i>Accessibility Characteristics</i>
• Car ownership • Education • Ethnicity • Income • Number of household members • Number of children • Occupation • Work schedule	• Air pollution • Average income • Crime rate • Density and openness of built form • Good area for children • Ethnicity composition • Pedestrian safety • School quality	• Accessibility to workplace • Accessibility to shopping • Accessibility to other activity • Accessibility to CBD • Public transit availability • Telecommute

2.3.1 Household socioeconomic and demographic variables

Household Income

Household income is the most important factor that affects a household's preference towards neighborhood characteristics and accessibilities. In the earliest monocentric model (Muth, 1969), travel cost $T(u,y)$ is expressed as a function of distance u and income y . Using Muth's model, Desalvo (1977) found that the income has multiple influences on residential choice. Higher income is associated with higher value of time, which indicated that the high-income households tend to live in a location closer to the center. On the other hand, households with higher incomes were also found to have higher household demand, which indicates high income households tended to choose locations far away from the center, where housing space is large. In making joint housing and job decisions, workers make trade-offs between commute distance, house quality, and wage. Guo and Bhat (2006) suggested that with higher household income, people are more willing to travel further in exchange for better housing.

Spousal incomes may play different roles in residential location decisions. Using husband and wife incomes for dual-worker families, Molin and Timmermans (2003) found that higher husband income encourages more detached housing. Given that suburbs have more detached houses, the above research result implies that households with husbands earning higher income

tend to locate in the suburbs with more detached houses and have lower value placed on commute time and travel time for shopping.

The smaller effect of the wife's income on residential location decisions may be the result of their traditional role in home. According to the "Household Responsibility Hypothesis" (Sermons and Koppelman, 2001), a wife usually has more household responsibilities than her husband does. Hence, she often works fewer hours and earns less income. Therefore, residential location decisions frequently rely more on the husband, who spends less time on family, more time on work, and earns a higher income. This Household Responsibility Hypothesis is supported by Hanson and Pratt (1995) and Turner and Niemeier (1997). Hanson and Pratt (1995) in particular, noted that changes in the male workplace lead to changes in the household's residential location. This "Household Responsibility Hypothesis" is also related to the common finding that females tend to have shorter commute distances than males (Guo and Bhat, 2006; Abraham and Hunt, 1997; Waddell, 1996). Since females tend to change their workplaces more often and work less, they would likely accept jobs with less income and shorter commute distance.

The Number of Household Members and Children

The number of household members and children is found to affect both housing space and the type of housing location (Bina et al, 2005; Molin and Timmermans 2003; So et al, 2001; Assadian and Ondrich 1993). Larger households and married couples prefer large housing space and ample parking spaces in the suburbs, while single-person households prefer shorter commute times and downtown locations (Bina et al., 2005; So et al., 2001). Larger numbers of children in a family is positively associated with the number of bedrooms in a house (Molin and Timmermans 2003; Assadian and Ondrich 1993). Larger households also value apartment enhancement more than access improvement (Bina et al, 2005).

Households with children may value neighborhood resources (e.g., park facilities) more than those without children because these neighborhood resources contribute to children's utilities, which are an important part of the household utility. Because of this, Kim et al (2005) argued

that child-rearing households (households having children under 18) prefer suburbs while non-child-rearing households prefer urban areas.

Ethnicity

Ethnicity plays a role in residential location choices (Bhat and Guo, 2004; Gabriel and Rosenthal, 1989). Bhat and Guo (2004) found that Hispanic households tend to live with their co-ethnics while African American households tend to locate in areas with poor work accessibility. Gabriel and Rosenthal (1989) found that the socioeconomic and demographic characteristics for African Americans are less elastic than Caucasians. Gabriel and Rosenthal suggested that the finding implied a possible racial segregation in the housing market. Compared with Caucasians, African Americans are more likely to face racial discrimination in the housing market and thus have fewer available residential alternatives. Their limited residential location alternatives may result in the observed smaller elasticity of African Americans.

Education

Sander (2004) found a positive effect of college education on the probability of living in large metropolitan areas in the U.S. This finding is probably related to the ample employment opportunities in large cities. This result is also consistent with So et al. (2001) whose study indicated that urban residents were more educated and have higher non-labor incomes and smaller families. As expected, higher education is positively correlated with higher income, and higher income households prefer large housing space (Bina et al., 2005; Molin and Timmermans, 2003; So et al., 2001). Bina et al. (2005) also found that individuals with Masters Degrees or higher are more likely to choose reduced travel times to shopping and a downtown area.

Neighborhood Profile

When making a residential location decision, households may want to locate in the areas whose residents match their own profiles. Guo and Bhat (2006) found that households tend to locate in an area with a high proportion of other households sharing similar household structure and size.

They also tend to select a location where people have similar income levels and tenure status (own or rent). However, this phenomenon may result from the fact that households with similar features would make similar decisions.

Work Schedule

Household work schedules may affect residential location decisions. A full-time worker is likely to face different time constraints from a part-time worker, or a retiree. For an employed person, commute time is an important factor to consider in residential location decisions, compared a retiree, whose commute time is no longer relevant. Empirical results verify this hypothesis. It was found that retired people are more interested in shopping access and are more likely to live in suburbs, while full-time workers value commute time savings (Bina et al., 2005; So et al., 2001). The effect of work schedules on residential location decisions between full-time and part-time workers can be ambiguous and bi-directional. Compared to full-time workers, part-time workers may be insensitive to commute time because they have more time and their schedule may be more flexible. On the other hand, part-time workers may be more sensitive to commute time, because it makes little sense to travel long distances in order to work for a short period. Many part-time workers may be homemakers, who often have more family responsibilities than their spouses do. The tight time constraint they face may make them more sensitive to commute time and select a shorter one.

Age and Life Cycle

Residential location choices are often prompted by major life events, such as getting married, having children, children leaving the home, and retirement (Krizek and Waddell, 2002). Glasgow (2000) showed that older people prefer non-metropolitan locations with amenities, while young and middle-aged people migrate to metropolitan areas to seek better job opportunities. Families with children may be less inclined to improve their job accessibility because of their willingness to trade-off job accessibility for a child-friendly environment.

Fernandez et al (2005) clustered residents in the Detroit metropolitan area into seven categories according to their preferences in location attributes. Cluster 1 included those residents who value good schools, housing costs, and social comfort (social comfort is defined by the familiarity with the area, closeness to family and friends, and similarity to other residents), and most importantly, have no strong preference in openness/nature (open space, close to natural areas, many recreational opportunities) in a neighborhood. Cluster 2 comprised of residents who value openness/nature, good school, short commute time, and residential aesthetics (attractive appearance of neighborhood as well as appearance and layout of the dwelling), and who have no preference for social comfort. Residents in Cluster 3 value closeness to family and friends and openness/nature as the most important. Those in Cluster 4 are most concerned about good schools and short commute times. Cluster 5 residents are those who are mostly concerned about social comfort, dwelling, and residential aesthetics. Cluster 6 consists of those who highly value openness/nature and residential and dwelling aesthetics, but not school quality and short commute time. Cluster 7 residents have a strong preference for openness/nature, good schools, and short commute time, but not aesthetics and social comfort.

Relating the preferences with the socioeconomic and demographic variables, Fernandez et al (2005) found that preferences over short commuting time and good schools are usually correlated with people who are married, with children, and under 40 years old. Households with less than \$75,000 annual income value social comfort more importantly than those with more income do. Race and education did not appear to be significant.

2.3.2 Neighborhood attributes

Neighborhood attributes that potentially affect residential location choices may include average or median level income, crime rate, ethnicity composition, air quality, school quality, pedestrian safety, density and openness of built environment, and open space for children. These neighborhood attributes, in fact, reflect a household's demand for quality of life. The consideration of crime rate and pedestrian safety relates to the purpose of obtaining personal safety. Air quality, density and openness of built form, and good areas for children reflect the

household consideration about obtaining a good natural environment. The school quality represents the concern about children's education.

School Quality

Barrow (2002) examined the effect of school quality on a household's residential location choices for Caucasians and African Americans. In his paper, school quality is measured by average SAT scores. On average, Caucasians households with children are more likely to live in an area with higher SAT scores than Caucasians households without children. Caucasian households' attitudes towards school quality are similar regardless their income level. This result, however, doesn't hold for African American households, who on average put less weight on school quality than Caucasians ones. Further analysis indicated that the poorest, the least educated, and the youngest population among African Americans primarily bring this difference. Interestingly, other African American households with children, higher income and education put similar weight on the school quality compared to corresponding Caucasians.

Land Use Pattern

Studies show that mixed land use is associated with a higher rate of non-automobile use. Most recently, land use has been viewed as a promising tool to reduce vehicle miles traveled. Through creating neighborhoods with clustered activities and quality transit services, city planners hoped that people would be attracted to such neighborhoods and reduce vehicle miles traveled. The empirical studies that analyze the relationship between land use patterns and residential location choices, however, do not show a strong correlation between the two. Guo and Bhat (2006) found that diverse land use patterns don't influence residential location choices except for families without a car. The higher the income a household has, the more likely it would locate in residential-oriented neighborhoods instead of mixed land use neighborhoods. Those who are attracted to mixed neighborhoods usually have no car or only one car. Guo and Bhat's results suggest that mobility plays an important role in the selection of neighborhoods. Families with good mobility do not value the advantages of being in mixed neighborhoods, while families with poor mobility are probably constrained to living in those neighborhoods. However, the above

correlation between low mobility and mixed land use may also result from the case that families who prefer mixed land use opt to have no or fewer cars. In this case, the mobility does not play a role in residential relocation. In reality, it may be a mixed situation, where some households with low mobility may want to locate in mixed land use neighborhoods, while other households with mobility may decide to sell their cars after relocating to mixed land use neighborhoods.

Population Density

Many research results (Bhat and Guo, 2004, Srour et al., 2002, Ben-Akiva and Bowman, 1998, Abraham and Hunt, 1997, etc.) suggest that households are more likely to locate in high population density area. The underlying reason for this might be that areas with high density are more established and offer more opportunities and accessibilities. In other words, population density may be a proxy of other variables. Not all studies support this result. Lerman (1975), for example, found that households prefer areas with low population density.

Crime Rate

Several studies (Ben-Akiva and Bowman, 1998; Nechyba and Strauss, 1998) found that crime rate has a negative impact on the probability of people choosing that neighborhood. Sermons and Koppelman (2001) suggested that this impact is not universal for all households but households with children only.

Neighborhood Housing Attributes

Some studies (Bhat and Guo, 2004, Srour et al., 2002, Molin, et al., 1999, Abraham and Hunt, 1997, Timmermans, 1996) investigated the effect of neighborhood dwelling type, average housing price, and average house size. They found that households prefer neighborhoods with detached houses to other housing types (Molin, et al., 1999). As expected, average housing price has a negative impact on the selection of the neighborhood (Abraham and Hunt, 1997), while larger housing space is preferred over smaller space (Bhat and Guo, 2004, Srour et al., 2002).

Open Space

A number of studies (Clark and Ware 1997, Farelly 1995) found a strong tendency of high-income households living in suburban or rural areas, while low-income population households live in cities. According to Brown et al. (1997), the preference toward suburban or rural areas has been unchanged in the past three decades. Brown used a household survey whose target population is nationwide. It was collected in November 1972, September-December 1988, and September 1992- February 1993. The sample sizes in three surveys are around 1477, 1233, and 1307 respectively. In 1988, about 30% of respondents preferred to live in a city with 50,000 people or more, 50% preferred a smaller city, a town, or a rural area within 30 miles of a larger city, and 20% preferred a small place with more than 30 miles from a city. These proportions of location preference didn't differ too much in three surveys. Also, most people are satisfied with their current residence locations, and those who are not are most likely to prefer living in smaller and/or less dense locations. Tu and Goldfinch (1996) suggested that single-person households are indifferent about the presence of a garden.

Geoghegan (2002) suggested there are two types of open space around a residential parcel: permanent open space, such as parks and lands that have conservation buildings; and developable open space, such as privately owned forested land and agricultural land. The preference toward open space can be observed by the high price for adjacent properties. All else being equal, people would prefer locations with more open space. Furthermore, permanent open space is valued more than developable open space.

Suburban Residents

Davis (1993) divided residents in the Portland, Oregon region into four categories: economy minded, family oriented, affluent, and long-distance commuters. Households that belong to economy minded have less than \$50,000 annual income, buy smaller lots, buy the least expensive homes, live farther from downtown, and have fewer children. These households prefer owning and living in rural area. Family oriented households are mainly two-earner families with children, living relatively close to downtown, and have larger lots than economy minded

households. Affluent households earn at least \$50,000 per year, have larger, more expensive properties, and are generally two-career managerial and professional couples without children. These households live closer to downtown than the other three categories of suburban residents. Long-distance commuters own the largest lots and pay the lowest prices for their lots. Their work trips are twice as long as the average of all four types. They tend to have only one worker in the family, have children, and strongly value living in a rural environment and having a large lot. These households tend to hold blue-collar, technical, or sales jobs and earn \$20,000 to \$50,000 per year.

2.3.3 Accessibility Attributes

Accessibility, defined as the ease of reaching opportunities, measures the connectivity between one location and the rest of the locations in the region. Connectivity may be measured depending on the opportunities of interest, for example, employment, retail and shopping, and recreation activities. Geographies in local and outlying areas can affect accessibility to opportunities.

Accessibility to Employment

In general, two types of measures have been applied to compute work accessibility. One is commute time. Commute time represents one-to-one (one residential location to one work location) accessibility for work opportunities. The second type of work accessibility is the accessibility index to employment, which represents the ability to access all potential job opportunities. This is a one-to-many (one residential location to many potential employment opportunities) measure.

Everything else being equal, residential locations with good accessibility to work are preferred by households (Alonso, 1964). Recent studies support this notion (Guo and Bhat, 2004; Srouf et al., 2002; Molin, et al., 1999; Waddell, 1996; Timmermans et al., 1996; Hunt et al., 1994) and they suggest that accessibility to employment has a positive effect of selecting a location. On average, high level of employment within 30 minutes of travel time from a location is associated with a high probability of being selected (Srouf et al., 2002). The longer the commute time is, the

smaller the probability that someone will select the location. Some studies found that households with different characteristics have varying preferences toward accessibility to employment. Guo and Bhat (2004, 2006) showed that households with over \$100,000 in household income tend to locate in areas closer to regional employment centers. This may be because they can afford a higher housing cost in those areas. Single-person households were found to locate closer to regional employment centers. As we have found, single-person households tend to rent apartments in urban areas. Childless couples tend to locate in areas with good access to shopping opportunities. Bina et al. (2005) found that those who view commute time as more important tend to be female, non-Caucasian, highly educated, and childless. Female household members may have more family responsibilities and thus prefer to select jobs with short commute distances and short working durations. Non-Caucasians on average have lower incomes and lower vehicle ownership, which encourages them to locate in urban areas and live closer to employment. Households without children usually don't have to consider children's preferences and can locate in areas closer to employment centers and have smaller housing size and costs than households with children. Two explanations exist for the shorter commute distance of female family members. First, females usually have more domestic responsibility and are more likely to tradeoff commute distance to accommodate their domestic role. Second, females are more likely to have low-paid jobs, which can't justify long commute distance.

Studies (Guo and Bhat, 2006; Abraham and Hunt, 1997; Waddell, 1996) suggested that husbands place less value on accessibility to work than wives. Sermons and Koppelman (2001) found that the difference in the preference toward commute time between wife and husband is the most profound in households with children. Freedman and Kern (1997) found that presence of children reinforces the impact of commute time for wives but not for husbands.

Accessibility to Shopping

While the findings on accessibility to employment are largely positive, empirical results on the accessibility to shopping are mixed and inconsistent. Some studies found that households prefer locations with good access to shopping (Guo and Bhat, 2004, Timmermans et al., 1996), while others found that accessibility to shopping is negatively associated with the probability of

selecting a residential zone (Srour et al., 2002, Hunt, et al., 1994). For example, Srour et al. (2002) found that the probability of selecting a Traffic Analysis Processing (TAP)¹ zone as the residential location is negatively associated with the retail employment within 30 minutes of travel time from the TAP. The contradictory results may come from the different accessibility measurements used by authors. Srour et al. (2002) used total retail employment within 30 minutes as a measurement of accessibility to shopping. Guo and Bhat (2004) measured the accessibility to shopping with a weighted linear function of the number of retail employment in

different zones: $A_i = \frac{1}{N} \sum_{j=1}^N \left(\frac{(\text{Number of Retail Employment}_j)^\gamma}{(\text{Impedance}_{ij})^\beta} \right)$, where A_i is the accessibility

to shopping for zone i , N is the total number of zones in the study region, Impedance_{ij} is a composite highway auto impedance measure of travel between zone i and zone j , and β and γ are coefficients obtained from a destination choice model (Guo and Bhat, 2004). Consider the land use pattern in which the commercial zone is surrounded by residential zones, with more residential zones located closer to the commercial zone. With the measurement used by Srour's study, we find much higher shopping accessibility in commercial zones and very low shopping accessibility (very low retail employment in residential zones) for most residential zones. This is because this method only calculates the retail employment within 30 minutes of travel time. The accessibility to shopping for those residential areas that are more than 30 minutes away from commercial zones would be the same, even though some zones are closer to commercial zones. The relationship, therefore, is negative between accessibility to shopping and the probability of being selected as a residential zone. However, if we use Guo and Bhat's method, the accessibility to shopping would decrease rather than remain static when the distance to commercial zone increases. In this case, we will find that higher shopping access would lead to higher probability of being a residential zone. Therefore, the relationship between shopping accessibility and the probability of being selected as a residential zone would be positive.

Preferences toward accessibility to shopping vary for people with different characteristics. Bina et al. (2005) found that being older or Hispanic, having few workers, and living with family members are associated with a higher value placed on accessibility to shopping. Guo and Bhat

¹ Defined by North Central Texas Council of Government.

(2006) found the attractiveness of shopping accessibility diminishes as household income increases. Lerman (1975) suggested that single-worker households are more interested in the level of shopping accessibility than multi-worker households are. Tu and Goldfinch found similar results: single young-person households prefer to live near shopping centers while young couples prefer to living far from the shopping areas.

Accessibility to Recreational Activity

Accessibility to recreational activities is rarely included in empirical residential location choice studies. Guo and Bhat (2006) found that accessibility to recreation doesn't have a significant impact. Colwell et al. (2002) used the differential equation approach to investigate how people trade-off between accessibility to employment and recreational activities. They based their study on a set of assumptions: identical job skill and time budget, and only one recreational place: beach, no intra-urban spatial variation and only two alternatives (city or beach). Based on these assumptions, Colwell et al. (2002) found that (1) the stronger the taste for recreation, the greater the attraction of living in close proximity to the recreation site; (2) city dwellers generally spend less time recreating than their beach dwelling counterparts; (3) if the same level of recreation consumption suddenly entailed shorter or more frequent visits to beach, some consumers living in city would prefer living at beach; and (4) if the same level of recreation consumption suddenly entailed higher housing costs and other expenses for recreation, some consumers living in the city would prefer living at the beach in order to avoid that component of recreation cost.

Accessibility to Transit

Access to public transit can have an effect on residential choices. Timmermans et al (1996) found that households prefer locations with shorter walking distances to transit. Hunt et al. (1994) found residents who already live within a walking distance of light rail transit perceive accessibility to transit more than twice as important as other households do. Molin et al. (1999) found that neighborhoods with higher bus frequencies have higher propensities to be selected than others. These studies suggested that the impact of transit accessibility might be larger for those transit users than those who mainly use vehicles due to residential self-selection. In other

words, travelers who prefer driving would not value the transit accessibility as important as those who prefer public transit.

Telecommuting

Telecommuting refers to the use of information technology at home or at a telecenter to partially or completely replace daily trips to and from regular workplaces (Nilles et al, 1976). In theory, the ability to telecommute widens one's temporal and spatial space and thus encourages the decentralization of residential locations (O'Sullivan 2000).

The empirical results, however, do not always support the above-mentioned hypothesis. Tayyran et al (2003) found significant correlations between telecommuting behavior and suburban residential location choice. They argued that "being able to telecommute is viewed more positively in the suburbs than in the central city, leading to higher probabilities of residential choices in those areas", which in turn "induce the decentralization of urban land development pattern".

On the contrary, using a survey data of 50,000 households from the Census Bureau's Current Population Study, Ellen and Hempstead (2002) investigated the impact of telecommuting on residential location choices. They found little evidence that telecommuting will lead to decentralization. Instead, current telecommuters appear less likely to live in rural areas and somewhat more likely to live in larger metropolitan areas. They admitted that a panel study of telecommuters' residential locations would be a better approach to address whether telecommuting will accelerate decentralization or not.

2.3.4 *Brief Summary*

Residential location decisions can be viewed as the process in which households weigh their preferences between attributes in order to maximize their total utilities. Three categories of attributes are important in a residential location decision: household socioeconomic and demographic attributes, neighborhood attributes, and accessibility attributes. Neighborhood

attributes define the quality of a neighborhood. Accessibility attributes, on the other hand, gauge the extent to which individuals are able to access various types of opportunities.

Past studies about residential choice have covered a wide variety of social-economics and demographics, neighborhood and accessibility attributes. We now have a better understanding about how households tradeoff between neighborhood and accessibility attributes. The studies also shed light on how households with different socio-economics and demographical attributes could have different attitude towards neighborhood and accessibility attributes.

The majority of the existing studies view location choice as a static choice, with no memory of the past. In reality, any residential choice is a relocation in space, a movement from one place to another. Given that human beings are highly adaptive, it is possible that previous location influences the current location choice. This perspective however is missing in existing studies.

2.4. Gap and My Research Goal of Location Study

The existing literature has made important contributions in understanding how people make residential location choices. Despite this, the current literature is deficient in accounting for the historical deposition influence on location choice. The majority of the existing studies analyzed residential location choices with no respect to previous locations. In other words, the previous studies did not investigate the influence of previous location in the residential relocation process.

In the existing residential location literature, the studies assumed that alternative locations are independent from each other. This assumption is in contrast to a nonprofessional's observation that areas clustered together often share common characteristics. This phenomenon is particularly true for accessibility, because by definition, a shorter distance between two zones results in a higher correlation with accessibility. Existing studies handling spatial correlation in residential location decisions are rare. The few recent studies assume constant spatial correlation between adjacent zones and zero correlation between nonadjacent zones, regardless their respective distance (Bhat and Guo, 2004; Bekhor and Parshker, 2008). I propose an alternative assumption by specifying spatial correlation as a function of distance between zones and by developing a distance-based GEV model.

In this study, I will use attributes of prior locations to measure the historical deposition influence and explore its interaction with lifecycle by sketching a conceptual framework and empirically quantifying it using the Puget Sound Panel Dataset. In particular, I am interested in two kinds of households: childbearing and non-childbearing, and am concerned about four types of accessibilities: accessibility to work, retail, open space, and recreational opportunities.

Chapter 3 Literature Review of Behavior Dynamics

3.1. Linking Residential Relocation and Behavior Dynamics

Hägerstrand (1970) noted there are three types of constraints that limit one's activity and travel pattern: capacity and coupling constraints, and authority constraints. Both the capacity and the authority constraints reflect the limit that the built environment puts on people's activity and travel patterns. Put simply, a change in the built environment would trigger changes in daily activity and travel behavior. I define traveling-related behavioral changes resulting from a trigger as "travel behavioral dynamics". There are two types of questions involved in travel behavioral dynamics. One is about the magnitude and the direction of a change. The other relates to response lags or response leads. The study of the magnitude and the direction of a change can refer to topics such as changes in mode split (e.g., Goodwin, 1989), changes in the number of daily trips, changes in daily vehicle miles traveled (e.g., Krizek, 2003), and changes in car ownership (Kitamura, 1989; Golob, 1990). The study of response lags or leads investigates the time between a trigger and the resulting behavioral change. In a response lag case, the trigger precedes the behavioral change, while in a response lead situation the order is reversed.

The last several decades of research in behavioral dynamics has almost exclusively focused on the magnitude and the direction of the behavioral change. A substantial amount of empirical studies have examined the magnitude and the direction of a change regarding different types of behavior (Kitamura, 1989; Krizek, 2003), different levels of observations (aggregate vs. disaggregate) (Goodwin, 1989; Golob, 1990), and different types of independent factors (Ma and Goulias, 1996, 1997; Goulias, 1999). A number of reviews and editorials have been written (Pas, 1985, 1996; Pendyala and Goulias, 2002; Kitamura, 2002; Pendyala, 2003), in terms of the modeling advances, about the various datasets collected and the amount of behavioral insights learned.

Little has been done in investigating issues related to response leads or lags. Kitamura (1990) suggested that a behavioral change is often not instantaneous and there exists a “response lag” or “response lead” between the occurrence of an external stimulus and a behavioral change. Response lag refers to the time lag between a trigger (e.g., a change in home location) and a response. Response lead refers to the situation that behavior change occurs in advance of a trigger. Several reasons may explain the existence of a lag. The individual may not immediately acquire the information of a trigger. Alternatively, the individual may ignore the initial and often small magnitude of a trigger until it accumulates over time to a significant threshold for the individual to perceive (Coombs, Dawes, and Tversky, 1970). In other cases, when a behavioral change involves large financial cost, the constraints of family resources may lead to a response lag. Finally, individuals may delay their behavioral responses until the next trigger (Clarke et al., 1982). This next trigger might be a change in home or job location. The association of the behavioral response with this next trigger is purely psychologically motivated. One can interpret it as the result of one’s resistance to change, or the habit of doing the same old thing. The influence of habit on travel behavior, in particular mode choice behavior, was recognized and demonstrated in several papers (Banister 1978; Hensher and Johnson 1980).

It is important to sort out these underlying reasons responsible for the response lag. Current studies pay insufficient attention to this aspect and consequently unsuitable policies may be put forth due to our lack of information on response lag. A single person with few social obligations may make a change on the second day of a policy implementation, while a married woman, who must plan accordingly before she can make changes, may not react until the second year. In the first case, the response lag is short and there is a single trigger, which is policy change. In the second case, the trigger is still policy change but the response lag is much longer because the subject is unable to make the change until she is relieved from family and social obligations one year later. Without understanding response lag, the individual in the second case might be viewed as unresponsive to the policy change.

The interest of the behavioral change in my dissertation is on the response lag following a significant change in discretionary time use. Three reasons contribute to the decision to focus on discretionary time use. First, the argument that discretionary activities bring joy to people is

undisputable. Unlike mandatory (i.e., work) and maintenance activities (i.e., eating, sleeping) that are conducted from a need to earn sufficient means and to maintain a normal life, discretionary activities exist because people enjoy performing them. Therefore, from a time use point of view, *ceteris paribus*: the more time spent on discretionary activities, the happier the person is. Secondly, discretionary activities relax one's mind and many of them require physical movement. Hence, from a health perspective, discretionary activities likely have a positive impact on health. Third, compared with mandatory and maintenance activities, discretionary activities are a category where an individual has the most autonomy.

It is reasonable to hypothesize that a change in one's discretionary time use does not happen instantaneously after the occurrence of a trigger. Response lags are expected due to various factors. A home or job relocation often involves a change in the built environment. If the change were large enough to trigger a behavioral response, we would expect a relatively short response lag. If this is true, it supports the "accumulative effect hypothesis." In other cases, a response lag occurs due to one's family and social obligations, or one's temporal constraints. Thus, if we observe significant effects from the built environment and the family, social or temporal constraints, it suggests to us that the built environment is acting as a trigger, but that the response is delayed due to these constraints. If significant effects are only observed from those constraints, it suggests that the forces regulating one's schedule play a more important role than the built environment. The "psychological delay hypothesis" states that a behavioral change is purposefully delayed until the next trigger, such as a change in home or job location. If this hypothesis is true, we should expect a negative link between a home/job relocation dummy and a response lag.

This dissertation does not deal with the issue of response lead. Conceptually, I feel that it is unlikely for a person to change his/her time allocation prior to location residence change, if indeed the residence change is the cause and the time allocation change is the effect. This is because one's time allocation is highly constrained by one's temporal and spatial environments (Hagerstrand, 1970).

The rest of this chapter is organized as follows. In Section 3.1, I will provide a review of past studies about the behavior dynamics using panel datasets, the major part of which is studies on the magnitude and the direction of a change. The review focuses on the study of panel datasets, because they provide us with sequential observations of trigger and response. We can infer a causal relationship with more confidence than with a cross-sectional dataset. My review on past studies in this paper therefore is limited to empirical studies that have used panel datasets to answer questions related to behavior dynamics. Despite studies on the magnitude and direction of behavior change, the studies about the response lead and lag are rare. My dissertation, therefore, focuses on response lag of a behavioral change. This will be discussed in Section 3.2.

3.2. Past Studies About the Behavior Dynamics

Analyses are based on cross-sectional datasets compared between observed states. Operationally, it means that a variety of behaviors observed at a specific time are compared against a set of states (described by a set of factors) observed at the same point in time. If a statistically significant and theoretically acceptable relationship is observed between the two, we can conclude that the factor has been found important in affecting the behavior of interest. Such analysis is static because it implicitly assumes that behavior happens instantaneously, and thus omits the dynamics that could possibly take place between a trigger and the actual behavioral change (Clarke et al. 1982; Kitamura 1990). Many studies have shown that behavioral responses lag. For instance, cross-sectional studies often show that households with young children make fewer transit trips than those without, due to the inconvenience of using transit when young children are present. Goodwin (1989) discovered that those households do not reduce their transit trips immediately after the birth of their first child, even though the birth of a child does eventually lead to a reduction in transit trips. Dynamics of travel behavior can be studied at three levels (Clarke et al. 1982). At the finest level, it is the microdynamics of people's 24-hour activity and travel pattern. These are individual choices, made out of one's preferences, subject to constraints and possible randomness. The next level is a macrodynamic modifier, which addresses behavioral responses that can lag or lead a stimulus several years later or earlier. The highest level of dynamics is the macrodynamic process, which is a lifelong process in which

people, with different cohorts and exposure to unique socio-demographic environments, age through lifecycle stages.

There are clearly interactions between levels of dynamics. Travel cost and speed may strongly influence an individual-level mode choice and delay it until the next life shock. This scenario involves three levels of dynamics. The influence of mode choice behavior by travel cost and speed is microdynamic at the individual level within a 24-hour span. The delay in the mode choice change is a microdynamic modifier, which may happen several years later. Last, the association of the delay with a life-shock event (e.g., a home or job location change or a change in lifecycle) situates the mode choice change in the aging process, which is uniquely defined by one's socio-demographic environment.

In this section, I am going to review the studies using two panel datasets: The Dutch Mobility Panel (1984-1988) and Puget Sound Panel Dataset (1989 to 2002), which are most commonly used in tracking behavior dynamics. The Dutch Mobility Panel was initiated in 1984 in The Netherlands, involving weekly travel diaries, as well as household and personal questionnaires collected at biannual and annual intervals. All household members over eleven years old completed the travel diaries. Initiated in 1989, the Puget Sound Panel Dataset is similar in design to the Dutch National Mobility Panel, but also adopted experience from the long line of cross-section urban travel surveys. It focuses more on transportation and transit policy issues, involving two-weekday travel diaries, and household and personal questionnaires collected at annual and biyearly intervals.

Behavior Dynamics Studies Using the Dutch Mobility Panel

Kitamura (1989) examined the causal relationships between car ownership, car use, and transit use using the first three waves of the Dutch Mobility Panel dataset. Although the data showed no significant relationship between car ownership and households' total trip generation, car ownership had significant impact on household mode use. As pointed out by Kitamura (1989), car ownership has a strong association with the "number of driver trip segments, total driver trip time expenditure, number of bus, tram and metro trip segments, and number of total transit trip

segments.” In order to understand the causal linkages among changes in car ownership, number of licensed drivers, and car use, Kitamura (1989) performed a cross-classification analysis. Four causal relationships are assumed for hypothesis testing: 1) a change in car ownership or being a licensed driver leads to a change in car use, which then leads to a change in transit use; 2) a change in car ownership or becoming a licensed driver leads to a change in transit use, which leads to a change in car use; 3) a change in car ownership or becoming a licensed driver leads to changes in car use and transit use, but the changes in car use and transit use are unrelated; 4) a change in car ownership or becoming a licensed driver leads to changes in car use and transit use, and the latter changes are related. It is suggested that the above four causal relationships may all exist and that the statistic testing result was not trying to reveal which hypotheses are accepted and which are not, but instead reveal which relationships are more representative than others.

Kitamura found that the number of car trips is negatively and significantly correlated with transit trips, suggesting the substituting relationship between the two. Higher levels of car ownership or more drivers in the household lead to more car trips and consequently fewer transit trips. The directional effect from transit trips to car trips, on the other hand, is rather weak. Compared with the number of drivers, the number of cars is a more significant factor for the number of car trips. The result implies that improving transit services may not lead to reductions in car use; instead, adopting policies to control car ownership may be more effective.

Using the same panel dataset, Goodwin (1989) investigated the relationship between a change in a circumstance (including life-cycle², employment status³, and income) and the change in mobility (car ownership). Using the first wave of the Dutch Mobility Panel Dataset as a cross-sectional dataset, Goodwin found that although both the cross-sectional dataset and the panel dataset reveal a negative relationship between public transport use and car ownership, and a positive one between car ownership and income, these relationships are much weaker in the panel dataset than in the cross-sectional dataset. Goodwin’s study also confirmed the existence of a lag between a change in the circumstance and the change in behavior. For instance, some

² In Goodwin (1989), life cycle is defined as the combination of characteristics reflecting the life stage of family. There are five life cycles are defined: marriage stage, birth of children stage, children growing up stage, children leaving home stage and retirement stage. See Goodwin (1989, P.122) for more details.

³ Six groups according to employment status: employed, unemployed, self-employed, housework, school and student, and retired.

people who change their life cycle from the childbearing stage to the children growing up stage do not change their usage of public transport in a short time. Goodwin suggested the response lag is about 5-15 years.

The existence of response lag was also noted by Golob. In his 1990 paper, he analyzed the contemporaneous and lagged relationships between car ownership and travel time by different modes using the Dutch Mobility Panel Dataset. Golob (1990) postulated four sets of contemporaneous causal relationships with regard to the number of cars, travel time by car, travel time by public transport and travel time by nonmotorized modes. 1) The number of cars has a positive effect on travel time by car and a negative effect on travel time by public transport and non-motorized modes; 2) given the number of cars available, travel time by car has a negative effect on travel time by public transport and non-motorized modes; 3) travel time by public transport has a positive effect on travel time by car; this relationship is postulated because households with a higher level of travel time by a slower mode may switch to faster mode in order to reduce time expenditures; 4) public transport travel time has a positive effect on the travel time by non-motorized modes. In addition to these four contemporaneous relationships, two lagged relationships are assumed: travel time by public transport has a positive lagged effect on future car ownership, and travel time by car has a positive lagged effect on future car ownership.

The estimation results reveal that all the postulated relationships are significant except for the contemporaneous effect of travel time by car on travel time by public transport. In addition to these relationships, Golob (1990) also found significant positive lag effects of the high-income dummy variable and the number of drivers on future car ownership. The study also found that there are significant positive lead effects between future high income and future number of drivers on present car ownership. In other words, if a household expects an increase in income or the number of drivers, the household may purchase the vehicle in advance. In other words, a higher car ownership may be a result of higher income or higher number of drivers, or it may be a result of anticipated future higher income or higher number of drivers.

As a summary, the principal conclusion from this 1990 study was that car ownership and travel time by car, public transport, and non-motorized modes are inter-related. Therefore, efforts on modeling only one of these variables will subject the data to an endogeneity bias. In addition, these interrelationships are not only contemporaneous but also dynamic involving leads and lags.

Behavior Dynamics Studies Using Puget Sound Panel

Ma and Goulias (1996, 1997) may be the one of the first researchers who utilized the Puget Sound dataset to investigate travel behavior dynamics related questions. In particular, they investigated the transition of travel pattern and activity pattern between waves. In their 1996 study, the activity patterns are defined by the combination of total activity time, frequency of mandatory, maintenance, and discretionary activities. The travel patterns are defined by the combination of total travel time, average length of trip chains, and frequencies of trips made by different modes. Using cluster analysis, four daily activity patterns – non-motorized, car/carpool, immobile, public transit - are classified. The non-motorized group is highly likely to make trips by non-motorized/other modes and is very efficient in trip chaining. The car/carpool group is most likely to use car and carpool as a travel mode, but less likely to use public transport or non-motorized mode. They have highest total travel time among the four groups. The immobile group has the shortest travel time and very low trip frequencies. The public transit group has the highest trip frequency by public transport but lowest number of trips by car and carpool. They also found four activity patterns in the Puget Sound Panel named “Worker-A, Shopper, Worker-B, and Inactive”. The Worker-A group has not only very high frequency and duration in mandatory activities, but also long total daily activity duration. The Worker-B group has very high duration in mandatory activities, but with far fewer frequencies. The Shopper group has the highest frequency and longest duration in maintenance and discretionary activities. Finally, the Inactive group has the lowest activity frequencies and activity duration.

In a follow-up 1997 study, they tried to reveal how people switch between these activity patterns and travel patterns. It is found that the Worker-B group is less likely to change the frequency and duration of activity, compared to Worker-A and Shopper. The Inactive group is most likely to change the frequency and duration of activity among the four activity pattern groups. The

Immobile group is also least likely to change their travel pattern in terms of trip frequency by mode, number of trip chains, and total travel time, while the Car/Carpool group is most likely to change.

Goulias (1999) extended the Ma and Goulias (1997) study on the analysis of activity and travel pattern dynamics by using mixed Markov latent class models. For day-to-day dynamics, the Inactive group has the lowest transition rate compared to other groups, suggesting that they may be satisfied with their activity pattern. People in the Shopper group have highest transition rate. Compared with activity patterns, travel patterns have a lower transaction rate, which indicated that people are more likely to modify activity patterns in day-to-day dynamics. For year-to-year dynamics, it was found that the transition rates of travel and activity are all higher than corresponding part of day-to-day dynamics.

Goulias (1999) indicated that two reasons could explain why year-to-year transition rates of activity and travel patterns are higher than they are from day to day. First, the day-to-day transaction rate cannot account for potential cyclic behavior (with a cycle shorter than a year). Second, some changes that could only change in a longer time frame such as residential relocation, workplace relocation, income increase or lifestyle stage transition, could change the activity and travel pattern significantly, and these changes could only be captured by a year-to-year transition.

Although Ma and Goulias analyzed the transition of travel behavior between waves, and the underlying reasons on exactly what causes were not investigated. One attempt of linking behavior dynamics and behavioral triggers is the study by Krizek (2003), in which the change of the built environment is viewed as the trigger. In his 2003 paper, Krizek used the Puget Sound dataset to investigate how relocations affect travel behavior. His dependent variables included vehicle miles traveled (VMT), number of tours, and number of trips per tour. For exogenous variables, he used both baseline values (values prior to the residential move) and change values, which is the difference between the value after and prior to the residential move. He found that the greater the baseline value is, the greater the resulting change. The households that shortened their commute distance are more likely to participate in more tours during the day. Regarding the

built environment, Krizek found that households that relocate to neighborhoods with high accessibility would reduce the vehicle miles traveled, increase the number of tours, and decrease the number of trips per tour. He also found that households originally located in areas with higher neighborhood accessibility have a greater propensity to decrease their vehicle miles traveled and the number of trips per tour, while increasing the number of tours.

Since all three constraints—capacity, coupling, and authority (Hägerstrand, 1970)—could influence the activity and travel pattern, what constraints might have a larger impact on behavior dynamics? Chen and Kwan (2006) answered this question from a perspective view of individual-based accessibility. They investigated how personal space-time accessibility changes with adjustments in the built environment using the Puget Sound Panel Dataset. In their paper, space-time accessibility is calculated based on observed individuals' daily travel patterns; therefore, the change in the accessibility reflects the change of activity and travel pattern. For this reason, an increase in accessibility is equivalent to an increase in travel and activity frequencies in this study. It was found that people who relocate from an area with a sparse transportation network to an area with a high-density network would not necessarily result in a significant increase in accessibility. Regarding socio-economic characteristics, a jump in the family income from below \$75,000 to equal to or above \$75,000 will result in a significant jump in individual accessibility. Children are found to have significant constraints in the accessibility. Reducing the number of children under 17 years old would lead to an improvement in accessibility, while increasing would do the opposite. Their results indicated that the built environment variables play a less important role than the variables characterizing the spatial and temporal constraints faced by the individual.

3.3. Gap and My Research Goal of Behavior Dynamics

Regarding behavior dynamics, what is missing from the large body of the literature is response lead/lag. Probably one reason for the lack of attention to the time question is the existence of vast number of cross-sectional datasets. Cross-sectional datasets are incapable of dealing with behavior dynamics because all inputs (the external stimulus and a behavioral change) are collected at a single point of time. Models estimated on cross-sectional datasets are static models

in that researchers can only relate those inputs (collected at a single time point) together and make inferences on the relationships.

To deal with behavior dynamics, a panel dataset is necessary. Panel datasets track people's behavior over time so that information on the external stimuli can be collected at a time prior to when the information on behavioral change is obtained. In this case, the information on the time between the occurrences of the external stimuli and behavioral change can be calculated, and inferences may be made in terms of what affects the response lag. In addition, with panel datasets, it also allows us to separate the effect of time progression from a structural change caused by an external stimulus (Greene, 2003).

I will use the Puget Sound Panel Dataset collected between 1989 and 2002 in this study. A behavioral change is an adjustment in the time allocated to discretionary activities and associated travel. The main question I hope to answer in this paper is what is responsible for the response lag of a significant change in discretionary time use: the built environment, family and social obligations, temporal constraints, or a psychological delay factor. The built environment is connected to the accumulative effect hypothesis. If this hypothesis were true, a large change in the built environment, brought by a home relocation, should result in a relatively short response lag. Family, social, and temporal constraints may complicate the length of a response lag. Thus, if I observe significant effects from the built environment and the family, social, or temporal constraints, it implies that the built environment is acting as a trigger, but the constraints are delaying a response. If significant effects are only observed from those constraints, it suggests that the forces regulating one's schedule play a more important role than the built environment. The psychological delay hypothesis states that a behavioral change is purposefully delayed until the next life shock. In this study, I define this next life shock as the next home or job relocation. If this hypothesis were true, I shall expect a negative link between a home/job relocation and response lag.

Chapter 4 Theoretical Framework and Methodology of

Residential Location Model

4.1 Introduction

The discrete nature of residential alternative locations dictates the use of a discrete choice model. In this chapter, I will provide a review of existing discrete choice models, point out their deficiencies, and propose methodology in the context of the historical deposition effect in residential location decisions. The rest of the chapter is organized as follows. In Section 4.2, I will describe the Random Utility Maximization framework from which discrete choice models are derived. In Section 4.3, I will review existing modeling frameworks. In Section 4.4, I will propose methodology and model configurations. In Section 4.5, I will describe the approaches for the measurement of neighborhood and accessibility attributes.

4.2 Random Utility Maximization Framework – A Theoretical Basis

The theoretical basis for discrete choice models is the Utility Maximization Framework, which assumes that a decision maker will choose an alternative with the maximum utility. Thurstone (1927) originally developed this concept in the study of psychological stimuli, in which the level of stimuli would determine the psychological response. Marschak (1960) interpreted stimulus as utility and derived a binary model based on the utility maximization principle. Discrete choice models based on the utility maximization assumption are also called Random Utility Models (RUM). The word “random” implies that there is an unobserved utility component that is known to the decision maker, but not to the researcher. This random term can encompass several sources of error, including unobserved attributes, unobserved taste variations, measurement of imperfect information, and instrumental variables (Ben-Akiva and Lerman, 1985).

A Random Utility Model (RUM) can be derived as follows. A decision maker faces a set of alternatives, with each alternative providing him/her a certain level of utility. Denote the total

utility that a decision maker obtains from alternative j as $U_{n,j}, j = 1, 2, \dots, C$. The decision maker will choose alternative i if and only if $U_{n,i} > U_{n,j}, i \neq j$. The total utility can be decomposed into two components: systematic utility $V_{n,j}$ and random error $\varepsilon_{n,j}$:

$$U_{n,j} = V_{n,j} + \varepsilon_{n,j}. \quad (4.1)$$

The systematic utility, $V_{n,j}$, can be observed and is specified as a function of the observed attributes $\mathbf{x}$, and the unknown parameters $\boldsymbol{\beta}$, which are to be estimated:

$$V_{n,j} = V(\mathbf{x}, \boldsymbol{\beta}). \quad (4.2)$$

Given that the decision maker faces C alternatives, there exist C error terms, or $\boldsymbol{\varepsilon}_n = (\varepsilon_{n,1}, \varepsilon_{n,2}, \varepsilon_{n,3}, \dots, \varepsilon_{n,C})$. Under the utility maximization assumption, the probability that alternative i is selected can be derived as follows:

$$\begin{aligned} P_{n,i} &= \Pr ob(U_{n,i} > U_{n,j}, \forall j \neq i) \\ &= \Pr ob(V_{n,i} + \varepsilon_{n,i} > V_{n,j} + \varepsilon_{n,j}, \forall j \neq i) \\ &= \Pr ob(\varepsilon_{n,j} - \varepsilon_{n,i} < V_{n,i} - V_{n,j}, \forall j \neq i). \end{aligned} \quad (4.3)$$

It is assumed that vector $\boldsymbol{\varepsilon}_n$ follows a probability density function $f(\boldsymbol{\varepsilon}_n)$. With this density function, I can calculate the probability in equation 4.3 by the following:

$$\begin{aligned} P_{n,i} &= \Pr ob(\varepsilon_{n,j} - \varepsilon_{n,i} < V_{n,i} - V_{n,j}, \forall j \neq i) \\ &= \int_{\boldsymbol{\varepsilon}_n} I(\varepsilon_{n,j} - \varepsilon_{n,i} < V_{n,i} - V_{n,j}, \forall j \neq i) f(\boldsymbol{\varepsilon}_n) d\boldsymbol{\varepsilon}_n. \end{aligned} \quad (4.4)$$

In equation 4.4, the probability is a multidimensional integral over $f(\boldsymbol{\varepsilon}_n)$, and $I(\bullet)$ is an indicator function, which equals to 1 if the expression in the parentheses is true and 0 otherwise.

The parametric approach to estimate β is to assume a density function for $f(\epsilon_n)$ and then calculate the above multidimensional integral based on this assumption. Then, equation (4.4) will become a central component of the log-likelihood function, which is to be maximized during estimation.

One typical feature of a RUM is that the scale of utility needs to be normalized, and the overall scale of utility does not affect the decision making. Consider two RUM models with utility functions being $U_{n,j}^0 = V_{n,j} + \epsilon_{n,j}$ (Model 1) and $U_{n,j}^1 = \lambda V_{n,j} + \lambda \epsilon_{n,j}$ (Model 2). The alternative with the highest utility in Model 1 will also have the highest utility in Model 2. If random errors are identically, independently distributed, the error variances are normalized to some number, which could be chosen for convenience. Since variances are the same, normalization sets all variances to the same number. For example, the error variances in a logit model are traditionally normalized to $\pi^2/2$. If the errors are correlated over alternatives, the normalization is applied to the error difference instead of the error term (see Train, 2003). A similar problem is applied when there are no constraints on the constant term. Consider the models with the utility functions for alternative i and j to be expressed as: $U_{n,i}^0 = V_{n,i} + \epsilon_{n,i}$; $U_{n,j}^0 = \alpha + V_{n,j} + \epsilon_{n,j}$ (Model 1) and $U_{n,i}^0 = \beta + V_{n,i} + \epsilon_{n,i}$; $U_{n,j}^0 = (\alpha + \beta) + V_{n,j} + \epsilon_{n,j}$ (Model 2). If alternative i has higher utility in Model 1, it would also have higher utility in Model 2. In the estimation, to avoid such problem, the constant for one of alternatives is set to zero.

4.3 Existing Model Structures

4.3.1 Multinomial Logit Model

The Multinomial Logit Model (MNL) is the most widely used discrete choice model. Luce (1959) derived the logit model from the *independence irrelevant alternatives* assumption (known as the IIA property). McFadden (1974) demonstrated that the expression for logit model implies that the random error is extreme value distributed.

In a MNL model, I assume that the unobserved random error has the following density function:

$$f(\varepsilon_{n,j}) = e^{-\varepsilon_{n,j}} e^{-e^{-\varepsilon_{n,j}}}, \quad (4.5)$$

and its cumulative distribution function has a closed-form:

$$F(\varepsilon_{n,j}) = e^{-e^{-\varepsilon_{n,j}}}. \quad (4.6)$$

Given this distribution, the probability of choosing alternative i can be expressed as (McFadden, 1974):

$$P_{n,i} = \frac{e^{V_{n,i}}}{\sum_j e^{V_{n,j}}}. \quad (4.7)$$

In practice, the observed systematic utility is usually specified as a linear in parameter function $V_{n,j} = V(\mathbf{x}, \boldsymbol{\beta}) = \boldsymbol{\beta}' \mathbf{x}$. It should be noted that the “linear in parameter” function can be used to reflect nonlinearity of some attributes. For example, if an attribute $x_{n,j}$ is believed to have a 2nd order effect, its specification in the systematic utility function can be $\beta_{j1}x_{n,j} + \beta_{j2}x_{n,j}^2$.

The MNL has several desirable properties. First, the probability $P_{n,i}$ is between zero and one, which is a necessary condition for probability expressions. Second, the sum of the probabilities of all alternatives is equal to 1: $\sum_{i=1}^C P_{n,i} = \sum_i e^{V_{n,i}} / \sum_j e^{V_{n,j}} = 1$. Third, the probability expression is closed-form and easy to apply. Most importantly, the log-likelihood function of the MNL model is globally concave if the utility is expressed as a linear in parameter function (McFadden 1974), which ensures that the estimated coefficients by a maximum likelihood estimation would be consistent.

The MNL is widely used due to its easily applied closed-form. However, it has limitations when facing three types of situations: taste variation, substitution patterns, and repeated choices over time (Train, 2003).

Taste Variation

Taste variation exists when decision makers value attributes differently. For example, households with many members might value housing space more importantly than a household with fewer members might. Households with young children may be more concerned about the quality of school than those without. The MNL model can capture systematic taste variations (taste variations that relate to observed attributes), but it cannot capture random taste variations (taste variations that relate to random errors).

Consider a logit model with two attributes x and y in the utility function. The coefficients for these two attributes vary over decision makers. The utility function for decision maker n for alternative j is expressed as:

$$U_{n,j} = \alpha_n x_n + \beta_n y_n + \varepsilon_{n,j}, \quad (4.8)$$

where α_n and β_n are coefficients for decision maker n .

The taste variation is represented by the variation of coefficients over decision makers. Assume that we recognize that the variation of coefficients is brought by another set of attributes w_n and z_n :

$$\alpha_n = \varphi w_n, \quad (4.9)$$

and

$$\beta_n = \gamma z_n. \quad (4.10)$$

Substituting equations 4.9 and 4.10 into 4.8, we have the following equation:

$$\begin{aligned} U_{n,j} &= \phi w_n x_n + \gamma z_n y_n + \varepsilon_{n,j} \\ &= \phi(w_n x_n) + \gamma(z_n y_n) + \varepsilon_{n,j}. \end{aligned} \quad (4.11)$$

Equation 4.11 is a MNL model with two recognized attributes $w_n x_n$ and $z_n y_n$, which demonstrate that an MNL model could capture systematic taste variation, or taste variation relating to observed attributes. However, if the taste variation is related to random errors, the MNL model is unable to capture it. Assume that variation of coefficients α_n and β_n are brought not only by recognized attributes w_n and z_n but also by certain random error ξ_n and ψ_n :

$$\alpha_n = \phi w_n + \xi_n, \quad (4.12)$$

and

$$\beta_n = \gamma z_n + \psi_n. \quad (4.13)$$

Substitute 4.12 and 4.13 into 4.8, we have the following equation:

$$\begin{aligned} U_{n,j} &= (\phi w_n + \xi_n) x_n + (\gamma z_n + \psi_n) y_n + \varepsilon_{n,j} \\ &= \phi(w_n x_n) + \gamma(z_n y_n) + \xi_n x_n + \psi_n y_n + \varepsilon_{n,j} \\ &= \phi(w_n x_n) + \gamma(z_n y_n) + \tilde{\varepsilon}_{n,j}, \end{aligned} \quad (4.14)$$

where $\tilde{\varepsilon}_{n,j} = \xi_n x_n + \psi_n y_n + \varepsilon_{n,j}$.

The covariance of $\tilde{\varepsilon}_{n,j}$ and $\tilde{\varepsilon}_{n,k}$ is not zero since

$\text{cov}(\tilde{\varepsilon}_{n,j}, \tilde{\varepsilon}_{n,k}) = x_n^2 \text{var}(\xi_n) + y_n^2 \text{var}(\psi_n) \neq 0$. In this case, the independence between the alternatives assumption, required by the MNL model, doesn't hold. Therefore, the MNL model cannot handle the taste variation relating to the random errors.

Substitution Patterns

Substitution pattern refers to the change in the odds ratio, which is the ratio of the probabilities of two alternatives. The substitution pattern reflected by the MNL is problematic when some alternatives share similar features. Consider any two alternatives i and k , the odds ratio is

$$\frac{P_{n,i}}{P_{n,k}} = \frac{e^{V_{n,i}} / \sum_j e^{V_{n,j}}}{e^{V_{n,k}} / \sum_j e^{V_{n,j}}} = e^{V_{n,i} - V_{n,k}}. \quad (4.15)$$

This odds ratio does not depend on attributes of any other alternatives. This property is called *independence from irrelevant alternatives (IIA)*.

The IIA property, in many cases, is not appropriate in reality. Consider a mode choice model with three alternatives: bus, walk, and vehicle. Assume that the choice probability for each of the three alternatives is the same: $P(\text{bus}) = P(\text{walk}) = P(\text{vehicle}) = 1/3$. The odds ratio of any two probabilities is one. It is assumed that transit (e.g., rail) is introduced and it has the same attractiveness as the bus in all features. Therefore, the ratio of probabilities of taking a bus and transit should be one. According to the IIA property, the ratio of probabilities of taking a bus, a vehicle, or walking will remain the same regardless the new transit mode. Then we have: $P(\text{bus}) = P(\text{walk}) = P(\text{vehicle}) = P(\text{transit}) = 1/4$.

This result, however, is not reasonable. Since both rail and bus belong to public transport and they share similar features, we would expect half of the travelers who used the bus would switch to rail, while travelers who walked or used a vehicle would not be unaffected by the newly introduced mode. The probabilities of these mode choices should be $P(\text{bus}) = P(\text{transit}) = 1/6$, and $P(\text{walk}) = P(\text{vehicle}) = 1/3$. In this example, the probabilities of choosing bus and transit are overestimated and the probabilities of driving or walking are underestimated because of the IIA property of the MNL model. The odds ratios of the three original modes: bus, walking, or

vehicle, should change because of the introduction of the transit mode, instead of remaining constant. The failure of IIA property comes from the correlation between bus and transit.

Repeated Choice over Time

In repeated choice scenarios, if unobserved random errors are independent over time, then the MNL model can be used in the same way as in a cross-sectional dataset. The utility that decision maker n obtains from alternative j at time t is:

$$U_{n,j,t} = V_{n,j,t} + \varepsilon_{n,j,t}, \quad (4.16)$$

where $\varepsilon_{n,j,t}$ is extreme value distributed, independent over n , t , and j . According to the MNL model, the choice probability for alternative j at time t is simply:

$$P_{n,j,t} = \frac{\exp(V_{n,j,t})}{\sum_k \exp(V_{n,k,t})}. \quad (4.17)$$

However, unobserved random errors are often correlated over time. In this case, a mixed logit model should be used instead of a MNL model.

4.3.2 Nested Logit Model

When the IIA property does not hold, a more appropriate model is the Nested Logit (NL) model. NL model allows correlation between alternatives in such a way that any two alternatives that are correlated with each other are grouped into the same nest, while any two alternatives that are not correlated with each other are grouped into different nests. In the NL model, the IIA property only holds within each nest. Using the same example in the previous section, assume that the utility for vehicles increases because of the improvement of level of service. Now $P(\text{vehicle}) = 1/2$, and walking and public transit split the remaining share: $P(\text{walk}) = 1/4$, and $P(\text{bus}) = P(\text{transit}) = 1/8$. Note that the odd ratios between nests (walking vs. vehicle,

walking vs. bus, walking vs. transit, vehicle vs. bus, and vehicle vs. transit) change, while the odd ratio within nest (bus vs. transit) remains the same.

The NL model assumes that all alternatives can be grouped into K non-overlapping nests denoted as $N_1, N_2, \dots, N_K$. The utility function still has the same form as the MNL model $U_{n,j} = V_{n,j} + \varepsilon_{n,j}$. NL model is different from MNL in that it assumes the distribution of $\varepsilon_{n,j}$ are correlated within nest, but uncorrelated between nests. Thus $\varepsilon_n = (\varepsilon_{n,1}, \dots, \varepsilon_{n,C})$ has the following cumulative distribution:

$$F(\varepsilon_n) = \exp\left(-\sum_{k=1}^C \left(\sum_{j \in N_k} \exp(-\varepsilon_{n,j} / \lambda_k)\right)^{\lambda_k}\right). \quad (4.18)$$

Parameter λ_k is a measure of the degree of independence within nest N_k . A value of 1 indicates complete independence within N_k . Given this assumption, the probability of selecting alternative j is:

$$P_{n,i} = \frac{e^{V_{n,i} / \lambda_k} \left(\sum_{j \in N_k} e^{V_{n,j} / \lambda_k}\right)^{\lambda_k - 1}}{\sum_{k=1}^C \left(\sum_{j \in N_k} e^{V_{n,j} / \lambda_k}\right)^{\lambda_k}}. \quad (4.19)$$

The above probability expression can be further split into the product of two probabilities: the probability of an alternative's locating in nest N_k (P_{n,N_k}) and the probability of choosing alternative i if the nest N_k is chosen ($P_{n,i|N_k}$). In a NL model, the observed systematic utility function can be split into two parts: $W_{n,k}$ which is constant for all alternatives within nest N_k , and $Y_{n,j}$ which varies over alternatives within the same nest. Mathematically, the probability that alternative j is chosen can be expressed as:

$$P_{n,i} = P_{n,i|N_k} P_{n,N_k}, \quad (4.20)$$

where,

$$P_{n,N_k} = \frac{e^{W_{n,k} + \lambda_k I_{n,k}}}{\sum_{k=1}^C e^{W_{n,k} + \lambda_k I_{n,k}}}, \quad (4.21)$$

$$P_{n,i|N_k} = \frac{e^{Y_{n,i}/\lambda_k}}{\sum_{j \in N_k} e^{Y_{n,j}/\lambda_k}}, \text{ and} \quad (4.22)$$

$$I_{n,k} = \ln \sum_{j \in N_k} e^{Y_{n,j}/\lambda_k}. \quad (4.23)$$

Function $I_{n,k}$ is called inclusive value and $\lambda_k I_{n,k}$ represents the expected utility received from all alternatives in nest N_k .

Compared to MNL, the NL model provides a more flexible setting for modeling correlation between alternatives. However, unlike MNL, NL does not hold the global concavity property.

4.3.3 Generalized Extreme Value Model

McFadden (1978) proposed the Generalized Extreme Value (GEV) model, which is a general form of MNL model or NL model. The GEV model provides a powerful tool for modeling correlation between alternatives and keeps the probability expression in a closed-form. The basic specification of the GEV model is a multivariate function $G = G(Y_1, Y_2, \dots, Y_c)$, where $Y_i = e^{V_i}$.

Let G_i be the partial derivative of the function G with respect to Y_i , or $G_i = \partial G / \partial Y_i$. If function G satisfies four conditions: 1) $G(Y_1, Y_2, \dots, Y_c)$ is nonnegative for any Y_j ; 2) $G(Y_1, Y_2, \dots, Y_c)$ is homogeneous of degree one, which means that $G(\rho Y_1, \rho Y_2, \dots, \rho Y_c) = \rho G(Y_1, Y_2, \dots, Y_c)$; 3) $\lim_{Y_i \rightarrow +\infty} G(Y_1, Y_2, \dots, Y_c) = +\infty$ for any i ; and 4) the partial

derivative function $\partial^k G / \partial Y_{i1}, \dots, \partial Y_{ik}$ is non-negative if k is odd and non-positive if k is even, then the probability of choosing alternative i can be written in the following form (subscript n for decision maker is suppressed):

$$P_i = \frac{Y_i G_i}{G}, \quad (4.24)$$

where $Y_i = e^{V_i}$.

The GEV model is so called in that it is the generalized form of the extreme value distribution model. Under certain conditions, the GEV model can evolve into a MNL or a NL model. When $G = \sum_{j=1}^C Y_j$, the GEV model becomes MNL model and if $G = \sum_{k=1}^C (\sum_{j \in N_k} Y_j^{1/\lambda_k})^{\lambda_k}$, the GEV model becomes NL model. The illustrations are shown in section 4.4.1.

The GEV model can accommodate a variety of substitution patterns by assuming different joint distributions for the random error. For repeated choices over time, if the random errors are correlated over time but are not correlated between different alternatives, the GEV model can be used in such a situation as well. I will explain why GEV can handle substitution patterns and repeated choices in detail in section 4.4. However, since the GEV model requires that the error term must have the extreme value distribution, it cannot handle random taste variance.

4.3.4 Probit Model

As mentioned in section 4.3.1, the MNL model has limitations in accounting for random taste variance, restrictive substitution patterns, and correlation over time. The GEV model can handle restrictive substitution patterns and correlations over time, given that error terms are extreme value distributed. However, it cannot account for random taste variance. The Probit model can deal with these three problems well.

The basic assumption of the Probit model is that the random errors $\boldsymbol{\varepsilon}_n = (\varepsilon_{n,1}, \dots, \varepsilon_{n,c})$ are jointly and normally distributed with a zero mean and covariance matrix Ω . We can assume that random errors take zero mean because the absolute value of utility has no effect on the decision makers' choices, and only the relative difference matters. In other words, if a constant is added to

the utilities of all alternatives, the probability of each alternative in the choice set will not change.

Given the normal distribution with zero mean and covariance matrix Ω , the density function of $\mathbf{\varepsilon}_n$ is:

$$\phi(\mathbf{\varepsilon}_n) = \frac{\exp(-\frac{1}{2} \mathbf{\varepsilon}_n' \Omega^{-1} \mathbf{\varepsilon}_n)}{(2\pi)^{C/2} |\Omega|^{1/2}}. \quad (4.25)$$

Thus, the probability of choosing alternative i is:

$$\begin{aligned} P_{n,i} &= \Pr ob(\varepsilon_{n,j} - \varepsilon_{n,i} < V_{n,i} - V_{n,j}, \forall j \neq i) \\ &= \int_{\mathbf{\varepsilon}_n} I(\varepsilon_{n,j} - \varepsilon_{n,i} < V_{n,i} - V_{n,j}, \forall j \neq i) \phi(\mathbf{\varepsilon}_n) d\mathbf{\varepsilon}_n. \end{aligned} \quad (4.26)$$

Define a set of new random errors that are differences between pairs of original random errors $\tilde{\varepsilon}_{n,ji} = \varepsilon_{n,j} - \varepsilon_{n,i}$, and also define $\tilde{V}_{n,ji} = V_{n,j} - V_{n,i}$. The probability of choosing alternative i can be rewritten as:

$$\begin{aligned} P_{n,i} &= \Pr ob(\varepsilon_{n,j} - \varepsilon_{n,i} < V_{n,i} - V_{n,j}, \forall j \neq i) \\ &= \int_{\mathbf{\varepsilon}_n} I(\tilde{V}_{n,ji} + \tilde{\varepsilon}_{n,ji} < 0, \forall j \neq i) \phi(\tilde{\mathbf{\varepsilon}}_n) d\tilde{\mathbf{\varepsilon}}_n, \end{aligned} \quad (4.27)$$

where $\phi(\tilde{\mathbf{\varepsilon}}_n) = \frac{\exp(-\frac{1}{2} \tilde{\mathbf{\varepsilon}}_n' \tilde{\Omega}^{-1} \tilde{\mathbf{\varepsilon}}_n)}{(2\pi)^{(C-1)/2} |\tilde{\Omega}|^{1/2}}$, and $\tilde{\Omega}_i$ is covariance matrix of $(n-1)$ dimensions vector

$$\tilde{\varepsilon}_{n,i} = (\tilde{\varepsilon}_{n,1i}, \tilde{\varepsilon}_{n,2i}, \dots, \tilde{\varepsilon}_{n,i-1i}, \tilde{\varepsilon}_{n,i+1i}, \dots, \tilde{\varepsilon}_{n,Ci}).$$

As I have mentioned in section 4.2, the differences in error terms are normalized. This method is used in the Probit model. In general, the original covariance matrix Ω has $(C+1)C/2-1$ unknown parameters after normalization and $\tilde{\Omega}_i$ has $(C-1)C/2-1$ unknown parameters after

normalization. In principle, the $(C-1)C/2-1$ elements in $\tilde{\Omega}_i$ can be estimated through the maximum likelihood estimation. Though $\tilde{\Omega}_i$ is derived from Ω , the unknown elements in Ω cannot be identified since we have $(C-1)C/2-1$ equations and $(C+1)C/2-1$ unknown parameters.

The major drawback in the Probit model is the excessive computation time required for the estimation. In most cases, there are no closed-form solutions for the integration, and the integration often has to be numerically approximated by the Monte Carlo Simulation. Many simulators have been suggested for the Probit model, such as the Crude Frequency Simulator (CFS), Normal Importance Sampling (NIS), Kernel-Smoothed Frequency Simulator (KFS), Stern's Decomposition Simulator (SDS), and Geweke-Hajivassiliou-Keane Simulator (GHK). Hajivassiliou et al. (1996) wrote a comprehensive review of these simulators and concluded that the GHK simulator appears to be the most reliable method among these simulators. Even given these simulators, when estimation involves a large set of alternatives, the choice probability would be a high-dimension integral and the simulation of this integral can be time consuming and sometimes infeasible.

4.3.5 Mixed Logit Model

McFadden and Train (2000) concluded that the Mixed Multinomial Logit (MMNL) model is a flexible model that can approximate any discrete choice model derived from Random Utility Maximization, and it provides a flexible and computationally practical approach to discrete response analysis.

A MMNL model is defined as a MNL model with random coefficients β drawn from a density distribution function $f(\beta; \theta)$:

$$P_{n,i} = \int L_{n,i}(\beta) f(\beta; \theta) d\beta, \quad (4.28)$$

where

$$L_{n,i}(\beta) = \frac{e^{V_{n,i}(\beta)}}{\sum_{j=1}^C e^{V_{n,j}(\beta)}}. \quad (4.29)$$

The above choice probability expression can be derived in two ways: random coefficients and error components.

Random Coefficients

Assume that decision maker n needs to make a choice among a set of alternatives. It is assumed that the utility for alternative j can be expressed as:

$$U_{n,j} = \beta_n' \mathbf{x}_{n,j} + \varepsilon_{n,j}. \quad (4.30)$$

As with any other discrete choice models, $\mathbf{x}_{n,j}$ is a matrix of observed attributes, β_n is a vector of parameters associated with $\mathbf{x}_{n,j}$, and $\varepsilon_{n,j}$ is a random error with an identical and independent extreme value distribution. The difference between MMNL and other discrete choice models is that β is not a set of constants but a set of random variables whose values may vary with decision makers. It is assumed that β follow some type of distribution with density function $f(\beta | \theta)$.

Decision maker n will choose among alternatives according to $\mathbf{x}_{n,j}$, β_n and $\varepsilon_{n,j}$, or in other words according to $U_{n,j}$. β_n and $\varepsilon_{n,j}$ are not observable. Given that $\varepsilon_{n,j}$ is extreme value distributed, the conditional probability of alternative i 's being chosen given β_n is:

$$L_{n,i} = \frac{\exp(\beta_n' \mathbf{x}_{n,i})}{\sum_j \exp(\beta_n' \mathbf{x}_{n,j})}. \quad (4.31)$$

Although the actual values of β_n are unknown, we can assume its density function. Therefore, to calculate the unconditional probability of alternative i 's being chosen is simply to integrate $L_{n,i}$ over all possible values of β_n :

$$\begin{aligned} P_{n,i} &= \int L_{n,i} f(\beta; \theta) d\beta \\ &= \int \frac{\exp(\beta'_n \mathbf{x}_{n,i})}{\sum_j \exp(\beta'_n \mathbf{x}_{n,j})} f(\beta; \theta) d\beta. \end{aligned} \quad (4.32)$$

Error Components

In the error components setting, random errors are assumed to be correlated with each other. The error term $\varepsilon_{n,j}$ is decomposed into two parts: uncorrelated $\tilde{\varepsilon}_{n,j}$ and correlated $\gamma_n' \mathbf{z}_{n,j}$. The utility function is assumed to be:

$$U_{n,j} = \alpha' \mathbf{x}_{n,j} + \gamma_n' \mathbf{z}_{n,j} + \tilde{\varepsilon}_{n,j}, \quad (4.33)$$

where $\mathbf{x}_{n,j}$, $\mathbf{z}_{n,j}$ are observed attributes, α are vectors of fixed parameters for all decision makers, and γ_n are random parameters with zero mean⁴ and density function $f(\gamma | \theta)$. $\tilde{\varepsilon}_{n,j}$ is identically and independently distributed with an extreme value distribution. The conditional probability of alternative i 's being chosen given β_n is:

$$L_{n,i} = \frac{\exp(\alpha' \mathbf{x}_{n,i} + \gamma_n' \mathbf{z}_{n,i})}{\sum_j \exp(\alpha' \mathbf{x}_{n,j} + \gamma_n' \mathbf{z}_{n,j})}. \quad (4.34)$$

The probability of choosing alternative i can be expressed as:

⁴ This is because if we assume γ_n have a mean of constant, $\gamma_n' \mathbf{z}_{n,j}$ can be decomposed into two part. One part is this constant value times $\mathbf{z}_{n,j}$, which is similar to $\alpha' \mathbf{x}_{n,j}$. The other part is equivalent to random coefficients with zero mean.

$$\begin{aligned}
P_{n,i} &= \int L_{n,i} f(\gamma; \theta) d\gamma \\
&= \int \frac{\exp(\alpha' \mathbf{x}_{n,i} + \gamma_n' \mathbf{z}_{n,i})}{\sum_j \exp(\alpha' \mathbf{x}_{n,j} + \gamma_n' \mathbf{z}_{n,j})} f(\gamma; \theta) d\gamma.
\end{aligned} \tag{4.35}$$

If $\mathbf{z}_{n,j} = \mathbf{x}_{n,j}$, the results of random coefficients and error components are equivalent to each other. In this case, α and γ in the method of error components are basically the mean and variance of β in the method of random coefficients.

The Monte Carlo Simulation performs the estimation of the Mixed Logit model. In a Monte Carlo Simulation, we first draw a set of random numbers β from the assumed density function $f(\beta | \theta)$. Then we calculate $L_{n,i}(\beta)$ based on newly drawn values of β . We repeat this step R times, and after R draws, we average the results. The average will be the simulated probability.

$$\hat{P}_{n,i} = \frac{1}{R} \sum_{r=1}^R L_{n,i}(\beta). \tag{4.36}$$

There is no mathematical function to decide the value of R . The decision on the value of R is based on experience. In theory, the bigger the R is, the closer the simulated values are to the true values. However, the value of R cannot be too big, as it will result in infeasible computation time. Therefore, the R value is usually the result of trading off between estimation precision and computation burden. Furthermore, the way to draw random numbers can also make a difference. A good random number generator needs fewer numbers of draws than a poor one.

Depending on the study objective, the density function $f(\beta | \theta)$ can be specified to be any distributions. Uniform distribution, triangle distribution, normal distribution, and lognormal distribution are most commonly used. The values of α and γ that maximize the likelihood will be the estimators.

4.4 Model Structure for Residential Location Study

The GEV model is selected in this residential location study for two reasons. First, typically there are many alternatives in a residential location model. Second, it is likely that the alternatives next to each other are correlated. The GEV model is superior to other discrete choice models (i.e., Nested Logit, Mixed Logit, and Probit models) as it provides a closed-form probability function and provides a flexible correlation structure for error terms. The Nested Logit model is a special case of the GEV model, with more restrictive assumptions. Both Mixed-Logit and Probit models involve integrals whose dimensions depend on the number of alternatives. The computational burden increases exponentially when the choice set expands. In our empirical dataset, the number of alternatives is more than 700, which makes it unrealistic to use the Mixed-Logit or Probit models.

4.4.1. Model Structure for Residential Location Problem

According to equation 4.3, the probability of living in location j is equal to: $\Pr(V_{n,j} + \varepsilon_{n,j} > V_{n,i} + \varepsilon_{n,i}, \text{ for any } j \neq i)$. Assume that the error terms ε are identically but not-independently⁵ distributed with multivariate extreme value distribution.

Suppose that error term ε_i (for notational clarity, n for household is suppressed in this section) represents the unobserved utility from neighborhood i . It is likely that neighborhoods follow the First Law of Geography postulated by Tobler (1970), which stated that near things are more related than distant things. Therefore, the unobserved utilities among different neighborhoods are correlated with respect to proximity in space. The closer the neighborhoods are with each other, the higher the correlation. Define ρ_{ij} as the correlation factor, describing the correlation between ε_i and ε_j ; ρ_{ij} is a function of the distance between neighborhood centers, $Dist_{ij}$, and an unknown positive parameter λ :

$$\rho_{ij} = \exp(-Dist_{ij} \cdot \lambda). \quad (4.37)$$

⁵ The non-independence is to handle correlation between alternatives.

This correlation factor ρ_{ij} has several properties. First, when ρ_{ii} is equal to 1, it indicates that the correlation for ε_i itself is 1. Second, when ρ_{ij} tends to be zero, it suggests that the distance between neighborhood j and i is very large. Third, ρ_{ij} always falls between 0 and 1. Last, ρ_{ij} decreases as distance between the two neighborhood centers increases.

Consider GEV model with G function to be:

$$G(y_1, y_2, \dots, y_c) = \sum_{i=1}^{c-1} \sum_{j=i+1}^c [(k_{i,ij} y_i)^{1/(1-\rho_{ij})} + (k_{j,ij} y_j)^{1/(1-\rho_{ij})}]^{1-\rho_{ij}}, \quad (4.38)$$

where,

$y_i \geq 0$, for all i ,

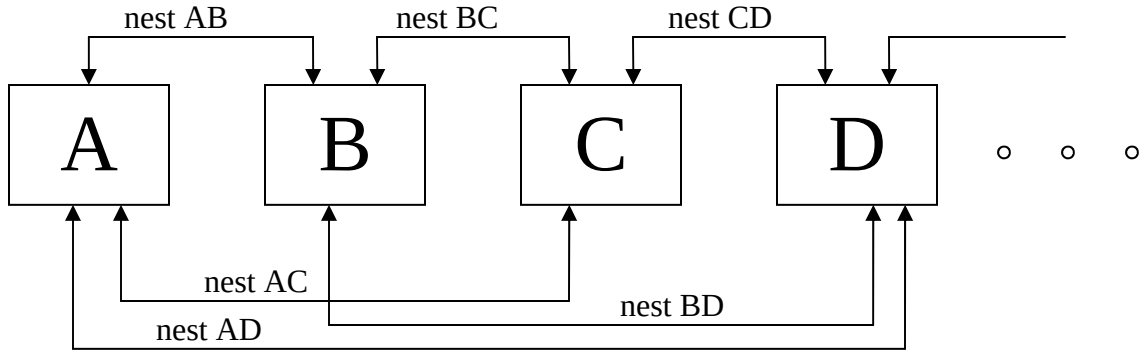
ρ_{ij} is the similarity index between zone i and zone j , and

$k_{i,ij}$ is the weighting parameter for alternative i in nest ij , $k_{i,ij} = \frac{I(\rho_{ij})}{\sum_{m=1, m \neq i}^c I(\rho_{im})}$, $0 \leq k_{i,ij} \leq 1$, and

by definition $\sum_{j=1, j \neq i}^c k_{i,ij} = 1$.

The similarity index ρ_{ij} describes the level of spatial correlation between zone i and zone j in nest ij . Shown in Figure 4.1, an alternative can appear in more than one nest if it is correlated with alternatives in other nests. Because the same alternative can appear in more than one nest, the weighting parameter $k_{i,ij}$ must be used to reflect to what extent alternative i belongs to nest ij .

Figure 4. 1: Assumed Correlation Structure



$G(y_1, y_2, y_3 \dots y_c)$ is a nonnegative, homogeneous-of-degree-one function of $(y_1, y_2, y_3 \dots y_c) \geq 0$. Since all y_i and $k_{i,j}$ are greater than zero, $G(y_1, y_2, y_3 \dots y_c)$ is nonnegative. To see that $G(y_1, y_2, y_3 \dots y_c)$ is homogeneous of degree one, we substitute ky_i for y_i into equation 4.38 (where ρ is a constant), we have

$$\begin{aligned}
 G(my_1, my_2, \dots, my_c) &= \sum_{i=1}^{c-1} \sum_{j=i+1}^c [(k_{i,j} my_i)^{1/(1-\rho_{ij})} + (k_{j,i} my_j)^{1/(1-\rho_{ij})}]^{1-\rho_{ij}} \\
 &= \sum_{i=1}^{c-1} \sum_{j=i+1}^c [m^{1/(1-\rho_{ij})} (k_{i,j} y_i)^{1/(1-\rho_{ij})} + m^{1/(1-\rho_{ij})} (k_{j,i} y_j)^{1/(1-\rho_{ij})}]^{1-\rho_{ij}} \\
 &= \sum_{i=1}^{c-1} \sum_{j=i+1}^c (m^{1/(1-\rho_{ij})})^{1-\rho_{ij}} [(k_{i,j} y_i)^{1/(1-\rho_{ij})} + (k_{j,i} y_j)^{1/(1-\rho_{ij})}]^{1-\rho_{ij}} \\
 &= \sum_{i=1}^{c-1} \sum_{j=i+1}^c m [(k_{i,j} y_i)^{1/(1-\rho_{ij})} + (k_{j,i} y_j)^{1/(1-\rho_{ij})}]^{1-\rho_{ij}} \\
 &= mG(y_1, y_2, \dots, y_c).
 \end{aligned}
 \tag{4.39}$$

Equation 4.39 shows that $G(y_1, y_2, y_3 \dots y_c)$ is a homogeneous-of-degree-one function. For the GEV model, the n th order cross partial derivatives of G must be non-negative for odd n and non-positive for even n . I will demonstrate this property in the following paragraph. To make statement simplified, let $1 - \rho_{ij} = D_{ij}$, then the first order partial derivative with respect to y_m is:

$$\begin{aligned}
\frac{\partial G}{\partial y_m} &= \frac{\partial \sum_{i=1, i \neq m}^c [(k_{i,im} y_i)^{1/D_m} + (k_{m,im} y_m)^{1/D_m}]^{D_m}}{\partial y_m} \\
&= \sum_{i=1, i \neq m}^c D_m [(k_{i,im} y_i)^{1/D_m} + (k_{m,im} y_m)^{1/D_m}]^{D_m-1} \frac{k_{m,im}}{D_m} (k_{m,im} y_m)^{(1/D_m-1)} \\
&= \sum_{i=1, i \neq m}^c k_{m,im} [(k_{i,im} y_i)^{1/D_m} + (k_{m,im} y_m)^{1/D_m}]^{D_m-1} (k_{m,im} y_m)^{(1/D_m-1)}.
\end{aligned} \tag{4.40}$$

Since $0 \leq \rho_{ij} \leq 1$, we have $0 \leq D_{ij} \leq 1$. Therefore, $\frac{\partial G}{\partial y_m} \geq 0$.

The second cross partial derivative is:

$$\begin{aligned}
\frac{\partial G}{\partial y_m \partial y_n} &= \partial \left(\frac{\partial G}{\partial y_m} \right) / \partial y_n \\
&= \partial \left(\sum_{i=1, i \neq m}^c k_{m,im} [(k_{i,im} y_i)^{1/D_m} + (k_{m,im} y_m)^{1/D_m}]^{D_m-1} (k_{m,im} y_m)^{(1/D_m-1)} \right) / \partial y_n \\
&= \partial (k_{m,nm} [(k_{n,mn} y_n)^{1/D_m} + (k_{m,mn} y_m)^{1/D_m}]^{D_m-1} (k_{m,mn} y_m)^{(1/D_m-1)}) / \partial y_n \\
&= k_{m,mn} (k_{m,mn} y_m)^{(1/D_m-1)} (D_m - 1) [(k_{n,mn} y_n)^{1/D_m} + (k_{m,mn} y_m)^{1/D_m}]^{D_m-2} \frac{k_{n,mn}}{D_m} (k_{n,mn} y_n)^{(1/D_m-1)} \\
&= k_{m,mn} k_{n,mn} \frac{(D_m - 1)}{D_m} (k_{m,mn} y_m)^{(1/D_m-1)} (k_{n,mn} y_n)^{(1/D_m-1)} [(k_{n,mn} y_n)^{1/D_m} + (k_{m,mn} y_m)^{1/D_m}]^{D_m-2}.
\end{aligned} \tag{4.41}$$

Since $0 \leq D_{ij} \leq 1$, $\frac{\partial G}{\partial y_m \partial y_n} \leq 0$.

Since the second order partial derivatives of G , $\frac{\partial G}{\partial y_m \partial y_n}$, is only a function of variables y_m and y_n , the higher order partial derivatives of G with respect to any variable other than y_m and y_n are all equal to zero. Then we have demonstrated that the n th order cross partial derivatives of G is non-negative for odd n and non-positive for even n . Since G function is nonnegative, homogeneous-of-degree-one, and its n th order cross partial derivatives is

non-negative for odd n and non-positive for even n , G represent a class of GEV model. According to McFadden (1978), the following function is a cumulative multivariate extreme value distribution:

$$F(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_c) = \exp(-G(e^{-\varepsilon_1}, e^{-\varepsilon_2}, \dots, e^{-\varepsilon_c})) = \exp(-\sum_{i=1}^{c-1} \sum_{j=i+1}^c [(k_{i,j} e^{-\varepsilon_i})^{1/(1-\rho_{ij})} + (k_{j,i} e^{-\varepsilon_j})^{1/(1-\rho_{ij})}]^{1-\rho_{ij}}). \quad (4.42)$$

This cumulative multivariate extreme value distribution $F(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_c)$ can evolve into MNL when there is no correlation between neighborhoods ($\rho_{ij} = 0$).

When $\rho_{ij} = 0$, and by definition $\sum_{j=1, j \neq i}^c k_{i,j} = 1$, the cumulative extreme distribution is:

$$\begin{aligned} F(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_c) &= \exp(-\sum_{i=1}^{c-1} \sum_{j=i+1}^c [(k_{i,j} e^{-\varepsilon_i}) + (k_{j,i} e^{-\varepsilon_j})]) \\ &= \exp(-\sum_{i=1}^c \sum_{j=1, j \neq i}^c k_{i,j} e^{-\varepsilon_i}) \\ &= \exp(-\sum_{i=1}^c e^{-\varepsilon_i}). \end{aligned} \quad (4.43)$$

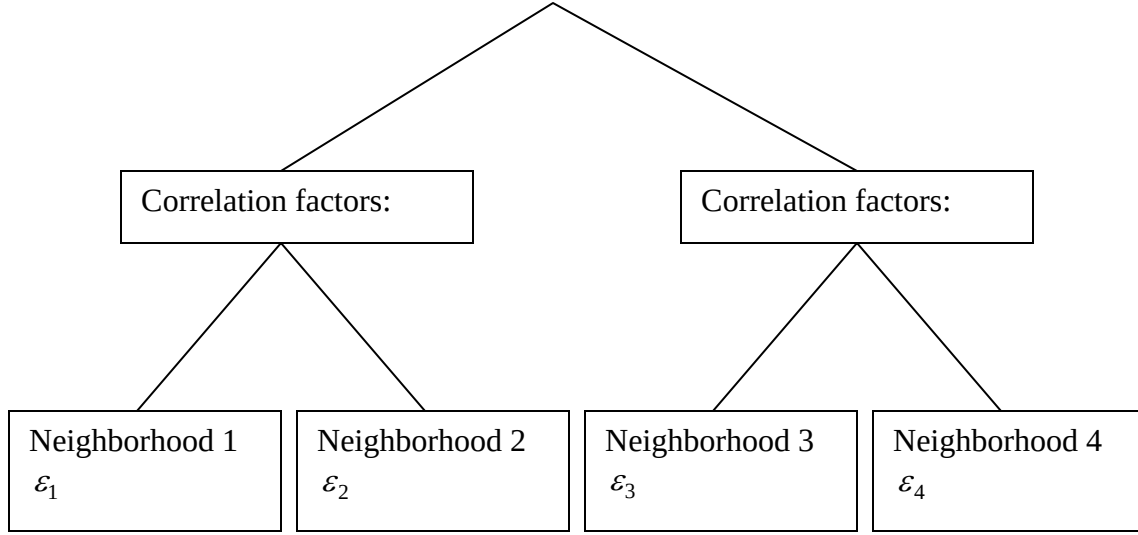
For the marginal probability of $x < \varepsilon_i$,

$$\begin{aligned} \text{Prob}(x < \varepsilon_i) &= F(\infty, \infty, \infty, \dots, \varepsilon_i, \infty, \infty, \dots) - F(-\infty, -\infty, -\infty, \dots, \varepsilon_i, -\infty, -\infty, \dots) \\ &= \exp(-e^{-\varepsilon_i}) - \exp(-\infty) \\ &= \exp(-e^{-\varepsilon_i}). \end{aligned} \quad (4.44)$$

This is the familiar setting of the MNL.

$F(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_c)$ can also evolve into Nested Logit model. Let us use an example of four alternatives in a Nested Logit model setting. The structure of the four alternatives is shown in Figure 4.2.

Figure 4. 2: An Example of Four Alternatives in a Nested Logit Model Setting



It is assumed that in this Nested Logit model, $\rho_{12} \neq 0, \rho_{34} \neq 0, \rho_{13} = \rho_{14} = \rho_{24} = \rho_{23} = 0$.

Therefore,

$$k_{1,12} = \frac{\rho_{12}}{\rho_{12} + \rho_{13} + \rho_{14}} = 1. \quad (4.45)$$

Similarly, we get

$$k_{2,12} = k_{3,34} = k_{4,34} = 1 \text{ and } k_{1,13} = k_{3,13} = k_{1,14} = k_{4,14} = k_{2,23} = k_{3,23} = k_{2,24} = k_{4,24} = 0.$$

The cumulative distribution is:

$$\begin{aligned}
 F(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_c) &= \exp\left(-\sum_{i=1}^{c-1} \sum_{j=i+1}^c [(k_{i,j} e^{-\varepsilon_i})^{1/(1-\rho_{ij})} + (k_{j,i} e^{-\varepsilon_j})^{1/(1-\rho_{ij})}]^{1-\rho_{ij}}\right) \\
 &= \{[k_{1,12} e^{-\varepsilon_1}]^{1/1-\rho_{12}} + [k_{2,12} e^{-\varepsilon_2}]^{1/1-\rho_{12}}\}^{1-\rho_{12}} + \{[k_{1,13} e^{-\varepsilon_1}]^{1/1-\rho_{13}} + [k_{3,13} e^{-\varepsilon_3}]^{1/1-\rho_{13}}\}^{1-\rho_{13}} \\
 &\quad + \{[k_{1,14} e^{-\varepsilon_1}]^{1/1-\rho_{14}} + [k_{4,14} e^{-\varepsilon_4}]^{1/1-\rho_{14}}\}^{1-\rho_{14}} + \{[k_{2,23} e^{-\varepsilon_2}]^{1/1-\rho_{23}} + [k_{3,23} e^{-\varepsilon_3}]^{1/1-\rho_{23}}\}^{1-\rho_{23}} \\
 &\quad + \{[k_{2,24} e^{-\varepsilon_2}]^{1/1-\rho_{24}} + [k_{4,24} e^{-\varepsilon_4}]^{1/1-\rho_{24}}\}^{1-\rho_{24}} + \{[k_{3,24} e^{-\varepsilon_3}]^{1/1-\rho_{34}} + [k_{4,34} e^{-\varepsilon_4}]^{1/1-\rho_{34}}\}^{1-\rho_{34}} \\
 &= \{[e^{-\varepsilon_1}]^{1/1-\rho_{12}} + [e^{-\varepsilon_2}]^{1/1-\rho_{12}}\}^{1-\rho_{12}} + \{[e^{-\varepsilon_3}]^{1/1-\rho_{34}} + [e^{-\varepsilon_4}]^{1/1-\rho_{34}}\}^{1-\rho_{34}}.
 \end{aligned} \quad (4.46)$$

This is a cumulative distribution for the NL model.

As a summary, equation 4.42 is a cumulative multivariate extreme value distribution function for neighborhoods choice with ρ_{ij} accounting for the correlation between the random error terms ε_i and ε_j .

4.4.2 Probability Expression using the $F(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_c)$

With the configuration of similarity index and weighting parameter in section 4.4.1, according to the theorem of McFadden (1978), the probability of choosing alternative location j for household n is:

$$P(j | n) = \frac{\sum_{i=1, i \neq a}^c (k_{j,ij} e^{V_{n,j}})^{1/1-\rho_{ij}} [(k_{j,ij} e^{V_{n,j}})^{1/1-\rho_{ij}} + (k_{i,ij} e^{V_{n,i}})^{1/1-\rho_{ij}}]^{\rho_{ij}-1}}{\sum_{i=1}^{c-1} \sum_{j=i+1}^c [(k_{j,ij} e^{V_{n,j}})^{1/1-\rho_{ij}} + (k_{i,ij} e^{V_{n,i}})^{1/1-\rho_{ij}}]^{1-\rho_{ij}}} \quad (4.47)$$

where, $V_{n,j}$ and $V_{n,i}$ are systematic utilities of alternatives j and i for household n ; ρ_{ij} accounts for the similarity between alternative i and j ; $k_{i,ij}$ is weighting parameter.

4.4.3 Similarity Index ρ_{ij} and Weighting Parameter $k_{i,ij}$

The similarity index ρ_{ij} is defined as the function of a constant λ (to be estimated) and distance between centers of alternative j and i :

$$\begin{cases} \rho_{ij} = \exp(-\lambda \times d_{ij}^\omega), & \text{if } d_{ij} \leq \theta \text{ miles} \\ \rho_{ij} = 0, & \text{if } d_{ij} > \theta \text{ miles} \end{cases} \quad (4.48)$$

The weighting parameter is defined as the function of similarity index:

$$k_{i,j} = \frac{I(\rho_{ij})}{\sum_{m=1, m \neq i}^c I(\rho_{im})} \quad (4.49)$$

where $I(\rho_{ij})$ is the dummy variable, which is equal to 1 if $\rho_{ij} \neq 0$ and 0 for otherwise. $k_{i,j}$ will always fall between 0 and 1. Furthermore, it should satisfy: $\sum_{j=1, j \neq i}^c k_{i,j} = 1$.

The use of the exponential form in equation (4.48) regulates that the values of ρ_{ij} will always fall between 0 and 1, where 0 indicates no correlation and 1 corresponds to perfect correlation. λ , which is inversely related to ρ_{ij} , will be estimated from the empirical data. The value of θ must be set prior to the estimation. The value of ω can be estimated from the model in theory, but we choose to set its value in advance to reduce the computational burden. We tested a wide range of possible values for the two parameters and set the value of 3.5 for θ and the value of 0.2 for ω . This combination provides the model with the best performance in terms of log-likelihood. It is also observed that the choice of these two values has a negligible effect on the coefficients of independent variables, suggesting that autocorrelation is highly independent from the other variables used in the model.

The similarity index ρ_{ij} describes the level of spatial correlation between zone i and zone j in nest ij . Shown in Figure 4.1, there are nests in a GEV model. Alternatives in a nest are correlated with each other. An alternative can appear in more than one nest, if this alternative is correlated with other alternatives in another nest. Because the same alternative can appear in more than one nest, the weighting parameter $k_{i,j}$ must be used to reflect to what extent alternative i is a member of nest ij .

In a residential choice model, the assumption regarding ρ_{ij} could be that: 1) any two alternatives are correlated; 2) any two alternatives that are within a certain distance are correlated 3) only alternatives that are adjacent to each other are correlated. The first assumption is more generous than the other two; however, the computation burden is also heavier than other two.

When the number of alternatives is large enough, it is even not feasible to estimate a base on the first assumption. The third assumption excludes some alternatives that might be correlated with each other. The second assumption relaxes the assumption that only alternatives adjacent to each other are correlated. Moreover, it reduces the computational burden by adding a limit: only alternatives that are within a certain distance are correlated. For this reason, I adopt a second assumption (see equation 4.48).

For weighting parameter $k_{i,j}$, it could be assumed to be:

$$1) \quad k_{i,j} = \frac{\rho_{ij}}{\sum_j \rho_{ij}} \quad (4.50)$$

$$2) \quad k_{i,j} = \frac{I(\rho_{ij})}{\sum_{m=1, m \neq i}^c I(\rho_{im})} \quad (4.51)$$

$$3) \quad k_{i,j} = \frac{L_{ij}}{\sum_j L_{ij}}, \quad (4.52)$$

where, L_{ij} is the shared boundary length between alternative j and alternative i .

I mentioned previously that the weighting parameter $k_{i,j}$ measures the extent alternative i belongs to nest ij . One assumption is that alternative i equally belongs to each nest containing it (e.g., Bhat and Guo, 2004). Bekhor and Parshker (2008) hypothesized that $k_{i,j}$ is a function of a shared boundary between alternative i and other alternatives (Equation 4.52). There is no evidence showing that any equation is more superior to the other two. I adopted equation 4.51 in the model. Note that equation 4.52 is actually only suitable when ρ_{ij} is assumed to be correlated for alternatives that are adjacent to each other. Equation (4.51) states that the extent that alternative i belongs to nest ij depends on the number of alternatives with which alternative i correlates. The more locations alternative i is correlated with, the smaller $k_{i,j}$ is.

4.4.4 Utility Function V_j and Parameter β_l

The total utility that location j derives can be decomposed into two components: systematic utility $V_{n,j}$ and random error $\varepsilon_{n,j}$. Or, $U_{n,j} = V_{n,j} + \varepsilon_{n,j}$. The systematic utility, $V_{n,j}$, can be specified as a linear function of a set of observed attributes x_l , weighted by corresponding β_l :

$$V_{n,j} = \sum_l \beta_l \times f(x_{n,j,l}).$$

To account for the historical deposition influence, I propose that the value of β_l changes as a household's prior location experience changes:

$$\beta_l = \alpha_{l1} + \alpha_{l2} \bullet \frac{\partial f(x_{n,a,l})}{\partial x_{n,a,l}}, \quad (4.53)$$

where,

β_l is the parameter of the l th accessibility attribute,

α_{l1} is the base parameter for β_l ,

α_{l2} is the adjustment parameter for β_l , and

$x_{n,a,l}$ is the l th accessibility attribute for household n with prior location a .

Function $f(x_{n,a,l})$ can be considered as a utility function. According to the Law of Diminishing Marginal Utility, $\partial f(x_{n,a,l}) / \partial x_{n,a,l}$ decreases as $x_{n,a,l}$ increases. Thus, $f(x_{n,a,l})$ should be concave down. I used the logarithm function to capture the concave down property of $f(x_{n,a,l})$. The explanatory variables are adjusted by adding one to the l th attribute: $f(x_{n,a,l}) = \log(x_{n,a,l} + 1)$, to avoid taking the logarithm of zero. The logarithm function has three appealing features. First, when the value of an attribute is zero, the corresponding $f(x_{n,a,l})$ is zero as well. Second, $f(x_{n,a,l})$ is always non-negative, which is consistent with the fact that $x_{n,a,l}$ is always non-negative. Third, when $x_{n,a,l}$ increases, $f(x_{n,a,l})$ increases as well.

4.4.5 Monte Carlo Simulation and Halton Sequence

Monte Carlo Simulation of an integral $\int g(\boldsymbol{\varepsilon}) f(\boldsymbol{\varepsilon} | \boldsymbol{\theta}) d\boldsymbol{\varepsilon}$ involves drawing a vector of random numbers $\boldsymbol{\varepsilon}$ from a density function $f(\boldsymbol{\varepsilon} | \boldsymbol{\theta})$, calculating corresponding values $g(\boldsymbol{\varepsilon})$ according to the drawn random numbers and averaging the result. Therefore, the first step of a Monte Carlo Simulation is to draw random numbers from a density function.

The procedures for drawing random numbers from a standard normal density and a standard uniform density (uniform random number between 0 and 1) are available in most of statistical packages, such as GAUSS, SAS, and SPSS. These random numbers are called pseudo-random numbers because they are generated with a seed number, which is randomly generated by statistical package.

Given that random numbers follow a standard uniform distribution, I can simulate other distributions including normal distribution. Available methods include linear transformation, inverse cumulative density function, Choleski transformation, accept-reject method, importance sampling, and Gibbs sampling. The right method for drawing random numbers depends on the density function we are going to draw.

Whatever statistical package is used, the start point of drawing random numbers is to draw a standard uniform random number. Among different standard uniform pseudo-random number generators, the Halton Sequence is by far one of the best, if not the best. Many studies (Bhat, 2001; Train, 2000; Hensher, 2001) have demonstrated by empirical research that the Halton Sequence provided an efficient means of estimation. Bhat (2001) found that 100 Halton Sequences provided more precise results for the Mixed Logit model than 1000 random number draws. With an additional 25 draws, the Halton Sequence became even better. The simulation error with 125 Halton Sequences was half as large as with 1000 random draws, and even smaller than 2000 random draws. If the computation time is linearly related to the number of draws, the Halton Sequence could save 90% of computation time compared to other random number generators.

4.4.6 Estimation of Parameters

After we get the probability expression of every alternative, we can now form the likelihood expression for individual n :

$$Li = \prod_{j=1}^C (P_{n,j})^{\delta_{n,j}}, \quad (4.55)$$

where

$$\delta_{n,j} = \begin{cases} 1 & \text{if decision maker } n \text{ choose alternative } j \\ 0 & \text{otherwise} \end{cases}.$$

Assume that each individual makes the decision independently, the joint probability of all N individuals in the sample choosing the observed choices:

$$L(\theta) = \prod_{n=1}^N \prod_{j=1}^C (P_{n,j})^{\delta_{n,j}}. \quad (4.56)$$

The log-likelihood function is:

$$LL(\theta) = \prod_{n=1}^N \prod_{j=1}^C \delta_{n,j} \ln(P_{n,j}), \quad (4.57)$$

where θ denotes the vector containing all parameters to be estimated. The estimation of θ is performed by the maximum likelihood (ML) estimation. The value of θ that maximizes the log-likelihood $LL(\theta)$ is the ML estimator.

4.5 Neighborhood and Accessibility Attributes

4.5.1. Neighborhood Measurement

One major issue in measuring neighborhood attributes is the definition of a neighborhood. Neighborhood is a vague concept. Many studies have attempted to identify a clear concept of neighborhood. Keller (1968) defined neighborhood as a place with physical and symbolic boundaries. In Morris and Hess (1975), these physical and symbolic boundaries are the perception of people who live in that place. Rather than defining neighborhood as a place, Warren (1981) defined neighborhood as a social organization of population. Downs (1981), Hallman (1984), and Galster (2001) argued that neighborhood is a combination of geographic units and certain social networks.

Galster (2001) pointed out that neighborhood is “a term that is hard to define precisely, but everyone knows it when they see it”. In sociological studies, researchers define a neighborhood as a natural boundary that contains a relative homogeneous population and/or land use pattern. When implementing this concept, neighborhood is usually set as a particular Census tract or a group of Census tracts. The similarity of the residents is a key in defining a neighborhood. From the geographic perspective, Lee (1976) investigated how households conceive their neighborhood boundaries by asking them to “draw a line around the part” they considered the neighborhood or district. Lee found that the average size of a neighborhood was about one hundred acres and that population density had little effect on neighborhood size. Natives and the working class tended to create a relatively small neighborhood while persons who lived in a neighborhood with mixed social classes tended to generate large neighborhoods.

In many residential location choice studies, a neighborhood is simply defined as a Census tract, a TAZ, or a zip code without adequate justification. Galster (2001) defined neighborhood as a complex commodity produced by households, businesses, property owners, and local governments with a bundle of spatially based attributes such as infrastructure, demographics, economics, public services, environmental, and political characteristics. Coulton et al. (2001) examined residents’ perception of a neighborhood. They asked household members to draw a map of what they considered as the boundaries of their neighborhoods. It was found that there were significant differences between perceived neighborhood boundaries and Census tract boundaries; households with similar demographic attributes often hold different ideas about their residential neighborhood boundaries.

As seen from the above discussion, the definition of neighborhood is conceptually and operationally difficult. In residential location studies, researchers have proposed using different types of methods to set the boundaries for neighborhoods. Kim et al. (2005) used a radius of $\frac{1}{4}$, $\frac{1}{2}$, and 1 mile to consider the environment surrounding the residential location. The variables calculated are the numbers of parks, retail stores, health care facilities, bus stops, and the size of nature and open areas, within a buffer zone of each respondent's home location. The space proximity is measured as the Euclidean shortest distance.

Guo and Bhat (2006) investigated three ways of defining a neighborhood: Census-unit representation, circular-unit representation, and network-band representation. The Census-unit representation of neighborhood used Census geography as the neighborhood boundary. Circular-unit representation uses the circle centered at the residential location. The network-band representation is to use the street network in calculating the "circle" centered at a residential location. In other words, the circular-unit neighborhood used straight distance as the radius, while the network-band neighborhood used street network distance as the radius. According to measures on the goodness of fit, Guo and Bhat (2006) concluded that the network-band representation of neighborhood has relatively superior estimation results than the other two types of definition.

4.5.2. Accessibility Measurement

4.5.2.1. The Definition of Accessibility

Many studies have suggested the definition for accessibility from different perspectives. Hansen (1959) defined accessibility as the potential opportunities for interaction. Dalvi and Martin (1976) suggested that accessibility represents the way with which people could reach any land-use activity from a given location. Ben-Akiva and Lerman (1979) connected accessibility with the benefits provided by a land-use pattern and associated transportation system. Genus (2006) summarized that accessibility is the extent to which land-use patterns and transportation systems

enable individuals to participate in different types of activities in different locations by means of a travel mode.

Genus (2006) concluded that four components are related to the accessibility concept: land-use, transportation, temporal constraints, and the individual. The land-use component reflects (a) the amount, the quality, and the spatial distribution of activity opportunities available at each destination (jobs, shops, health, social and recreational facilities, etc.), (b) the demand for these opportunities at origin locations, and (c) confrontation of supply and demand for opportunities. The confrontation of supply is also known as “competition effects”, which indicates a restricted capacity for opportunities due to the limitation of land. The transportation component reflects the cost for an individual to travel between origin and destination using a particular travel mode, including travel time, and travel costs. The reliability, level of comfort, and accident risk are also components of a transportation system. The temporal component reflects the temporal constraints associated with the individual and the facility. Temporal constraints associated with a facility represent the availability of opportunities at different times of day. Temporal constraints associated with the individual are the time available for the individual to participate in certain activities. The individual component reflects the needs, abilities, and opportunities of individuals in pursuing different activities.

4.5.2.2. The Measurement of Accessibility

A crucial issue is the how to measure accessibility attributes. Ingram (1971) categorized two types of measures: relative accessibility and integral accessibility. The relative accessibility measurement is used in a monocentric city form, while the integral accessibility measurement can be used in a polycentric layout. Handy and Niemeier (1997) summarized three types of accessibility measures: cumulative opportunity measures, gravity-based measures, behavioral measures, and space-time measures. As I mention above, accessibility involved four components. An ideal accessibility measurement should consider all four components. I will briefly review these four types of accessibility measurements and discuss how they incorporate the four components.

Cumulative Measures

Cumulative measure of accessibility, also known as contour measure, counts the number of opportunities that could be reached from the origin location within specified travel distances or times. The cumulative measure of accessibility is a special case of the gravity-based measure, when the travel time to opportunities outside the boundary is set to zero.

Gravity-Based Measures

The gravity-based method is the most widely used in empirical studies. Developed in early 1950s (Hansen 1959, Reilly 1953, and Huff 1963), gravity based measures include two components: land use, representing the spatial distribution of opportunities, and transportation, describing the cost of traveling. A typical expression for the gravity-based measurement is:

$$X_{ij} = \sum_{j=1}^n O_j \times f(C_{ij}), \quad (4.58)$$

where,

X_{ij} is accessibility of zone (i) to the considered type of opportunities (j),

O_j refers to the opportunities of the considered type in zone j ,

C_{ij} is generalized time or cost to travel from zone i to zone j , and

$f()$ is the impedance function.

The impedance function could be power, Gaussian, logistic, or the negative exponential function, the last of which is most often used and most closely related to travel behavior theory (Handy and Niemeier, 1997).

The gravity-based measurement has been implemented in many studies (e.g. Hansen, 1959; Vickerman, 1974; Wilson, 1971). However, there are two shortcomings associated with this measure. First, this measurement computes accessibility based on a single fixed zone i , which

implicitly assumes that all trips start from that zone. In reality, a traveler can often make linked trips, which involve several zones. Second, this measure doesn't take into account the temporal constraints and the characteristics of an individual.

Utility-Based Integral Measure

Ben-Akiva and Lerman proposed the utility-based integral accessibility measure in 1978. This approach interprets accessibility as the outcome of a set of travel choices. There are two types of behavioral measures. The first one is the well-known logsum approach, based on the Random Utility Theory. Using a logit model, individuals are assumed to assign a utility to each alternative in a choice set. If they select the alternative that maximizes their utility, accessibility can be defined as the denominator of a multi-nominal logit model. The logsum serves as a summary measure, indicating the desirability of the full choice set. The utility-based integral measure has the form:

$$A_i = \ln\left(\sum_{k=1}^m e^{V_k}\right), \quad (4.59)$$

where,

A_i is the expected utility for person i , and

V_k is the systematic utility of alternative k .

The second approach of behavioral measurement is based on the doubly constrained entropy model. This approach takes the following form (Martínez, 1995):

$$A_i = -\frac{1}{\beta} \ln(a_i), \quad (4.60)$$

$$A_j = -\frac{1}{\beta} \ln(b_j), \quad (4.61)$$

$$A_{ij} = -\frac{1}{\beta} \ln(a_i b_j), \quad (4.62)$$

$$a_i = \sum_{j=1}^n \frac{1}{b_j} O_j(b) f(C_{ij}), \quad (4.63)$$

$$b_i = \sum_{j=1}^m \frac{1}{a_i} O_j(a) f(C_{ij}), \quad (4.64)$$

where,

A_i is the expected utility generated per trip from zone i ,

A_j is the expected utility generated per trip to zone j ,

A_{ij} is the expected utility generated per trip between zone i and j ,

a_i, b_j are balancing factors, related with zone i and j ,

$O_j(a), O_j(b)$ are opportunities available in zone i and j , and

$f(C_{ij})$ is the impedance function between zone i and j .

The second approach has the advantage over the logsum approach in that it accounts for the competition effect by incorporating balancing factors. The balancing factors in the second approach are mutually dependent and should be estimated iteratively.

Utility-based measures have advantages over gravity-based because they can handle competition effects and they are sensitive to changes in the supply capacity (e.g., hospital beds and job vacancies etc.). Furthermore, the measure is based on the expected utility for every individual; therefore, the individual component is taken into account to some extent.

Space-time Measure

The utility-based measure incorporates land use, transportation, and individual components, but not the temporal component. Given the same location, a full-time worker and a retired individual will have different accessibilities, because of their different time constraints. One way to incorporate the temporal components is by calculating potential path space (PPS). This method is based on the concept of space-time prisms (Hagerstrand, 1970; Lenntorp, 1976; Burns, 1979;

Miller, 1991). The implementation of this measure is infeasible only until the tool of GIS and information from travel dairies become available. Kwan and her colleagues (Kwan and Hong, 1998; Kwan, 1998, 1999; Kwan and Weber, 2003; Kim and Kwan, 2003) described the detailed information on implementation. Briefly, the PPS for a fixed pair of locations i and j can be defined as:

$$PPS = \{(k, t) \mid t_i + \frac{d_{ik}}{v} \leq t \leq t_j - \frac{d_{kj}}{v}\}, \quad (4.65)$$

where,

t_i is the latest ending time of the activity at location i ,

t_j is the earliest starting time of the activity at location j ,

v is the average travel speed on the network, and

d_{ik} and d_{kj} are the distances from location i to location k , and from location k to location j respectively.

Generally speaking, PPS is a three-dimensional conception, with x-y coordinates representing the geographical position and z coordinate representing time of day. For ease of calculation, the three-dimensional PPS can be projected into a two-dimensional space, Potential Path Area (PPA). The PPA represents the locations that individuals could reach under temporal constraints.

The PPA only exists when there is an available time budget between two consecutive fixed activities conducted in location i and j . This time budget is the difference between the end of first fixed activity and the required starting time of second fixed activity. The space-time measure includes not only the land-use and transportation components, but also temporal and individual components. Although the space-time measure has the advantage of incorporating all components in accessibility, the computation can be burdensome.

4.5.2.3 The Selection of Accessibility Measurement

No measurement discussed above is a superior to others in all aspects. Genus (2006) concluded that an accessibility measure should satisfy five conditions. First, if the travel time or travel cost of any travel mode in an area increases (decreases), the measured accessibility in that area should decrease (increase). On the contrary, if reliability and safety of any travel mode in an area increases (decreases), the measured accessibility in that area should increase (decrease). Second, if the number of opportunities for an activity increases (decreases) anywhere, accessibility to that activity should increase (decrease) from anywhere. Third, if the demand for opportunities for an activity with certain capacity restrictions increases (decreases), accessibility to that activity should decrease (increase). Fourth, an increase of opportunities for an activity should not change the accessibility to that activity for any individual whose time schedule doesn't allow him to participate that activity. Finally, an improvement in one travel mode, such as reduced driving time, should not change the accessibility for any individuals who are not able to use that travel mode. For simplification purposes, the gravity formula is adopted to measure the accessibility of one location. The gravity formula satisfies the first and second conditions while not for the left three. Using an accessibility measurement approach satisfying all five conditions in residential location choice will make an already complex model more sophisticated, and I will leave it for the future work.

Chapter 5 Theoretical Framework and Methodology of Response

Lag

5.1. Introduction

Where one lives influences one's daily activity and travel pattern. How one's daily activity and travel pattern is shaped in time and space depends on not only one's needs and desires, but also the spatial constraints of the built environment. As I discussed earlier, a change in activity and travel pattern can be described in two dimensions: one relates to its magnitude and direction of the change and the other relates to a response lag, or time between a trigger and the resulting change. Almost all of the existing studies (Kitamura, 1989; Goodwin, 1989; Golob, 1990; Ma and Goulias, 1996, 1997; Goulias, 1999; Krizek, 2003) focus on the former: the magnitude and the direction of a change. The latter question remains largely unanswered. Namely, how long does it take individuals to change their activity patterns given a change in the built environment? To answer this question, I propose a survival model. In the following sections, I will provide a review of the current survival model, discuss the nature of censored data with multi-event exits, and develop a theoretical framework for my study of response lag.

This chapter is organized as follows. In Section 5.2, I will review of survival analyses. In a survival analysis, according to the way the data was collected, data can be non-censored, left-censored, right-censored, and left & right censored. Also, if there are multiple events that could define the end of survivor time, the survival analysis is called a multi-event survival model. The discussion about data censoring and multi-event survival analysis is provided in Section 5.3. In Section 5.4, I will discuss the theoretical framework of the survival model for the time allocation of discretionary activity.

5.2. Survival Analysis

5.2.1. Introduction

Survival analysis, also known as duration analysis, event history analysis, or reliability or failure time analysis, was originally developed in medical and biological areas, then also widely in the social, economic, and engineering sciences. The variable of interest is the length of time that elapses from the beginning of an event until either its end or the time when the measurement is taken, which may follow termination.

A typical example of applying the survival analysis in medical areas is how long a patient survives after receiving a treatment. Other applications might involve understanding the duration of a strike, the length of unemployment, and the lifetime of an electronic component. In my dissertation, I want to understand the response lag between residential relocation and the time when a significant change takes place the time people allocate to discretionary activities and travel.

In survival analysis, there needs to be a trigger that indicates the start of the study period, and a termination that signals the end of the study period. In the example of the length of a strike, the trigger is the onset of the strike and the termination is the end of the strike. If we observe both the trigger and the termination, we call the observation a “complete cell.” In cases when only the trigger is observed, the observation is right-censored; in cases when only the termination is observed, the observation is left-censored. In some cases, neither the trigger nor the termination are observed; this is called a left & right-censored cell. When the dataset is censored, it is important to deal with them correctly, as omitting them will bias estimation results (Kiefer, 1988). I will discuss ways to deal with censored data in survival analyses in Section 5.3.

Another problem in survival analyses is associated with multiple entrances and/or exits at the trigger and the termination. For instance, the trigger at the start of unemployment could be either a lay-off or a voluntary quit. The termination at the end of unemployment could be either being

employed or leaving the labor market. Variation in the duration under different types of triggers and termination is expected. For instance, people who voluntarily quit jobs are likely to be unemployed for a shorter time than those who are laid off. A mathematical way to deal with a multi-event survival model is called a Competing Risks Model, which will be discussed in Section 5.3.

5.2.2. Survival Analysis and Hazard Rate

Let random variable duration T have the cumulative probability distribution function $F(t) = \text{Pr ob}(T < t)$. Its corresponding density function is $f(t) = dF(t)/dt$. Therefore the probability of duration T is larger than t and can be written as $S(t) = 1 - F(t)$. The function $S(t)$ is called survival function. The hazard rate function can be expressed as:

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{\text{Pr ob}(t \leq T < t + h | T \geq t)}{h}. \quad (5.1)$$

The hazard rate function in equation 5.1 is essentially the conditional probability density function of duration T , given that it lasted until t . The relationship between hazard rate, survival function, density function, and cumulative distribution function can be written as follows:

$$\lambda(t) = f(t) / S(t) = F'(t) / S(t) = -S'(t) / S(t) = -(\ln S(t))'. \quad (5.2)$$

If we integrate equation 5.2, we have

$$S(t) = \exp\left(-\int_0^t \lambda(s) ds\right). \quad (5.3)$$

Parametrically, we can assume a distribution form for the hazard rate function. The hazard rate may be time independent or time dependent. In the time independent case, let $\lambda(t) = \gamma$, then $S(t) = \exp(-\gamma t)$. Accordingly, the cumulative density function and density function are $F(t) = 1 - \exp(-\gamma t)$ and $f(t) = \gamma \exp(-\gamma t)$ respectively. In the time dependent case, let

$\lambda(t) = \gamma\alpha^{\alpha-1}$, then $S(t) = \exp(-\gamma t^\alpha)$, $F(t) = 1 - \exp(-\gamma t^\alpha)$, and $f(t) = \gamma\alpha^{\alpha-1} \exp(-\gamma t^\alpha)$. The survival model with a constant hazard rate is called the exponential model while the model with the time dependent hazard rate introduced here is called Weibull model. It should be noted that other distributions could also be made on the hazard rate, for example, log-normal. I introduced the Weibull model here because it is widely used. The broad usage of the Weibull model may come from its flexible and simplistic hazard function. Flexible, because its hazard function can accommodate an increasing, decreasing, or constant hazard rate; simplistic, as compared with other models such as log-normal). In the Weibull model, when $\alpha > 1$, it indicates that the probability of ending the spell increases as time progresses; when $\alpha < 1$, it indicates that the probability of ending the spell decreases as time progresses. If $\alpha = 1$, the Weibull model becomes the exponential model.

A primary interest in survival analysis is to investigate what factors contribute to the length of a spell. One way to incorporate the explanatory factors in the survival model is the proportional hazard function. Let the hazard rate be written as:

$$\lambda(t, \mathbf{x}, \boldsymbol{\beta}, \lambda_0) = \phi(\mathbf{x}, \boldsymbol{\beta}) \lambda_0(t), \quad (5.4)$$

where $\lambda_0(t)$ is called the baseline hazard, $\lambda_0(t) = \gamma$ in the exponential model and $\lambda_0(t) = \gamma\alpha^{\alpha-1}$ in the Weibull model. $\boldsymbol{\beta}$ and $\mathbf{x}$ are vectors of parameters and explanatory variables respectively.

Since the hazard rate can never take a negative value (since there is no negative probability), function $\phi(\mathbf{x}, \boldsymbol{\beta})$ can be expressed as $\phi(\mathbf{x}, \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}' \mathbf{x})$. It should be noted that constant γ is standardized to 1 in practice. This is because $\exp(\boldsymbol{\beta}' \mathbf{x})\gamma$ is equivalent to $\exp(\ln(\gamma) + \boldsymbol{\beta}' \mathbf{x})$, γ can not be separately identified from the constant in $\boldsymbol{\beta}$. Under these specifications, a typical setting of the Weibull model with proportional hazard function is listed as follow:

$$\lambda(t, \mathbf{x}, \boldsymbol{\beta}, \lambda_0) = \exp(\boldsymbol{\beta}' \mathbf{x}) \alpha^{\alpha-1}, \quad (5.5)$$

$$S(t) = \exp(-\exp(\beta' \mathbf{x}) \int_0^t \alpha^{\alpha-1} dt), \quad (5.6)$$

$$F(t) = 1 - \exp(-\exp(\beta' \mathbf{x}) \int_0^t \alpha^{\alpha-1} dt), \quad (5.7)$$

$$f(t) = \exp(\beta' \mathbf{x}) \alpha t^{\alpha-1} \exp(-\exp(\beta' \mathbf{x}) \int_0^t \alpha^{\alpha-1} dt). \quad (5.8)$$

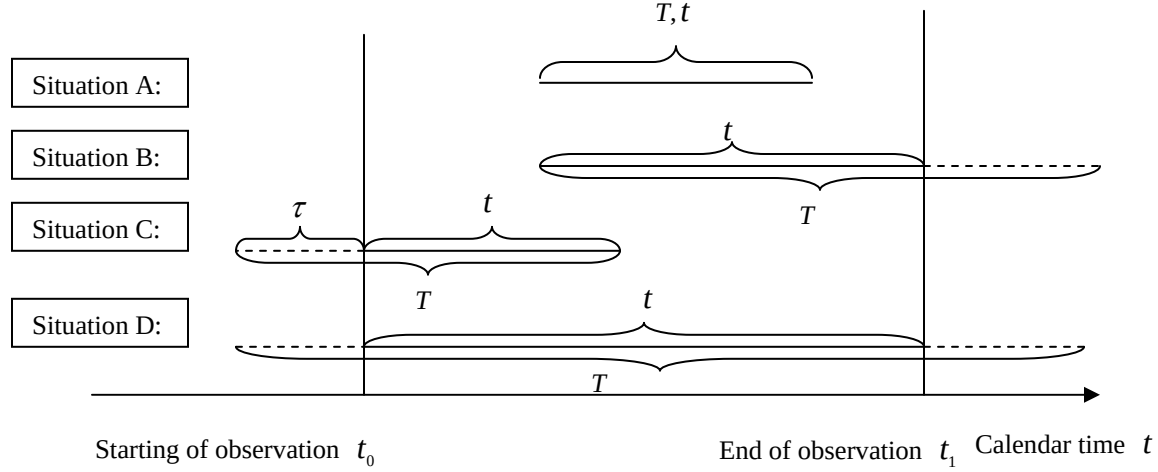
5.3. Censored Data, Partial Observation and Multi-event

There are three important issues we need to address in the survival model of response lag. First is about the censoring of data. This is because not everyone in the sample is observed with a home location change or a time allocation change. We have observations that are non-censored, right-censored, left-censored, or left & right censored. The standard survival model provides probability expressions for non-censored data and right-censored data, while left-censoring is not commonly treated. The second issue is also about censoring, which is a different type of censoring from the previous type. The response lag, by nature, is a continuous variable, while it is only observed at discrete time intervals. In other words, the dependent variable that we observe is censored due to partial observation. A methodology needs to be developed to address this issue. Both of these types of censoring need to be addressed within the context of a competing risks model framework, as an individual may choose to increase, decrease or have no change. In other words, there are multiple possible exits. In the following, we discuss each of these three issues separately.

5.3.1. Likelihood Expression for Censored Data

The parameter estimation of a survival model can be performed by the Maximum Likelihood Estimation (MLE). Before we discuss the likelihood expression, let's first discuss the censoring issue that often appears in the dataset. Censoring results in different likelihood expressions. There are four types of data points in survival analysis. They are exhibited in Figure 5.1.

Figure 5. 1: Censoring Types



In situation A, data is complete. T is observed in full length and is equal to observed time length t . The right-censored data arises in situation B, in which the end of the observed time length t proceeds the termination of the spell. If the trigger is not observed, as in situation C, the data is left-censored and the observed time length t is also shorter than actual duration T . Finally, if both the trigger and the termination are not observed, it is both left and right-censored.

For non-censored data, the likelihood is: $\Pr(t < T < t + \Delta) = \lambda(t)S(t) = f(t)$ and for right-censored data, the likelihood is: $\Pr(T > t, \mathbf{x}(T)) = S(t)$. The likelihood functions for both left-censored and left & right-censored data are more complicated.

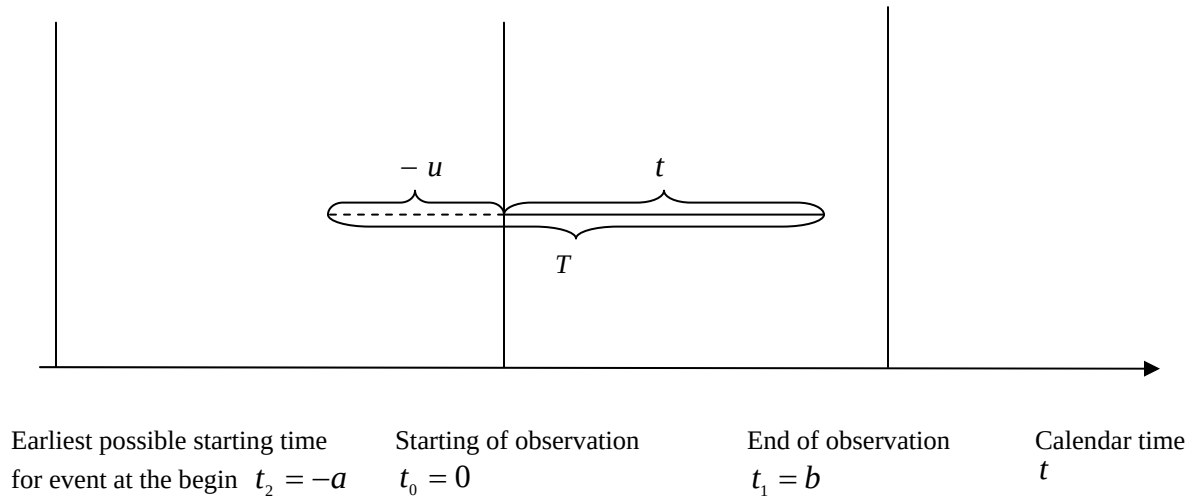
The likelihood function for complete and right-censored spells can be easily derived. The complete spell with duration $T=t$ means that it has already survived for time length t and it would end right after this time. Therefore, the probability of a complete observation is the product of two probabilities: surviving for time length t and ending right after time t given that it has survived for time t . The former probability is just $S(t)$ and the latter probability is simply hazard rate $\lambda(t)$. Therefore the likelihood expression for a complete spell is $\Pr(t < T < t + \Delta) = \lambda(t)S(t) = f(t)$. For the right-censored spell, it simply indicates that the

duration T is longer than t , but there is no information about its exact length. The likelihood expression for the right-censored spell, then, can be expressed as $\Pr(T > t) = S(t)$.

The case for left-censored data, however, is much more complicated than the two types just discussed. Note that even though both left-censored and right-censored data indicate that duration T is longer than t , the left-censored cell cannot use the same likelihood expression as the right-censored one. This is because that duration T being longer than t in the left-censored case is conditioned on that it has survived for an unknown amount of time length τ . If the hazard rate increases as time passes, the probability of this left-censored spell shall be smaller than $S(t)$.

Typically, these two types of data were not included in empirical studies, as “this type of censoring is difficult to handle in hazard-based models of duration” (Hensher and Mannering, 1994, p. 65). Discarding these two groups of data, however, is not an option for this study due to two reasons, since this will bring a biased result (Kiefer, 1988). Amemiya (1999) suggested using the conditional probability to deal with the left-censored data. Following Amemiya (1999), let the start of a study period be equal to zero, the termination be equal to b , and the earliest possible starting time for a spell to start be equal to $-a$.

Figure 5. 2: Left-Censored Data



The actual starting time of duration is $u(u < 0)$, which is unknown to researchers. Assume that u follows a density function, $g(u)$. The duration T still follows a density function $f(t)$, a cumulative function $F(t)$ and a survival function $S(t)$. It is assumed that u and T are independent. The observation of this left-censored data indicates that the spell starts at time u and its duration T is longer than $-u$. Let $A = \{\text{observing the left censored spell}\}$, then

$$\text{Pr ob}(A) = P(T > -u, -a < u < 0) = \int_{-a}^0 g(u)S(-u)du. \quad (5.9)$$

If random number u is already known, the conditional probability of observing the length of the spell being T is $f(t - u)$ (Noted that the duration before the starting of observation is $-u$, therefore T is equal to $t + (-u)$, see Figure 5.2.) The unconditional probability of observing the spell length being T is to integrate $f(t - u)$ with density function $g(u)$ over interval $[-a, 0]$, which is:

$$\text{Pr ob}(t - u < T < t - u + \Delta) = \int_{-a}^0 g(u)f(t - u)du. \quad (5.10)$$

Finally, the probability of observing a left-censored spell to have an observed length of t (note that actual duration is T) is expressed as a conditional probability:

$$\text{Pr ob}(t - u < T < t - u + \Delta \mid T > u, -a < -u < 0) = \frac{\int_{-a}^0 g(u)S(-u)du}{\int_{-a}^0 g(u)f(t - u)du}. \quad (5.11)$$

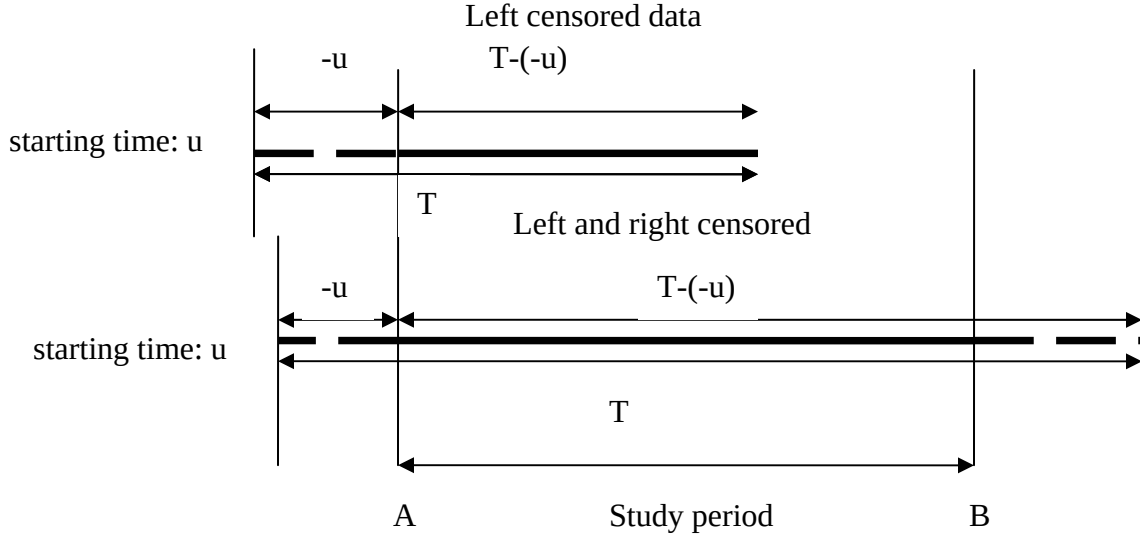
Amemiya (1999) provides probably the most complete treatment to date on left-censoring. Using the above-introduced approach, he studied a scenario where the start of a duration after the starting point of the study period is random. One example of this scenario is to study the duration of the survival time after a particular element in one's blood system is elevated. This scenario, however, is not applicable in the context of this study. I do not have the luxury of setting up a

common starting point and then observing the starting time of the duration of an event. Instead, the starting point for different individuals in the sample varies as the time of their first participation in the survey varies. In addition, once an individual participates in the survey, our observation starts. In other words, given the starting point of the survey participation, the start of the duration of the event in which I am interested (response lag) is certain, not random. Bhat (1996) was probably the first pioneering study in transportation that proposed a generalized multiple-duration proportional hazard model, which can accommodate left-censoring and left & right-censoring data. However, Bhat provided no empirical implementation for left-censoring.

I will now describe our treatment on left-censoring and left & right-censoring as follows. Figure 5.3 illustrates these two cases, where the solid line refers to the observed and the dotted line refers to the unobserved. Point A is the starting point of the study period and Point B is the ending point. The length of study period is then AB . In the context of this study, Point A is the wave when the subject participates in the survey and Point B is the fifth year of survey participation⁶. For both left-censoring and left & right-censoring, a home location change takes place prior to the survey participation time and is unobserved by us. We assume that the starting time of a home location change, U , follows a probability density function $g(u)$. If we set Point A as the starting point and it is equal to zero, the value that the starting time of a home location change, U , can take, is always negative. The duration since the home location change, T , is assumed to follow a probability density function $f(t)$. Note that in left-censored data, T includes two parts: $-u$ and the duration we observed since the survey starts. In addition, in left & right-censored data, T includes three parts: $-u$, the total study period, and the period after point B. We further assume that u and T are independent with each other.

⁶ This implies that AB is always equal to 5.

Figure 5. 3: Left-Censored Data and Left & Right-Censored Data



Following Amemiya's method introduced in this section, given that a home location change occurs prior to the start of the study period, the likelihood of a left-censored observation is $\Pr(t < T + u < t + \Delta \mid T > -u, u < 0)$ and the likelihood of a left and right-censored observation is $\Pr(T + u > AB \mid T > -u, u < 0)$. Let event $A = \{\text{observe a left censored observation}\}$, which is: $\{T > -u, u < 0 \mid u, T\}$, the probability of event A is:

$$\Pr(A) = \Pr(T > -u, u < 0) = \int_{-\infty}^0 g(s)[1 - F(-s)]ds = \int_{-\infty}^0 g(s)S(-s)ds, \quad (5.12)$$

where $S(-s) = 1 - F(-s)$. Based on theorem of conditional probability, the likelihood of observing a left-censored observation in Figure 5.3:

$$\Pr(t < T + u < t + \Delta \mid A) = \frac{\int_{-\infty}^0 g(s)f(t-s)ds}{\int_{-\infty}^0 g(s)S(-s)ds}. \quad (5.13)$$

The derivation for the likelihood function of a left & right-censored observation is similar to the one derived above for the left-censored observation. It is:

$$\Pr(T + u > AB \mid A) = \frac{\int_{-\infty}^0 g(s)S(AB - s)ds}{\int_{-\infty}^0 g(s)S(-s)ds}. \quad (5.14)$$

Remember that we do not observe the starting time of the home location change at all. Therefore, no information is available about $g(u)$. Therefore, we need to make a further assumption about $g(u)$. We assume $g(u)$ is a constant when $s \in [-100, 100]$ and zero when $s \in (-\infty, -100) \cup (100, \infty)$.⁷ With this assumption, the probability of observing a left-censored observation, given that the home location change takes place prior to the study period is:

$$\Pr(t < T + u < t + \Delta \mid A) = \frac{\int_{-100}^0 f(t - s)ds}{\int_{-100}^0 S(s)ds}. \quad (5.15)$$

Similarly, the probability of obtaining a left & right-censored observation, given that the home location change takes place prior to the study period is:

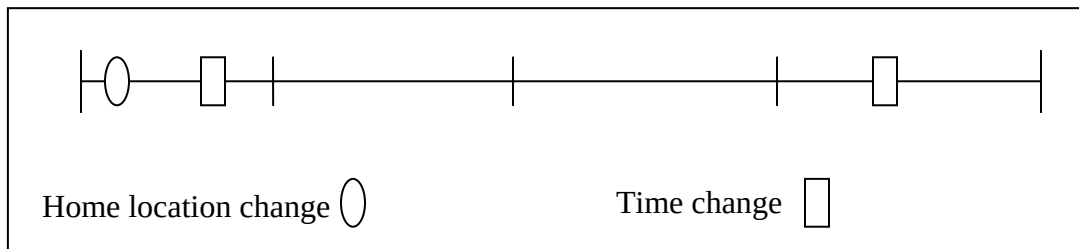
$$\Pr(T + u > AB \mid A) = \frac{\int_{-100}^0 S(AB - s)ds}{\int_{-100}^0 S(s)ds}. \quad (5.16)$$

5.3.2. Treatment on Partial Observation

⁷ For a continuous integration $\int_{-\infty}^{\infty} g(s)ds$, the value of integration is infinitive when $g(s)$ is constant for $s \in (-\infty, \infty)$. Since human's life usually is less than 100 years, therefore it is reasonable to assume the density function of starting time of home location change is constant when $s \in [-100, 100]$ and zero when $s \in (-\infty, -100) \cup (100, \infty)$. Please note that here “-100” means 100 years prior to study period.

By nature, the dependent variable we are interested in (the response lag) is a continuous variable. However, the Puget Sound Panel Survey was only conducted at discrete time intervals, and home location changes are only observed at discrete intervals. In other words, due to partial observation, the response lag observed as the result of these discrete time intervals is censored. The actual time of occurrence for a change takes place at any point between two survey points. Figure 5.4 illustrates this type of censoring due to partial observation. The oval shape in the figure refers to a home location change, while the rectangular shape refers to a change in time allocation. If a time allocation change happens during the same time interval as the home location change (as is shown between Wave 1 and Wave 2 in the figure), the true response lag would be less than 1 year. If a time allocation change happens in a different time interval compared with the home location change, the true response lag would fall in an interval of two years. For example, if a time allocation change happens between Wave 4 and Wave 5 (as is shown in the figure), because we do not know the exact positions of these two changes in time, the true response lag would be between 2 and 4 years.

Figure 5. 4: Illustration of Censoring Due to Partial Observation



Mathematically, we can express this type of censoring as follows. Let T_i be the true response lag for individual i . T_i is not observed by us. What we observed is the discrete time interval, w , which can only be one year, two years, three years, four years, or five years, given the 5 consecutive year criteria. Generally speaking, the probability of observing a response lag of k is equivalent to the probability of having the true response lag, T_i , to be between $k-1$ and k , if we observe a time allocation change and a home location change during the same time interval; and to be between $k-2$ and k , if we observe a residence change first and a time allocation change after a subsequent wave. Mathematically, we can express the following: $\Pr(w = k) = \Pr(k - 1 < T_i \leq k) = \Pr(T_i \leq k) - \Pr(T_i \leq k - 1)$, if we observe a time allocation change

and a home location change during the same time interval and $\Pr(w = k) = \Pr(k - 2 < T_i \leq k) = \Pr(T_i \leq k) - \Pr(T_i \leq k - 2)$ if we observe a home location change first and a time allocation change in a subsequent wave.

For a non-censored observation, the likelihood is simply $\Pr(w = k) = F(k) - F(k - e)$. For a left-censored observation, the likelihood is simply:

$$\Pr(w = k) = \frac{\int_{-\infty}^0 g(s)F(k - s)ds}{\int_{-\infty}^0 g(s)S(-s)ds} - \frac{\int_{-\infty}^0 g(s)F(k - e - s)ds}{\int_{-\infty}^0 g(s)S(-s)ds}, \quad (5.17)$$

where $e = 1$ or 2 , depending on whether we observe a home location and a time allocation change during the same time interval or not. For observations that are right-censored or left & right-censored, the likelihood functions would be unchanged, since we did not observe any change in time allocation. In this case, the likelihood for a right-censored observation would be: $S(t)$ and the likelihood for a left and right-censored observation would be:

$$\Pr(T + u > AB \mid A) = \frac{\int_{-100}^0 S(AB - s)ds}{\int_0^{100} S(s)ds}. \quad (5.18)$$

5.3.3. Multi-event

Another issue that we need to deal with is the multi-type event. Since a time allocation change can be either an increase or a decrease, we should further modify the above-developed probability expressions such that an exit can be either an increase exit or a decrease exit. When there are multiple events representing the trigger or termination of survival period, it is called a Multi-Event Survival Model. If there are multiple entrances, it is called a Multi-entrance Survival Model; if there are multiple exits, it is called Multi-Exit Survival Model. A Competing Risks Model is designed to handle these situations (Hachen, 1988).

Assume that there are m exits at the termination⁸ in a survival model. We use the same definition for the overall hazard rate $\lambda(t)$ (shown in Equation 5.1), which is the conditional probability density function of observing a spell in a short time interval, given that the spell has at least survived until time t . This hazard rate can be broken down into a set of type-specific rates $\lambda_j(t)$, which indicates the probability of observing exit j , given that the spell has survived at least t .

$$\lambda_j(t) = \lim_{h \rightarrow 0} \frac{\text{Prob}(t \leq T < t + h, \text{Exit} = j \mid T \geq t)}{h}, \quad (5.19)$$

for $j = 1, 2, \dots, m$. For instance, divorce the passing of a spouse may cause the termination of a marriage. There are two important relationships between type-specific rates and the overall hazard rate. First, the sum of type-specific rates should be equal to the overall hazard rate (equation 5.20) and, second, each type-specific rate is equal to the overall hazard rate times the conditional transition probability m_j (equation 5.21).

$$\sum_j \lambda_j(t) = \lambda(t), \quad (5.20)$$

$$\lambda_j(t) = \lambda(t)m_j. \quad (5.21)$$

The conditional transition probability m_j is the probability of a type j event, given that the spell makes j th exit.

The estimation of a Competing Risks Model could be done by assuming a distribution for the type-specific hazard rate $\lambda_j(t)$ or for the overall hazard $\lambda(t)$ and conditional transition probability m_j . For example, the hazard rate for the type-specific hazard $\lambda_j(t)$ under the Weibull distribution assumption is:

⁸ The derivation is same for the case that there are m events at the start.

$$\lambda_j(t, \mathbf{x}, \boldsymbol{\beta}, \lambda_0) = \exp(\boldsymbol{\beta}_j' \mathbf{x}_j) \alpha t^{\alpha-1}. \quad (5.22)$$

$\boldsymbol{\beta}_j$ and $\mathbf{x}_j$ are vectors of parameters and explanatory variables for event j respectively. Alternatively, we can estimate the model with $\lambda(t)$ and m_j . Note that $\sum_j m_j = 1$, we assume that m_j takes the form as:

$$m_j = \frac{\exp(\boldsymbol{\chi}_j' \mathbf{y}_j)}{\sum_j \exp(\boldsymbol{\chi}_j' \mathbf{y}_j)}, \quad (5.23)$$

where $\boldsymbol{\chi}_j$ and $\mathbf{y}_j$ are vectors of parameters and explanatory variables for the conditional transition probability. The overall hazard rate $\lambda(t)$ has the following form:

$$\lambda(t, \mathbf{x}, \boldsymbol{\beta}, \lambda_0) = \exp(\boldsymbol{\beta}' \mathbf{x}) \alpha t^{\alpha-1}. \quad (5.24)$$

The proper method to estimate the model depends on the individual case. Estimation with type-specific rates implies that these rates don't affect each other and they can be estimated separately, while estimation with conditional transition probability and overall hazard implies dependency between events.

For time allocation change, let function $m(t)$ be the probability of having an increase in time allocation, given that a change has occurred at time t . Therefore, the probability of $1-m(t)$ is the probability of having a decrease in time allocation, given that a change has occurred at time t .

Let $\ln \frac{m(t)}{1-m(t)} = \boldsymbol{\chi}' \mathbf{y}$ (Hachen, 1988), where $\boldsymbol{\chi}$ is parameter vector and $\mathbf{y}$ is the explanatory variables vector (note that in a binomial case, this is another form of formula 5.23). The positive sign of $\boldsymbol{\chi}$ suggests an increase while the negative sign of $\boldsymbol{\chi}$ indicates a decrease. In terms of the magnitude, the larger a value is, the more likely an increase (decrease) would take place.

Reorganizing terms in $\ln \frac{m(t)}{1-m(t)} = \boldsymbol{\chi}' \mathbf{y}$, we have $m(t) = \frac{e^{\boldsymbol{\chi}' \mathbf{y}}}{1 + e^{\boldsymbol{\chi}' \mathbf{y}}}$ and $1-m(t) = \frac{1}{1 + e^{\boldsymbol{\chi}' \mathbf{y}}}$.

Multiplying $m(t)$ or $1 - m(t)$ with the likelihood functions developed in Section 5.4.2 will give the final likelihood functions in the context of a multi-type event.

5.3.4. General Likelihood Expression

Let us now turn to writing a General Expression of Likelihood function for i^{th} observation, which may belong to any of the four groups (non-censored, right censored, left censored, or left & right censored) and may have either an increase exit, a decrease exit, or no exit. Let E_{pq} be the dummy variable indicating the group membership and the type of exit for an observation;

$E_{pq} = \begin{cases} 1 & \text{if the observation belongs to p group and has a type q exit} \\ 0 & \text{otherwise} \end{cases}$. Given that we have only

four groups (non-censored, right-censored, left-censored, and left and right-censored), p can only take one of four values ($p=1$ for non-censored data; $p=2$ for right-censored data; $p=3$ for left-censored data and $p=4$ for left & right-censored data). For q , it can only take one of the two values, since an exit can only be either an increase or a decrease ($q = 1$ for an increase in time allocation, $q = 2$ for a decrease in time allocation, $q = 3$ for no change in time allocation). For i^{th} observation, its likelihood function would be:

$$\begin{aligned} & \prod_{p=1}^4 \prod_{q=1}^3 (F(k) - F(k - e))^{E_{1q}} \bullet S(t)^{E_{2q}} \bullet \left(\frac{\int_{-\infty}^0 g(s)F(k - s)ds}{\int_{-\infty}^0 g(s)S(-s)ds} - \frac{\int_{-\infty}^0 g(s)F(k - e - s)ds}{\int_{-\infty}^0 g(s)S(-s)ds} \right)^{E_{3q}} \\ & \bullet \left(\frac{\int_{-100}^0 S(AB - s)ds}{\int_0^{100} S(s)ds} \right)^{E_{4q}} \bullet \left(\frac{e^{x'y}}{1 + e^{x'y}} \right)^{E_{p1}} \bullet \left(\frac{1}{1 + e^{x'y}} \right)^{E_{p2}} \bullet 1^{E_{p3}}. \end{aligned} \quad (5.25)$$

In this expression, $e = 1$ or 2 , depending on when we observe a home location change and a time allocation change.

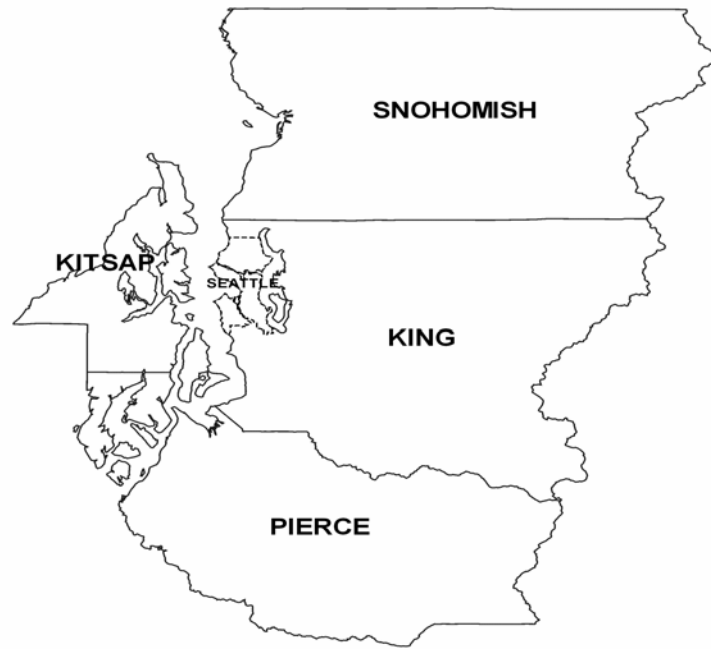
Chapter 6 Empirical Residential Location Choice Study

6.1 Empirical Dataset

Area

The dataset used in the empirical residential location choice study is from the transpiration panel survey conducted in Puget Sound area, which is located in the west side of Washington state (see fig 6.1). There are four counties in the Puget Sound region, including King, Kitsap, Pierce and Snohomish. Facing the Pacific Ocean, the Puget Sound is the bay area in state of Washington. King County is located in the center of Puget Sound. According to 2000 census, there were more than 1.7 million people living in King County. It is the largest county by population in Washington state. Seattle is located by the seashore and it is the largest city in the county. Kitsap County is on the other side of the bay and there are over two hundred thousand people living there. Situated in the south of King , Pierce County is the second-most populous county in the State of Washington. As of 2000, over 700,000 people lived there. Snohomish County is north of King County, and its population was about 600,000 in 2000. The Puget Sound is composed of these four counties and had over 3 million people as of 2000.

Figure 6. 1 Puget Sound Region



Dataset

The survey dataset used in the residential choice study is the Puget Sound Transportation Panel (PSTP), conducted from 1989 to 2002. Initiated in 1989, the purpose of the panel survey was to monitor changes in household composition, location, and employment characteristics, to monitor changes in travel behavior and responses to changes in the transportation environment, and to examine the effects of changes in attitudes and values on mode choice and travel behavior. Until 2002, there were ten waves of survey, conducted consecutively in 1989, 1990, 1992, 1993, 1994, 1996, 1997, 1999, 2000, and 2002.

The survey contains information on household and individual trips in a 2-day diary for participants more than 15 years old. The information available in PSTP is summarized in Table 6-1.

Table 6- 1 PSTP Dataset Information

Category	Primary information
Household level	➤ Lifecycle (adults with children, adults without children and younger than 35, etc.) ➤ Household size and household compositions (number of adults, number of children) ➤ Income ➤ Auto ownership ➤ Number of employed household members ➤ Residential location (county, city, zip code, census tract, and traffic analysis zone)
Personal level	➤ Age ➤ Sex ➤ Employment status ➤ Occupation ➤ Work location (county, city, zip code, census tract, and traffic analysis zone)School location (city) ➤ Commute mode ➤ Commute distance
Travel dairy	➤ Total trips ➤ Trip purpose* ➤ Trip mode ➤ Trip time (starting time/ending time) ➤ Trip duration ➤ Activity duration ➤ Trip destination
Attitude	➤ Attitude toward travel model ➤ Attitude toward neighborhood ➤ Attitude toward transportation service

* The trip purposes used in the two surveys were slightly different. For wave 1 to 7, the following trip purposes were recorded: work, shopping, school, visiting, free-time, personal, appointments and home. Among them, visiting, free-time and personal were classified as discretionary. For wave 8 to 10, the following trip purposes were recorded: home, commute to work, other work-related travel, commute to school, taking child to/from school, lesson etc., visiting friends or family, errands/picking up/dropping of others, doctor appointments/other medical-related, delivery, shopping, dining/coffee/etc., recreation/exercise/lessons/personal business, volunteering, going to another travel mode, other appointment or meeting, meet/pick up/drop off carpoolers, linked commute trip to/from work; commute from work, commute from college, commute from school and other. Among them, visiting friends or family, dining/coffee/etc, recreation/exercise /lessons/personal business, and volunteering are classified as discretionary.

Sample Size and Sample Selection

The entire dataset comprises 7,659 households. In each wave, socio-demographics, home location and work location were recorded. Census tracts of the Puget Sound are used as residential alternative locations in the current analysis for their easy identification and homogeneity. There are a total of 713 alternative locations in the Puget Sound region.

Not all records were used in the analysis. Only movers (households who relocated at least once) were included. In addition, if a household head or his/her spouse was unemployed, or employed but did not report his/her work location, his/her corresponding record was removed. Furthermore, if a household did not report its home location or household size, the record was also removed. According to these criteria, the resulting sample contains data from 566 households.

6.2 Explanatory Variables

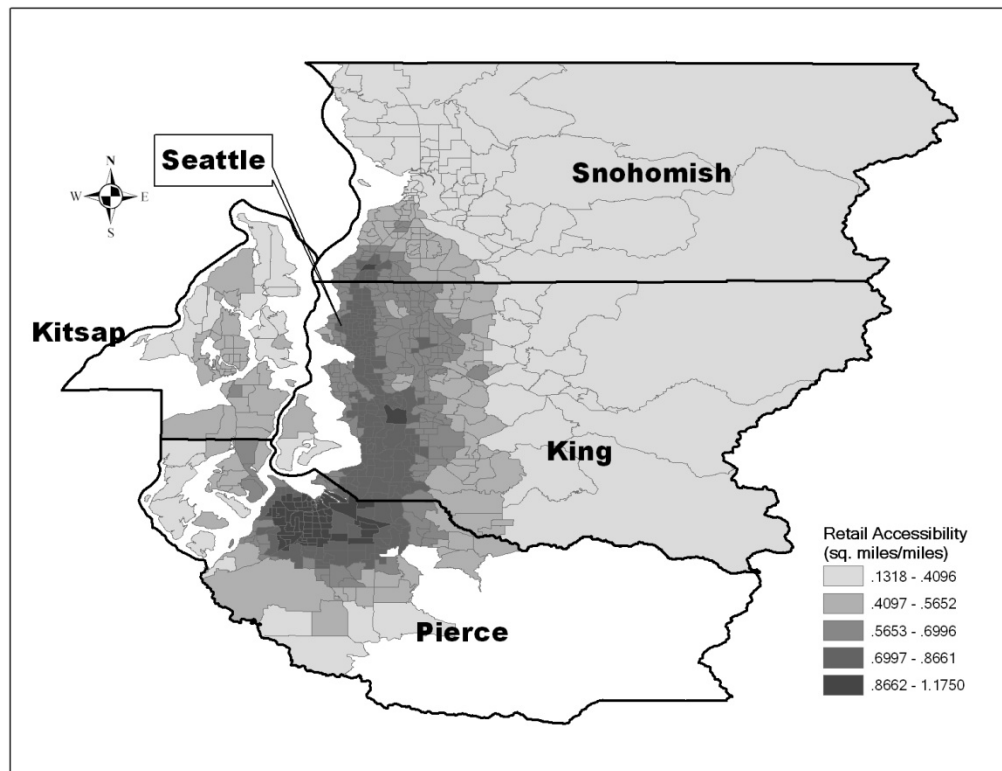
Accessibility variables

Four types of accessibility are considered as influential factors in the residential choice model – accessibility to work, accessibility to open space, accessibility to retail, and accessibility to recreation opportunities. The measurements of four types of accessibility are described as follows.

There are two types of measurements for accessibility to employment. The first type is the accessibility index, which describes the access to all types of job opportunities around ones' residential locations. Travel time or distance is the impedance factor, which is to penalize those opportunities located far away from a location. The second type is the commute distance or commute time, which describes the access to current job location. Many empirical studies (Timmermans et al., 1996; Waddell, 1996; Molin, et al., 1999; Srour et al., 2002; Guo and Bhat, 2004) have demonstrated the significant role of commute distance in residential location choice, while Deitz(1998) found that the accessibility index measure (often computed as a gravity-based measure) is not a significant explanatory variable for location choice. Probably commute distance is an important attribute considered by households while the accessibility index measure is not a good proxy of commute distance. A large accessibility index doesn't necessarily correspond to a short commute distance. In this study, I used the commute distance as the measurement of accessibility to employment.

Accessibilities to open space, retail, and recreational opportunities are measured by the gravity-based formula. Or, $A_j = \sum_{t=1}^T G_{ti} / D_{ij}$, where A_j is the accessibility of zone j , G_{ti} is the total floor area of t^{th} facility, and D_{ij} is the distance between t^{th} facility and zone j . The open space variable accounts for opportunities such as parks, golf courses, resorts, camps, vacant/undeveloped lands, wetlands, forest, reforestation, tidelands, ski areas, greenbelts, wilderness areas, athletic areas, and water areas (such as lake, river, bay). The retail land-use category includes opportunities such as community shopping centers, grocery stores, malls, gas stations, furniture and appliance stores, and clothing stores. Recreational opportunities include parks, restaurants, lounges, taverns, entertainment bars, sports facilities, libraries, theaters, amusements, golf courses, resorts camps, neighborhood/community/regional centers, public assemblies, bowling allies, ski areas, skating rinks, art galleries, museums, and clubs. Figure 6.2 depicts the distribution of accessibility to retail opportunities. In general, Census tracts close to the coast have more retail opportunities.

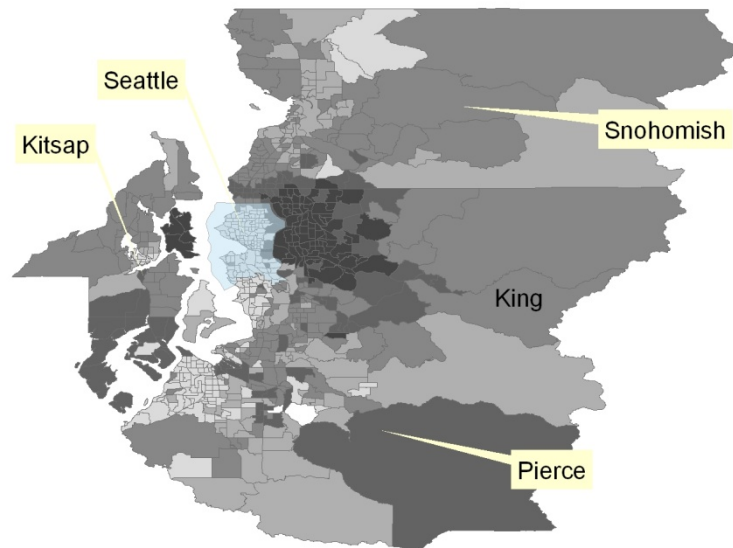
Figure 6. 2: Retail Accessibility in Puget Sound (2000)



Neighborhood Variables

The neighborhood attributes considered in our residential choice model include school quality, housing value per room, and population density. School qualities were obtained from the Washington Assessment of Student Learning 2002 (WASL, 2002). I defined the variable as the percentage of grade 10 students passing the standard requirement of three subjects (reading, math, and science). The level of school quality is shown in Figure 6.3. Suburban areas of Seattle seem to have relatively higher school quality in comparison to other Puget Sound areas.

Figure 6. 3: School Quality in Puget Sound (2000)



The housing value per room is calculated by the average housing value, divided by the average number of rooms in one census tract. The distribution of housing value and population density is shown in Figures 6.4 and 6.5. For population density, areas along the coast have higher densities than those in other parts of the Puget Sound. Areas with higher housing value are mostly located in and around Seattle, as well as parts of Kitsap County. Notice that Census tracts with higher housing value usually correspond to a higher school quality.

Figure 6. 4: Population Density in Puget Sound (2000)

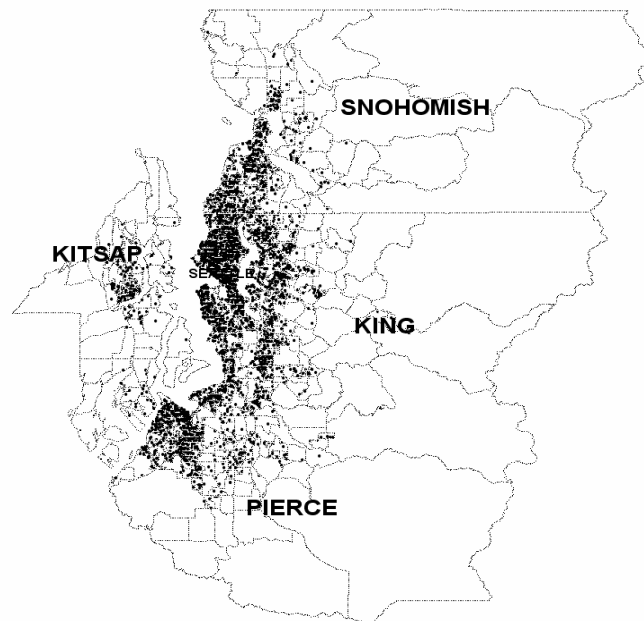
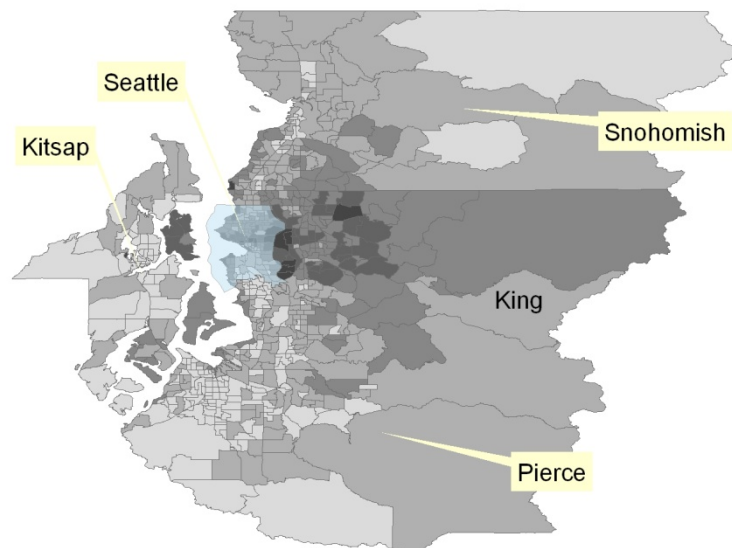


Figure 6. 5: Distribution of Housing Value per Room in Puget Sound (2000)



Household Level Variable

The Household Level Variable is defined by childbearing and non-childbearing households. The assumption is that households in different stages of a lifecycle may show variations in terms of the historical deposition influence.

The socio-economic and demographic attributes of the childbearing households and non-childbearing households are shown in Table 6-2. Two-hundred fifteen households have children under 17-years-old, and the rest do not. Comparing between pre- and post-relocation, both types of households experienced longer commute distances, an indication for suburbanization. Changes in the other three accessibilities are not. Comparing between households with and without children, the former have more household members, longer commute distances, and fewer retail and recreational opportunities than the latter, both before and after the relocation. As for open space, childbearing households have more open space than those without children, pre-relocation. The order is reversed post-relocation: households with children have less open space than those without children. Non-childbearing households also face higher housing prices per room, and higher density but lower school quality than childbearing households do, suggesting that the former are more likely to reside in cities.

Table 6- 2 Descriptive Statistics of the Sample and the Study Region

Variables	Prior to relocation		After relocation	
<u>Non-child-bearing households (n=346)</u>	Means	Std	Means	Std
Household size	2.22	1.48	1.81	0.68
Number of children under 17 years old	0.18	0.56	0	0
Commute distance (miles)*	10.63	9.15	12.29	10.93
Open space (miles)	22.03	4.23	22.14	4.53
Retail (miles)*	0.64	0.18	0.63	0.18
Recreation (miles)*	1.57	0.42	1.57	0.31
Avg. housing price per room			29,643	12,630
School quality			20.06	2.48
Population density			4,157	4,294
<u>Child-bearing households (n=215)</u>	Mean	Std	Mean	Std
Household size	3.41	0.99	3.67	0.87
Number of children under 17 years old	1.46	0.87	1.66	0.74
Commute distance (miles)*	11.92	9.96	14.44	12.15
Open space (miles)	23.05	15.13	21.50	4.18
Retail (miles)	0.61	0.19	0.57	0.20
Recreation (miles)	1.54	0.34	1.48	0.31
Avg. housing price per room			28,555	11,144
School quality			20.76	2.4
Population density			2,863	2,366

* Prior-relocation means are significantly different from after-relocation means.

6.3 Model Estimation Results

Before the model estimation, I conducted a correlation analysis, which suggested a substantial correlation between retail, open space, and recreational opportunities. This observation supports the view that the housing market is highly constrained (Kreibich and Petri, 1982). The implication in modeling is that a simultaneous estimation of these three accessibilities in one model will disguise the true effect of a single variable. I thus estimated three models, including variables on retail, open space, and recreational opportunities separately. The other variables contained by each of the three models are the same.

I first discuss variables that are not accessibility related (Table 6-3). Two school quality variables were estimated, one for households with children and the other for households without children.

Both are positive and significant, indicating that locations with better schools are preferred. Furthermore, the estimate for households with children is larger in magnitude than that for households without children, implying a higher value placed by the former on school quality.

Table 6- 3 Estimation Results of the Distance-based GEV Models

	Model 1 (retail only)		Model 1 (open space only)		Model 3 (recreation only)	
Variables	With children	W.O. children	With children	W.O. children	With children	W.O. children
<i>Commute distance</i>	β_{1c}	β_{1nc}	β_{1c}	β_{1nc}	β_{1c}	β_{1nc}
α_{11}	-1.13	-1.24	-1.10	-1.25	-1.13	-1.25
α_{12}	-1.42	-0.84	-1.49	-0.85	-1.37	-0.81
<i>Retail</i>	β_{2c}	β_{2nc}	β_{2c}	β_{2nc}	β_{2c}	β_{2nc}
α_{21}	2.90	3.03				
α_{22}	-6.33	-5.02				
α_{23}	-1.00*	-0.64				
<i>Open space</i>	β_{3c}	β_{3nc}	β_{3c}	β_{3nc}	β_{3c}	β_{3nc}
α_{31}			0.38*	0.64		
α_{32}			-3.53*	-3.74*		
α_{33}			-0.12	-0.08		
<i>Recreation</i>	β_{4c}	β_{4c}	β_{4c}	β_{4c}	β_{4c}	β_{4c}
α_{41}					2.15	2.27
α_{42}					-5.06	-4.12
α_{43}					-0.61	-0.43
<i>School quality</i>	β_{5c}	β_{5c}	β_{5c}	β_{5c}	β_{5c}	β_{5c}
β_5	0.10	0.05	0.11	0.06	0.10	0.06
<i>Population density</i>	β_6		β_6		β_6	
β_6	-0.02*		-0.19*		-0.13*	
<i>Avg. house value per room</i>	β_7		β_7		β_7	
β_7	-0.008*		-0.007*		-0.007*	
λ	4.61		3.86		3.79	
Log likelihood (β)	-3232.59		-3222.17		-3227.01	

* Coefficients not statistically significant at 5% level.

The negativity of the population density variable in all three models seems to accord with the continued growth of the suburbs in metropolitan areas (Clark et al., 2006), but the estimates are insignificant. It is likely that density acts as a proxy for a combination of factors that might affect one's residential location choice (Chen et al., 2008).

The coefficient for the housing variable: average housing price per room, is negative and insignificant. This observation is in contrast with findings in other studies (Clark et al., 2006), which indicate an upward preference in relocation. It is possible that housing price does not fully account for the concept of an upward move. On the other hand, residential relocation choice is highly constrained, thus even though an upward move is preferred, affordability may prevent households from actually executing relocation decisions based on their preferences.

Figures 6.6 to 6.9 depict the relationship between one's preference toward commute distance, retail, open space, and recreational opportunities and the corresponding accessibility at the prior location. As expected, the estimates of commute distance are all negative, regardless the commute distance at the prior location. This is consistent with our expectations. Estimates of accessibilities to retail and recreational opportunities fall in both positive and negative domains, although most are positive. Behaviorally, it indicates that retail and recreational opportunities can contribute both positively and negatively, probably corresponding to the mixed results found in the literature (e.g, Hunt, et al., 1994; Timmermans et al., 1996; Srouf et al., 2002). Even though they are more likely to contribute a positive utility than a negative one, a low level of accessibility at the prior location could drive one's preference toward that attribute into the negative domain. In both cases, the point at which one's preference becomes positive is approximately 1 mile. We may view this point as the minimum threshold value to make any accessibility meaningful and valuable to people. The case for open space is more complicated in that the estimate changes from being negative to positive when the accessibility at the prior location falls in the 5-8 mile range. Thus, open space appears to require more abundance than retail and recreational opportunities in order for it to be meaningful.

Figure 6. 6 Parameter Estimates for Commute Distance

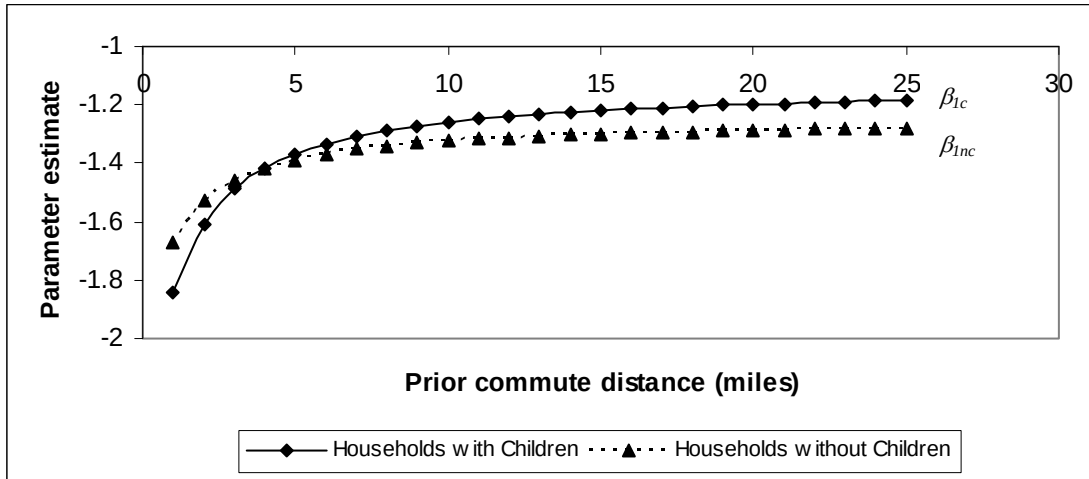


Figure 6. 7: Parameter Estimates for Retail Land Use

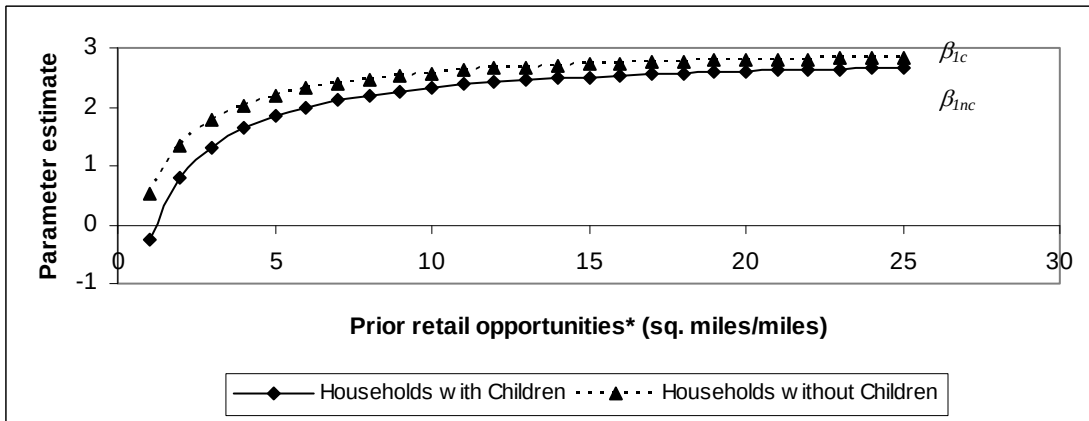


Figure 6. 8: Parameter Estimates for Open Space

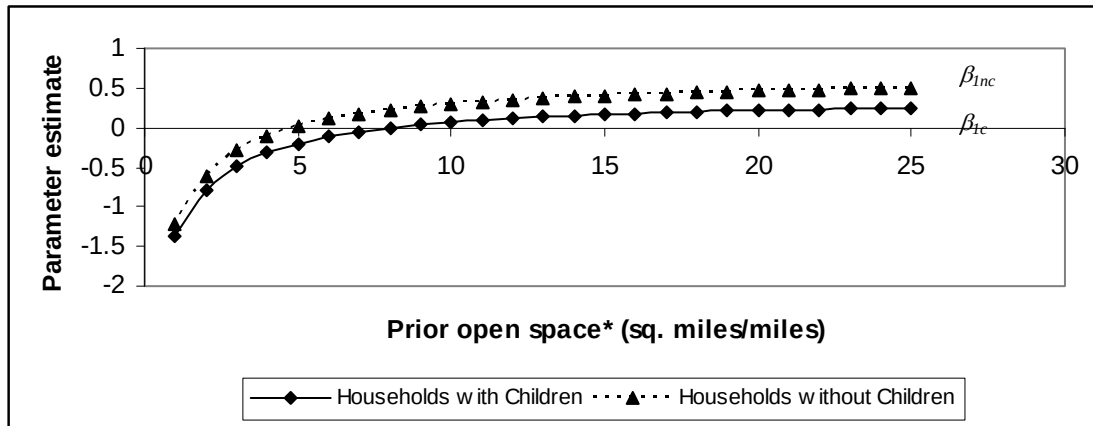
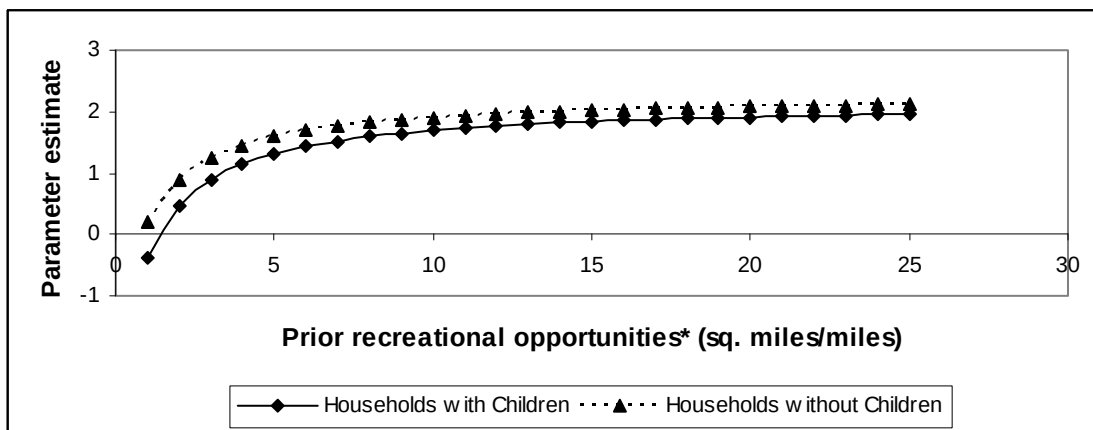


Figure 6. 9: Parameter Estimates for Recreational Land Use



We now examine how one's preference will change with respect to the accessibility experienced at the prior location. For commute distance, we observe that as prior commute distance increases, the estimate for commute distance becomes less negative. Or, people are less sensitive to long commute distances if they have lived with long commutes before. For the other three accessibility attributes, we observe that one values them more (or less negative if the estimate falls in the negative domain), as the accessibility at the prior location increases. The behavioral interpretation is that if one has lived in a location with many retail, open space, or recreational opportunities, he/she would value them more than someone who has not been exposed to them in a large quantity. To summarize, we are more tolerant if we have suffered and more acquisitive if we have enjoyed.

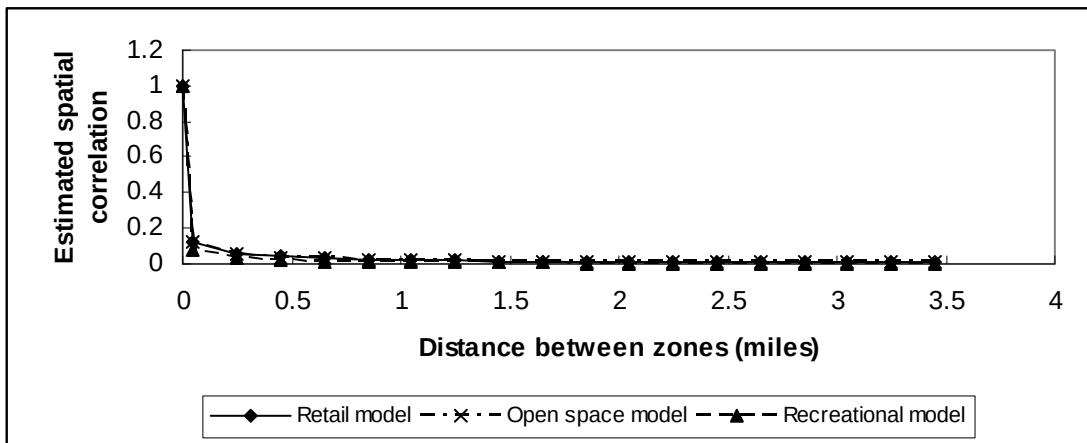
We further notice that the effect prior location has on preferences diminishes as prior accessibility increases. This is true for all four accessibility attributes. In particular, 10-25 miles seems to be the range after which the additional weight brought by prior accessibility approaches zero. Behaviorally, it implies that a person whose prior commute distance was more than 10 miles would value commute distance equally to one whose prior commute distance was exactly 10 miles. A person who has been exposed to more than 20 miles of retail, open space, or recreational opportunities would value them the same as one who has been exposed to these opportunities at the exact level of 20 miles.

To assess the interaction between prior location and lifecycle, we compare the preference curves between childbearing and non-childbearing households as the accessibility at the prior location changes (Figures 6.6 to 6.9). In the case of commute distance, the two preference curves cross at approximately 4 miles. Prior to a 4 mile commute distance at the prior location, childbearing households are more sensitive to commute distance than non-child-bearing households, while after the 4 mile threshold, the former becomes less sensitive than the latter. After about 16 miles, the difference between the two stabilizes. For retail, non-childbearing households value more than childbearing households do. The differences between the two close in as retail opportunities at the prior location increase. Similar observations apply to open space and retail opportunities: non-childbearing households always value more than childbearing households do. The disparities between childbearing and non-childbearing households decrease as prior accessibilities increase, however, they are quite different for these two types of accessibilities. For open space, the differences between the two enlarge and stabilize when prior open space reaches 9 miles. For recreational opportunities, the difference closes in as prior accessibility increases.

Estimates of the quadratic term (α_{13}) are negative for retail, open space, and recreational opportunities. This is in contrast to the general positivity observed in one's preference toward these accessibilities. This finding provides support to our hypothesis that having the right amount is the most preferred.

Lastly, λ in the similarity index function ρ_{ij} is estimated to be 4.61, 3.86, and 3.80 for retail, open space, and recreational models respectively. All are significant at the 10% level; λ in the recreational model is significant at the 5% level. These estimates confirm the existence of spatial correlation. Figure 6.10 shows the relationships between the estimated spatial correlation and the distance between zones, using the three estimated values of λ . The three curves are highly similar to each other. The estimated correlation drops significantly initially. At 3.5 miles, the estimated correlation is about 0.005, which is close to zero and suggests that our use of the threshold distance of 3.5 miles is appropriate.

Figure 6. 10: Estimated Spatial Correlation vs. Distance between Zones (miles)



Chapter 7 The Empirical Study of the Response Lag Question

7.1 Objective

The study objective of response lag is to investigate the time lag of a significant change in the daily activity and travel pattern given a residential relocation. The trigger is the residential relocation, the termination is change of activity pattern, and the duration is the time interval between residential relocation and behavioral change, which is called response lag. More specifically, change in activity pattern is defined as the change in the time allocated to discretionary activities and their associated travel. I focus on this particular issue for three reasons. First, discretionary activities bring joy to people. Unlike activities such as mandatory (e.g., work activities) and maintenance (e.g., eating, sleeping) that can be conducted from a need to earn sufficient means and to maintain a normal life, discretionary activities are conducted mostly because people like performing them. Therefore, from a time use point of view, *ceteris paribus*: the more time spent on discretionary activities, the happier the person is. Second, discretionary activities relax one's mind and many of them require physical movement. Hence, from a health perspective, the performance of discretionary activities is likely to play a positive role in promoting health. Third, time allocation on discretionary activities is more flexible compared to mandatory and maintenance activities.

7.2 Dataset

Data Source

The data set used in the analysis is Puget Sound Transportation Panel (PSTP), a diary survey conducted in the Puget Sound region from 1989 to 2002. The entire dataset comprises 7,659 households and 16,499 individuals who are over 15 years old. In each wave, each person in the dataset was surveyed for two days. All activity and trip related information during these two days was collected, along with household and personal, socio-economic, and demographic

characteristics. The first record of each diary is the location where the individual started the day, and each following record constitutes a trip. Information on the destinations, the purpose of the trips, and the duration of the trips and activities are recorded. In this chapter, I will focus on understanding how people change their behavior in response to a residential relocation. In particular, I want to answer the question: how long does it take people to change time allocation to discretionary activities and travel?

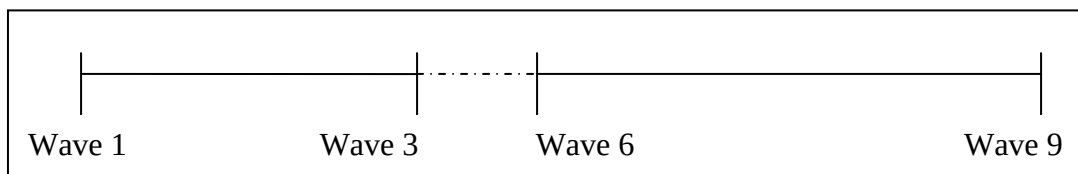
I am interested in discretionary activity, which the survey defines in slightly different ways. Two types of survey questionnaires were used in PSTP, one for Wave 1 to 7 (from 1989 to 1997) and the other for Wave 8 to 10 (from 1999 to 2002). The trip purposes used in the two surveys were slightly different. For Waves 1 to 7, the following trip purposes were recorded: work, shopping, school, visiting, free time, personal errands, appointments, and home. Among them, visiting, free time, and personal errands were classified as discretionary. For Waves 8 to 10, the following trip purposes were recorded: commutes home, commutes to work, other work-related travel, commute to school, taking child to/from school, lessons etc., visiting friends/family, errands/picking up/dropping of others, doctor appointments/other medical-related errands, deliveries, shopping, dining/coffee/etc., recreation/exercise//personal business, volunteering, going to another travel mode, other appointments or meetings, meet/transfer of carpoolers, linked commute trip to/from work, commute from work, commute from school/college. Among them, visiting friends or family, dining/coffee/etc, recreation/exercise lessons/personal business, and volunteering were classified as discretionary. The time allocated to discretionary activities and associated trips was calculated as the sum of both recorded days in each wave.

Sample Selection

For this study, we require that all must have joined the survey continuously for at least five waves. This criterion is important in our model. Since we are interested in the response lag, there must exist a common and continuous time frame during which we can track changes in the external stimuli and in the behavior. Figure 7.1 illustrates the importance of continuity. Suppose an individual first participated in the survey from Wave 1 to Wave 3. Then he dropped out of the survey between Wave 4 and Wave 5. In Wave 6, he returned and his participation continued until

Wave 9. Suppose in Wave 6, we observe a significant change in the time allocation as compared to Wave 1. Because the individual is absent from the survey in Waves 4 and 5, there exist several alternatives to calculate the response lag and categorize this particular observation. If there is no home relocation change between Waves 1 6, the response lag will be 5 years (wave 6 – wave 1 = 5 years). The observation is left-censored, since no home relocation change is observed. In another case, a home location change can take place in Wave 4 5. If that is the case, the time allocation in each of the subsequent waves (e.g., if a home relocation change takes place in Wave 4, then the subsequent waves should include Wave 5 to Wave 9) should be compared to the wave when a home relocation change takes place. If a significant change is found, then the response lag should be the difference between the two waves (e.g., if a change is found in Wave 7 as compared to Wave 4, then the response lag should be 3 years) and the observation should belong to the non-censored category (meaning that both a home location change and a time allocation change are observed during the study period). If a change is not found in any of the subsequent waves, then the response lag should be the difference between the wave when a home relocation takes place, and the last wave in which the individual participates. The observation should belong to the right-censored category (meaning that a home relocation takes place, but a time allocation change does not occur). The lack of information on the individual in Wave 4 and Wave 5 prevents us from determining which should be the correct calculation for response lag and the categorization of its group membership.

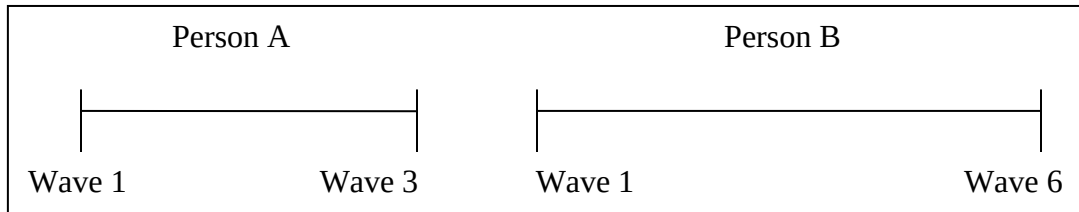
Figure 7. 1: Illustration of the Importance of Continuous Participation in Survey



Not only the study period has to be continuous as explained above, but also it has to be based on a common study period that is of the same length to every individual. Figure 7.2 shows two individuals' participation in the survey. Person A participated in three waves: wave 1 to wave 3 and Person B participated in six waves: Wave 1 to Wave 6. For Person A, we observe no significant changes in time allocation throughout three waves; for Person B, we observe a significant change in time allocation in Wave 6. In this case, it is incorrect to state that the

response lag for Person A is 2 years, while the response lag for Person B is 5 years, because the two are not based on a common study period.

Figure 7. 2: Illustration of the Importance of A Common Study Period



7.3 Variable Definition

In this research, we want to study the impact of a home location change on the response lag for two types of behavior change: a significant increase and a significant decrease in the time allocated to discretionary activities/travel. Therefore, our first task is to identify which observations experienced a significant increase and which ones had a significant decrease. The leftover observations would be the ones who maintained their time allocation to discretionary activities and associated travel.

Before we proceed with our task, we first need to determine the starting wave for each observation in the sample, against which time allocations in subsequent waves can be compared. Four-hundred seven (407) individuals have changed home locations. For these individuals, the start of study period was the wave prior to their move, and the end of study period is the fifth year after that wave. For these individuals, the wave when the home location change took place is viewed as the starting wave. For the rest of the people, for whom we did not observe a home location change, the start of study period is the first wave of survey participation, and the end of study period is the fifth year following the first wave. After the determination of the starting wave, we can calculate the differences in time allocation between the starting wave and each subsequent wave. We then input the calculated differences into the cluster analysis to determine which ones had a significant increase or decrease.

We used the cluster analysis technique with the drift option in SAS software. This is essentially performed in the following steps (SAS, 2002). In the first step, observations called “cluster seeds” are randomly selected. Then, we formed temporary clusters by assigning each observation to the cluster with the nearest seed. Each time an observation is assigned, the cluster seed is updated as the current mean of the cluster. This procedure is repeated until all observations are assigned and the change in cluster means does not exceed a pre-specified value.

Because not everyone in the sample can be observed with a home location change or a time allocation change, we have four types of censoring in the sample. Group 1 is the non-censored group, meaning that both a stimulus (a home location change) and an event (a significant increase or decrease in the time allocated to discretionary activities and associated travel) takes place within the 5-year study period⁹. Group 2 is the right-censored group, meaning that a home relocation change is observed but a significant change in the time allocated to discretionary activities and associated travel is not observed during the 5-year study period. Group 3 is the left-censored group, meaning that a significant change in the time allocated to discretionary activities and associated travel is observed, but a home relocation change is not observed during the 5-year study period. Group 4 is the left & right-censored group, meaning that neither a home relocation change nor the time allocated to discretionary activities and associated travel is observed during the 5-year study period.

According to the set of criteria explained above, the sample consists of 1,956 individuals. Among them, 407 of them changed home location and 1,549 of them didn’t change home location.

⁹ Group 1 individuals are the only group we do not need to strictly impose the 5-year study period. This is because as long as a significant change in the time allocation is observed within 5 years, we can calculate the response lag. For example, a person participates in waves 1 to 3. In wave 2, we observe a home location change and in wave 3 we observe a significant change in time allocation as compared to wave 2. Then, the response lag can be calculated as one year, even though the person’s total participation in the survey is less than 5 years.

7.4 Descriptive Analysis

Table 7-1 provides descriptive statistics of the four groups in the sample. Note that an increase or a decrease can only be distinguished for observations that are non-censored or left-censored. The average response lag for Groups 1 and 3 (non-censored and left-censored) is between two and two-and-a-half years in the increase case, and less than two years in the decrease case. Because of right-censoring, the average response lag for Groups 2 and 4 (right-censored and left & right-censored) is 5 years for both increase and decrease cases.

Descriptive statistics are provided for three types of variables: socio-economic, built environment, and temporal constraints. On average, there is no significant age difference between groups. The average age of Groups 1 and 2 is slightly lower than the average age of Groups 3 and 4. The average household size of Groups 1, 2, and 4 is around 2.7, while the average age of Group 2 is almost 3. Interestingly, subjects in Group 2 (the right-censored group) have the lowest car ownership. These are the people who have had home location changes, but not changes in the time allocation to discretionary activities. It is possible that low vehicle ownership delays changes in their time allocation. This group of individuals is also more likely to have kids under 5. In addition, Group 2 individuals are the poorest among all groups in both samples, with the highest percentage of low-income households (households with less than or equal to \$25,000 per year) and lowest percentage of high-income households (households with equal to or more than \$75,000 per year).

Within the built environment category, we provided characteristics of at three scales: 1, 2 and 5 miles. The variable “opp. at H. within 1 miles” is computed as the total area of facilities (in acres), which are suitable for discretionary activities, within 1 mile from home location.

Table 7- 1: Descriptive Statistics of the Four Groups in the Sample
(Mean Values)

Variables	Group 1		Group 2	Group 3		Group 4
	Non-censored		Right-censored	Left-censored		L&R-censored
	IN.	DE.		IN.	DE.	
Response lag	2.16	1.76	5	2.55	1.84	5
N	183	189	35	695	615	239
<i>Change in time allocation on discretionary activities (Cluster analysis result)</i>						
Average (minutes)	327	-262	-7	288	-276	-11
Standard Deviation(minutes)	208	144	47	153	162	239
<i>Household Socio-economic & Demographic Characteristics</i>						
Age	44.44	45.32	42.46	48.20	49.59	47.26
Household size	2.62	2.74	2.97	2.72	2.78	2.83
Vehicles per person	0.90	0.84	0.67	0.95	0.92	0.98
% of hhlds. who have children under age 5 (%)	19	24	31	14	13	12
% of low income ¹ families	16	14	34	17	13	12
% of high income ² families	18	22	14	21	20	15
<i>Change of Household Socio-economic & Demographic Characteristics</i>						
Change of household size	0.017	0.038	0.057	-0.059	-0.060	-0.228
Change of vehicles per person	0.007	0.042	0.073	0.003	0.023	0.003
<i>Built Environment Characteristics</i>						
Opp. at H. w. 1 mile (acres)	69.4	82.1	57.3	75.8	74.6	67.8
Opp. at H. w. 2 miles (acres)	322.2	311.6	304.5	304.5	326.2	270.7
Opp. at H. w. 5 miles (acres)	1697.3	1715.3	1667.6	1907.9	1825.2	1839.1
% of cds. w. 1 mile from H. ³	22	25	18	24	24	25
% of cds. w. 2 mile from H. ³	22	25	19	24	24	25
% of cds. w. 5 mile from H. ³	24	25	23	24	24	25
Div. within 1 mile from H.	28	31	27	31	30	31
Div. within 2 mile from H.	32	37	30	38	37	39
Div. within 5 mile from H.	33	41	33	39	39	40
<i>Change of Built Environment Characteristics</i>						
Opp. at H. w. 1 mile (acres)	8.8	-20.2	7.6	0	0	0
Opp. at H. w. 2 mile (acres)	-45.1	-20.1	18.1	0	0	0
Opp. at H. w. 5 mile (acres)	14.7	34.5	47.1	0	0	0
% of cds. w. 1 mile from H.	2.4	2.3	2.8	0	0	0
% of cds. w. 2 mile from H.	2.7	1.9	2.8	0	0	0
% of cds. w. 5 mile from H.	1.1	0.3	1.2	0	0	0
Div. within 1 mile from H.	-1.2	-3.7	2.1	0	0	0
Div. within 2 mile from H.	-0.3	-3.1	4.9	0	0	0
Div. within 5 mile from H.	1.5	-1.2	3.1	0	0	0
<i>Temporal constraints</i>						
Mandatory Time (min)	638	512	755	602	428	764
Maintenance Time (min)	101	89	46	103	105	74

Table 7- 2: Descriptive Statistics of the Four Groups in the Sample (Con't)

Variables	Group 1		Group 2	Group 3		Group 4
	Non-censored		Right-	Left-censored		L&R-
	IN.	DE.	censored	IN.	DE.	censored
Discretionary Time (min)	91	343	50	90	361	57
# of trips for mandatory act.	1.93	1.76	2.23	1.73	1.40	1.92
# of trips for maint. act.	1.61	1.33	0.97	1.52	1.75	1.12
# of trips for discret. act.	1.74	4.77	1.4	2.15	5.05	1.72
<i>Change of Temporal constraints</i>						
Mandatory Time (min)	-148	58	-120	-157	112	-76
Maintenance Time (min)	-32	-7	18	-26	-11	-21
Discretionary Time (min)	327	-262	-6	288	-275	-11
# of trips for mandatory act.	-0.41	0.10	-0.77	-0.14	0.46	0.08
# of trips for maint. act.	-0.22	0.19	0.48	-0.25	-0.17	-0.10
# of trips for discret. act.	2.15	-3.08	-0.14	2.40	-2.95	-0.43

¹ Low income: income of \$25,000 or under. ² High income: incomes of \$75,000 or above. ³ Percentage of cul-de-sacs within 1, 2, or 5 miles from home.

The same definition also applies to variable “opp. at H. within 2 miles” and “opp. at H. within 5 miles”.

For built environment characteristics, individuals in Group 2 stand out again; they have the lowest number of opportunities in radius scales of 1 and 5 miles. The variable “% of cul-de-sacs within 1 mile from H.” is calculated as the total number of cul-de-sacs divided by total number of street intersections within a 1-mile radius from home location. Our attention again turns to the Group 2 individuals. These individuals tend to live in areas with the fewest number of cul-de-sacs, an indication that many of them live in traditional cities with a grid-like street network. Diversity is computed using the following formula (Kockelman, 1997; Zhang, 2003): $div = -\sum (P_j \times \ln(P_j)) / \ln(J)$, where P_j is the proportion of jth type discretionary activity facility, J is the total number of the types of discretionary activity facilities. For this paper, J=5, includes facilities for eating, indoor activity, outdoor activity, amusements, and miscellaneous types. The Group 2 individuals tend to have lower levels of diversity compared to other groups.

We can also depict variables on temporal constraints for Group 2 individuals: working long hours, and having little time for maintenance and discretionary activities and trips. For example, Group 2 individuals spend nearly 100 minutes or more working than individuals in Group 1 and

Group 3, both of whom had experienced a significant increase in the time allocated to discretionary activities. Group 2 and Group 4 individuals made significantly fewer discretionary trips than Group 1 and Group 3 individuals, who had a significant decrease in the time allocation to discretionary activities.

7.5 Model Results

According to the response lag model developed in previous sections (see discussion in Chapter 4), we estimated a Weibull-based survival model. The derived likelihood functions were estimated in Gauss. The model results are presented in Table 7.2. In terms of the goodness of fit, the model has a 24% improvement over the constant-only model, which suggests that the performance of the model is satisfying (Greene, 2005).

Table 7- 3: Estimation Results of the Weibull Based Survival Model
(Standardized Estimates)

Variables	Response lag		Type of event	
	Coefficient	p-values	Coefficient	p-value
Intercept	0.6484	0.0112	0.1818	0.6902
Household Socio-economic & Demographic Characteristics				
Age	0.3267	0.3267	-0.1954	0.8458
Age ²	-0.2324	0.4913	0.2479	0.7993
Have children under age 5	-0.01	0.9609	0.1234	0.8441
Average vehicle ownership	-0.0300	0.6494	0.0806	0.6950
Low-income ($\leq$ \$25,000)	-0.1207	0.4774	0.3022	0.5480
High-income ($\geq$ \$75,000)	-0.2922	0.0489	-0.1302	0.7788
Change in the presence of ≤ 5 kids in the household	0.4284	0.1013	0.2227	0.7604
Change in average vehicle ownership	0.1201	0.1764	0.0559	0.7667
Built Environment Characteristics				
Diversity within 2 miles from home	0.0171	0.7763	-0.0427	0.8075
Percentage of cul-de-sac within 2 miles from home	-0.0551	0.3649	-0.2353	0.2359
Log of Opp. within 1 miles from home	-0.0158	0.6819	-0.0412	0.7114
Change in log of Opp. within 1 mile from home	-0.0097	0.9023	0.0329	0.7529
Change in diversity within 2 miles from home	0.0290	0.7105	-0.0224	0.8879
Change in percentage of cul-de-sac within 2 miles from home	-0.0544	0.6362	-0.1556	0.4130
Home/work relocation				
Indicator of change of home location	-0.3761	0.2563		
Indicator of change of work location	0.0993	0.6174		
Temporal constraints				
Discretionary activities time	-0.5604	<1E-04	-2.9973	<1E-04
Time spent on mandatory activities	0.0665	0.3800		
Time spent on maintenance activities	-0.0406	0.5248		
Change in time spent on mandatory activities			-0.7025	0.0003
Change in time spent on maintenance activities			-0.3565	0.0387
Scale (α)	0.8960	<1E-04		
Log likelihood (constant only)	-1872			
Log likelihood	-1430			

Because the model estimation is essentially a joint probability of response lag and types of exits, consequently, the independent variables include two sets: explanatory variables for response lags,

and types of events. Each set includes three kinds: socio-demographic variables, built environment variables, and temporal constraints. Each of these variables then includes both the baseline values of the variables (taken prior to the relocation) and the change values (taken as a difference between after and prior-moves). Variables in the response lag partially depict the impact on response lag; positive values would indicate a positive association between response lag and the variable, while the negative values would suggest a negative association. Variables in the type of exit partially suggest whether the exit was more likely to be an increase, a decrease, or no change. Statistically significant and positive values on the variables indicate a positive relationship between the variable and the increase exit; statistically significant negative values on the variables suggest a negative relationship between the variable and the decrease exit. Statistically insignificant variables suggest that the individual is likely to maintain the time allocation.

At 5% significance level, none of the socio-economic variables are statistically significant, except the “high income” dummy variable. The results suggest that people with higher incomes (household income equal to or greater than \$75,000) tend to have shorter response lags and are more likely to decrease the time allocated to discretionary activities and associated travel. This is not beyond our expectations. As we discussed in Section 3, the forces that income can exert can be bi-directional. While higher income provides a means to enjoy discretionary activities/travel, the temporal constraints associated with higher income people often prohibit it. Another variable that is statistically significant at the 10% level is the presence of children under the age of 5 in the household in the response lag part. For this variable, a value of 1 suggests that a household changes from having no kids to having kids; a value of 0 suggests no change; and a value of -1 suggests that a household changes from having kids to not having kids. The positive value on the variable suggests that a change from having no kids to having kids would prolong the period in which they change their time allocation.

In addition to socio-economic variables, we also examined temporal constraints. Variables of “time spent on mandatory and maintenance activities” are designed to capture the temporal constraints that an individual faces. For Waves 1 to 7, mandatory activities include those trips related to work or school; for Waves 8 to 10, mandatory activities include those trips related to

work commutes, other work-related travel, and school and college commutes. For maintenance activities, it includes shopping or appointments for Waves 1 to 7, and taking children to/from school, lessons etc., errands/picking up/dropping off others, doctor appointments/other medical-related errands, deliveries, shopping, other appointments/meetings, and meets/pick-ups/drop offs of carpoolers for waves 8 to 10. Mandatory activities are generally more spatially and temporally fixed than discretionary activities. In terms of time allocation, they are less flexible than discretionary activities. The more time spent on mandatory activities, the less time is available for discretionary activities. At the base values, none of these two variables is statistically significant at a 5% significance level. The change values of the two variables are, however, statistically significant. In the types of exits, both are negative, suggesting that an increase in the time spent on mandatory or maintenance activities would likely lead to a decrease in the time allocated to discretionary activities. Another temporal constraint is the time spent on discretionary activities. For people who already have spent a large amount of time on discretionary activities, it would take them less time to make changes, and the changes are more likely to decrease.

Another interesting set of variables is the built environment characteristics. These variables originally motivated this study. Unfortunately, none of the built environment variables we tested (base values and change values) are statistically significant in this study. This is true in the case of either response lag or type of exit. It is worthy to note that all of these built environment variables are point-based, meaning that they are calculated assuming a single reference point. For example, diversity is calculated within 2 miles from home. Apparently, these point-based measures have little impact in determining the response lag or the type of exit.

Given the findings that changes in the built environment have little impact on response lag or type of exit, we can further revise our hypothesis by saying that “maybe it is not the change in the built environment, but the simple factor of home relocation or job relocation, that triggers the change in time allocation”. Human beings are highly adaptive to their environment. They are also highly habitual in their daily routines (Chen, Kitamura and Garling, 2004). In other words, a trigger is necessary to prompt a behavior change. Due to the habitual nature of human behavior, a small trigger will probably not do, and a large trigger is necessary. A home location change

often accompanies changes in a life stages (having newborns, kids leaving home, getting married etc.). We can view such changes as large triggers, as they change one's surroundings completely. The triggering effects of home location changes may be large enough to cause changes in time allocation. Unfortunately, neither of the two variables appeared significant.

The scale parameter, α , is estimated to be less than 1, suggesting that the probability of having a significant change in time allocation decreases as time progresses. This finding is quite consistent with the current view in the literature that human beings are highly adaptive (citations). Once we adjust ourselves to new environments, it becomes increasingly harder for us to change, if no trigger is present. From the descriptive analysis (Table 1), we can see that for Groups 1 and 3 (those who have a change), the average response lag is less than 3 years. In other words, the first 3 years period following a home relocation are probably the most prudent period for a time allocation change in discretionary activities and associated travel. After that, the probability of change decreases even further.

Chapter 8 Conclusions, Policy Implications, and Future Directions

8.1 Residential Location Choice Study

Conclusions and Policy Implications

Most of the existing studies are cross-sectional in nature, speculating a contemporaneous relationship between attributes and residential location choice. In this study, I took a retrospective perspective. I hypothesized a historical deposition influence in residential location decisions. The study results support this hypothesis. As discussed in Chapter 2, the way historical deposition influences one's preferences may be bidirectional. Under an adaptive human behavior hypothesis, one may become more tolerant toward a dis-utilized attribute or more acquisitive of a beneficial attribute. Under an alternative hypothesis, one may become even more averse to an unwelcoming attribute or fatigued with a welcoming attribute such one seeks that an opposite environment. The results of my study appear to support the adaptive human behavior hypothesis. In particular, I found that people have become more tolerant to long commute distances and have viewed attributes like retail, open space, and retail opportunities more valuable after previous heavy exposure. This result, however, does not totally invalidate the alternative hypothesis and only suggests that adaptive human behavior appears to prevail.

I also found evidence of an interactive effect between historical deposition and lifecycle. During parenthood, this historical deposition influence weakens. Indirectly, this finding affirms the overriding importance of the lifecycle concept in explaining individual heterogeneity in residential location choice.

The study results imply challenges we must face if we are to reduce “wasteful travel” by investing in the urban environment and enticing people to move back to the city and be closer to their workplaces. People who have spent much time commuting probably have adjusted and accepted the long journeys as part of their lives. Furthermore, moving back to the city is often

accompanied by sacrifices on other attributes such as housing space and tranquility. Thus, a combination of a weakened disutility toward a long commute distance and a spoiled taste toward other attributes can make the back-to-the-city move particularly difficult. Does that mean that we are bound to fail in the higher-density initiative? The answer is probably no. But to achieve great success, we ought to be more sophisticated in devising the right set of policies in a local context. In devising a marketable product, we must pay attention to the target population, understand where they are coming from, and comprehend their current and future needs. This discussion naturally leads to issues on data needs. It is not sufficient to obtain only the current locations of sampled households, as most household travel surveys do. Panel datasets can be costly and require substantial technical expertise on the MPO (Metropolitan Planning Organization) staff. An alternative within the current survey format is to add questions that inquire about people's prior locations. This additional information will greatly improve our modeling capability in residential location choices. Previous research (Behrens and Del Mistro, 2006; Beige and Axhausen, 2008; Verhoeven et al, 2008) suggests that retrospective surveys, especially with regard to housing careers, are reliable.

We are still far from a full understanding of residential location choice. Currently available datasets often provide snapshots or incomplete information. In this study, the only housing attribute we can obtain from the Census data is the average housing price per room, while leaving many other housing stock variables in the error component. Even though our study is not about housing choice, but neighborhood choice, we cannot dispute the importance of the former. The current results therefore pertain to the choice of neighborhood, and can be validly interpreted in those terms as long as there is no substantial confounding of housing effects on the effects of neighborhood and accessibility attributes.

The notion of historical deposition is based on the argument that people change their preferences over time. A parallel notion is that cities evolve. Thus, to track the evolution in preferences over time, information on prior and current locations must be accompanied by land use files that track the respondents' tenure. Unfortunately, the only available time to us is year 2000. This represents a significant data limitation, and we should interpret the results with caution. Adding to our previous discussion on the data needs, MPOs should take a proactive

role in obtaining and guarding chronological land use files that record changes in the region over time.

On the other hand, I feel that the main findings from the current study would sustain if proper years were used. The Puget Sound region adopted the Growth Management Act in 1990, which set growth boundaries and limited growth to urban areas. Thus, areas originally rich in opportunities will likely maintain that status during the ten-year period. In other words, when comparing between the prior and the current location, the relative order of opportunities between the two locations will likely remain the same. Furthermore, our confidence derives from the observation that the study results are consistent with behavioral theories on adaptive human behavior (Anderson and Milson, 1989).

Gaps and Future Direction

In the current study, we assumed that each individual faces over 700 alternative zones (Census tracts). Many studies view this perfect information assumption as unreal (Simon, 1984). Studies in residential search behavior have consistently shown that households only consider a very small number of alternatives prior to the selection of a new residence (Barret, 1973; Huff, 1982; 1986). The use of real estate brokers may not only reduce the choice set, but also bias a set of alternatives considered (MacLennan and Wood, 1982). Furthermore, these administratively defined zones are not likely to match the actual alternatives in a real world scenario. Studies in environmental cognition (Timmermans, 1982; Golledge, 2003; Han et al, 2008) suggest that the limited knowledge people have about their environment is constantly modified through their encounters with various channels (e.g., friends, relatives, and media). The connection between residential search space and location choice should be a focus of our future studies.

Residential location choice is a household level decision; it is a decision reached by a negotiation process in which household members come from a completely or at least partially different past. While the current study has demonstrated the influence of the prior location that is mostly shared by all household members, the underlying negotiation process remains unknown. This negotiation process should be analyzed in a temporal and spatial context that

encompasses the pasts of individual household members. Knowledge gleaned from these types of analyses will provide important insights to residential location choice at the household level.

Finally, and typical of all studies of residential choice based on actual moves, the current study is a snapshot of a complex process in which individual households explore their residential options in a dynamically evolving housing market, under the influence of many other external factors. One may view the current study (and other location studies) a partial equilibrium where various constituent elements of a housing market are exogenous. Individual households' preference and behavioral response may well change under a dynamic environment. Therefore, one should not interpret estimated utility functions as a true utility function, expressing people's preferences in the sense that the effects of typical disequilibria in housing markets, non-availability of preferred housing in certain parts of the study area, and specific spatial distributions of attribute variables will be confounded with true underlying preferences. In light of this reservation, future work on the development of quasi-dynamic or dynamic models, involving reference points and benchmarks and reflecting spatial externalities (Werczberger, 1991), seems warranted.

8.2 Response Lag Study

Conclusions and Policy Implication

The response lag study makes methodological contributions to the field of travel behavior dynamics. Although survival models are not new in the field of transportation, it is rare to handle left-censoring and left & right-censoring. In fact, a review of the literature recovered no empirical studies in transportation that treated left-censoring and left & right-censoring. Typically in empirical studies, these two groups of data are excluded as they are "difficult to handle in hazard-based models of duration" (Hensher and Mannering, 1994, p. 65).

In addition to handling left-censoring in the sample, the observed response lag variable also involves another type of censoring, called "partial observation." Although the true response lag is a continuous variable, the observed response lag is censored at discrete intervals (1, 2, 3, 4,

or 5 years), because a panel survey is only collected at discrete time points and we do not observe the change of home location at the time it is changed. Therefore, the observed response lag variable is censored, due to partial observation. Although probing techniques and additional survey questions in future surveys may help correct this problem (Kitamura, Yamamoto and Fujii, 2003), I offer a mathematical treatment and an empirical implementation in the current study.

Both of the above-mentioned censoring types must be handled in the context of a competing risk model. This is because there involves more than one exit. A person may opt for a significant increase in time allocation, a significant decrease in time allocation, or no exit at all. I therefore develop a competing risk survival model with treatment on left-censoring and partial observation.

In addition to the methodological contributions, this study sheds new lights in activity and travel behavior by focusing on the response lag, instead of the actual behavioral change. To the author's knowledge, this is one of the first studies examining the response lag of a travel behavioral change.

While the existing studies suggest that the built environment can have a modest but significant impact on actual behavioral changes (Chen and McKnight, 2007; Chen, Gong, and Paaswell, 2008), the result of the current study indicates that the response lag is largely immune to the surrounding environment. The insignificant finding on the impact of the built environment may seem disappointing but insightful. The accumulative effect offered by Coombs, Dawes, and Tversky (1970) is not confirmed to the impact of the built environment. The individual probably rightfully perceives a change in the built environment. Indeed, prior to a home location change, one probably has studied the attributes of the new environment and compared them to the old surroundings. If a change in the built environment triggers a behavioral change, it should happen quite promptly. If there is a delay, it is not likely related to the built environment.

The combination of insignificant findings on the built environment and significant findings on other causes, such as family and social obligations, or temporal constraints, suggests that the latter contributes to the response lag of a significant change in discretionary time use. I would like to answer the earlier question of what is responsible for the response lag of a significant change in discretionary time use: the built environment, family and social obligations, temporal constraints, or a psychological delay factor. Among the hypothesized factors, family and social obligations and temporal constraints appear to play a more important role than the other two. This finding has significant policy implications. Current policies mostly concentrate on using external triggers (e.g., land use tools) to encourage people to change their behaviors, while paying insufficient attention to the various constraints that people face (e.g., family, social, or temporal constraints). The results of this current study indicate that policies can be made more effective if they can address these real constraints.

The estimation of the scale parameter, or α , turns out to be less than 1 in the model, suggesting a negative relationship between the probability of having a significant change in time allocated to discretionary activities/travel and time progression. We view this finding as a manifest of the adaptability of human beings. Human beings are highly adaptive and habitual (Chen, Kitamura, and Garling, 2004). As time progresses, human beings become used to their environments and their routines. Therefore, the chance of having a significant increase in the time allocated to discretionary activities reduces over time. The average response lag for an increase or a decrease is roughly between 1.5 and 3 years. We may view the first three years as the most prudent period for a significant change in time allocation. After that, as one becomes gradually accustomed to the environment, a significant change becomes unlikely.

Gaps and Future Direction

One issue that may arise for this study relates to residential self-selection. One may argue that people expecting a significant change in discretionary time use may choose to move to neighborhoods with resources facilitating such a change. In other words, it is the desire of wanting a time allocation change that causes the relocation choice, not the other way around. In our opinion, the problem associated with residential self-selection in our study was expected to

be minimal. Suspicion of residential self-selection arises in particular when there arises a strong association between the built environment and the activity/travel behavior. In our study, our response variable is not the change in activity/travel behavior. Rather, our response variable is the response lag between a home relocation and a significant change in time allocation. Furthermore, we found no significant association between the built environment and the response lag.

Although in most cases a trigger precedes a response, this usual temporal order may be reversed (Kitamura, 1990; Golob, 1990) in some cases. Or, an anticipatory change in time allocation may take place prior to the relocation of the household. This is an issue of response lead, instead of response lag. The current paper does not cover the issue of response lead. Conceptually, we feel that it is unlikely for a person to change his/her time allocation prior to home location change, if indeed the home location change is the cause and the change in time allocation is the effect. This is because one's time allocation is highly constrained by one's temporal and spatial environments (Hagerstrand, 1970).

The current study has limitations. An immediate one relates to the dataset in use. The Puget Sound Panel Dataset only collects two weekdays' time use information in each wave. This two-day period may not be enough for discretionary time use. Weekly or even monthly data would be more desirable. The consequence of this two-day period might lead to misclassification of a significant change in discretionary time use as well as biased model results.

Another limitation relates to the independence assumption between the response lag and the type of exit. It is possible that the two are related, meaning that an increase in time allocation may be more likely related with longer response lag, while a decrease may be associated with shorter response lag. Conceptually, this can be handled by a joint distribution between the two, and then approximating the joint distribution by Monte Carlo simulations (Bhat, 1996). This is not handled in the current paper, due to the complexity of the forms derived from left-censoring and partial observation.

One future direction relates to the relationship between response lag and the magnitude of change. One may argue that these two variables might be simultaneously determined. We counter-argue that conceptually, synergy between the two is unlikely, because there exists an inherently sequential order between the two. When a significant change in time allocation takes place, the response lag is automatically determined. Conversely, when one decides on the response lag, it will result in a significant change in the time allocation. A more meaningful question appears to be which comes first? In other words, do people decide on the response lag first, then on the change in time allocation, or vice versa? This question is beyond the scope of the current study, but certainly worthy of future pursuit.

References

- Aarts, H.; Verplanken, B. and van Knippenberg, A. (1998) Predicting Behavior From Actions in the Past: Repeated Decision Making or a Matter of Habit? *Journal of Applied Social Psychology* 28(15) , 1355–1374.
- Abraham, E. and Hunt, D. (1997) Specification and Estimation of a Nested Logit Model of Home, Workplace and Commuter Mode Choice by Multiple Worker Households, *Transportation Research Record* 1606, 17-24.
- Alonso, W. (1964) *Location and Land Use*, Harvard University Press, Cambridge.
- Anderson, J. R. and Milson, R. (1989) Human Memory: An Adaptive Perspective. *Psychological Review* 96, 703-719.
- Anderson, J. R. (1990) *The Adaptive Character of Thought*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Amemiya, T. (1999) A Note on Left Censoring. In *Analysis of Panels and Limited Dependent Variable Models*. Hsiao, C. Pesaran, M.H., Lahiri, K., Lee, L.-F.(eds.). Cambridge University Press, 7-22.
- Assadian, A. and Ondrich J.(1993) Residential Location, Housing Demand and Labour Supply Decisions of One- and Two- Earner Household: The Case of Bogota, Colombia. *Urban Studies* 30(1), 73-86.
- Banister, D. (1978) The Influence of Habit Formation on Modal Choice – A Heuristic Model. *Transportation* 7, 19-33.
- Barrow, L. (2002) School Choice Through Relocation: Evidence From the Washington, D.C. area. *Journal of Public Economics*, 86(2), 155-189.
- Bekhor, S. and Prashker J. N. (2008) GEV-based Destination Choice Models That Account For Unobserved Similarities Among Alternatives. *Transportation Research Part B* 42, 243-262.
- Bell, W. L. (1968) *The City, Suburb and A Theory of Social Choice*. *The New Urbanization*, edited by S. Greer, St. Martin's Press, New York, 132-168.
- Böheim, R. and Taylor, M. (1999) Residential Mobility, Housing Tenure and the Labor Market in Britain. Institute for Social and Economic Research and Institute for Labor Research, University of Essex, UK.

Ben-Akiva, M. and Lerman, R. (1979) Disaggregate Travel and Mobility-choice. Models and Measures of Accessibility. in D. A. Hensher and P. R. Stopher. (eds.) Behavioural Travel Modelling, London: Croom-Helm, 654-679.

Ben-Akiva, M. and Palma de. A. (1986) Analysis of A Dynamic Residential Location Choice Model with Transaction Costs, *Journal of Regional Science* 26(2), 321-341.

Ben-Akiva M.; Bowman J. (1998) Activity Based Travel Demand Model Systems. In: Marcotte, P.; Nguyen, S.; eds. Boston/Dordrecht. Equilibrium and Advanced Transportation Modeling, Kluwer Academic Publishers.

Baht, R. C (1996) A Generalized Multiple Durations Proportional Hazard Model With an Application to Activity Behavior During The Evening Work-to-home Commute. *Transportation Research Part B* 30(6), 465-480.

Bhat, C. (2001) Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. *Transportation Research Part B* 35, 677-693.

Bhat, C. and Guo J. (2004) A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling. *Transportation Research Part B* 38, 147-168.

Bina, M.; Kockleman, K. and Suescun, D. (2005) Location Choice vis-à-vis Transportation: The Case of Recent Homebuyers, manuscript, http://www.ce.utexas.edu/prof/kockelman/public_html/TRB06HomeChoice.pdf.

Brown, L.; Fuguitt, V.; Heaton, B. and Waseem S. (1997) Continuities in Size of Place Preferences in the United States, 1972-1992. *Rural Sociology* 62, 408-428.

Burns, L. (1979) Transportation, Temporal, and Spatial Components of Accessibility. Lexington, MA: Lexington Books.

Buhaug, H. (2005). Dangerous Dyads Revisited: Democracies May Not Be That Peaceful after All. *Conflict Management and Peace Science* 22(2), 95-111.

Camerer, C. (1998) Bounded Rationality in Individual Decision Making. *Experimental Economics* 1, 163-183.

Cao, X.; Mokhtarian, P. and Handy S. (2006) Examining The Impacts of Residential Self-selection On travel Behavior: Methodologies and Empirical Findings. Institute of Transportation Studies, University of California, Davis, Research Report UCD-ITS-RR-06-18.

Chen, C.; Garling, T.; and Kitamura, R. (2004). Activity Rescheduling: Deliberate or Habitual? *Transportation Research Part F* 7(6), 351-371.

Chen, C.; Gong, H.; and Paaswell, R. (2008) Role of the Built Environment on Mode Choice Decisions: Additional Evidence on the Impact of Density. *Transportation* 35(3), 285-299.

Chen, C. and Lin, J. (2006) Making an Informed Vehicle Scrappage Decision. *Transport Review* 26(6), 731-748.

Chen, C. and Lin, H. (2008) Dialogues with Homeowners and Renters in the New York Metropolitan Region. Unpublished working paper.

Chen, C. and McKnight, C.E. (2007) Does the Built Environment Make a Difference? Additional Evidence From the Daily Activity and Travel Behavior of Homemakers Living in New York City and Suburbs, *Journal of Transportation Geography*.

Chen, C. and Mokhtarian, P. (2006) Tradeoffs between Time Allocations to Maintenance Activities/Travel and Discretionary Activities/Travel. *Transportation* 33, 223-240.

Chen, C. and Niemeier, D. (2005) A Mass Point Vehicle Scrappage Model. *Transportation Research Part B* 39(5), 401-415.

Chen, C.; Li, W. and Kwan, M. (2006) Point-based Accessibility and Individual-based Accessibility. Presented at IATBR, Kyoto, Japan, August 2006.

Clarke M., Dix M., and Goodwin P. (1982) Some Issues of Dynamics in Forecasting Travel Behavior – A discussion Paper. *Transportation* 11(2), 153-172

Clark, W. and Burt, J. (1980) The Impact of Workplace on Residential Relocation. *Annals of the Association of American Geographers* 70(1), 59-67.

Clark, W. A. and van Lierop, W. F. (1986) Residential Mobility and Household Location Modeling. *Handbook of Regional and Urban Economics*, Volume 1, Chapter 3, edited by P. Nijkamp. Elsevier Science Publisher BV.

Clark, W. and Onaka, J. L. (1983). Life Cycle and Housing Adjustment as Explanations of Residential Mobility, *Urban Studies* 20(1), 47 – 57.

Clark, W.; Huang, Y. and Withers, S. (2003) Does Commuting Distance Matter? Commuting Tolerance and Residential Change. *Regional Science and Urban Economics* 33, 199-221.

Clark, W.; Deurloo, C. and Dieleman, M. (2006) Residential Mobility and Neighborhood Outcomes, *Housing Studies* 21(3), 323-342.

Clark, W. and Ware, J. (1997) Trends in Residential Integration by Socioeconomic Status in Southern California. *Urban Affairs Review* 32(6), 825-843.

Clark, W. and Withers, S. (1999) Changing Jobs and Changing Houses: Mobility Outcomes of Employment Transitions. *Journal of Regional Science* 39(4), 653-673.

- Clarke M., Dix M., and Goodwin P. (1982) Some Issues of Dynamics in Forecasting Travel Behavior – A discussion Paper. *Transportation* 11(2), 153-172.
- Crane, R. and Crepeau, R. (1998) Does Neighborhood Design Influence Travel? A Behavioral Analysis of Travel Diary and GIS data. *Transportation Research Part D – Transport and Environment* 3(4), 225-238.
- Cervero, R. and Gorham, R. (1995) Commuting in transit versus automobile neighborhoods. *Journal of the American Planning Association* 61(2), 210-226.
- Coombs, H.; Dawes, M. and Tversky, A. (1970), *Mathematical Psychology: An Elementary Introduction*, Prentice-Hall, Englewood Cliffs, NJ.
- Colwell, F.; Dehring, A. and Turnbull, K. (2002) Recreation demand and residential location. *The Journal of Urban Economics* 51, 418-428.
- Coulton, C.; Korbin, J.; Chan, T. and Su, M. (2001) Mapping Residents' Perceptions of Neighborhood Boundaries: a Methodological Note. *American Journal of Community Psychology* 29(2), 371-383.
- Curtis, C. (1995) Reducing the need to travel: strategic housing location and travel behaviour. In: Earp, J.H.; Headicar, P.; Banister, D. and Curtis, C. *Reducing the need to travel: some thoughts on PPG13*. *Oxford Planning Monographs* 1(2), 29-47.
- Dalvi, Q. and Martin, M. (1976) The Measurement of Accessibility: Some Preliminary Results. *Transportation* 5, 17-42.
- Davis, J. (1993) The Commuting of Exurban Home Buyers. *Urban Geography* 14(1), 7-29.
- DePalma, A. and Ben-Akiva, M. (1981) Interactive Dynamic Modeling of Residential Location Choice. Umea. Sweden. Paper presented at the international conference on structural and economic analysis and planning in time and space.
- Desalvo, J. (1977) Urban Household Behavior in a Model of Completely Centralized Employment. *Journal of Urban Economics* 4, 1-14.
- Deng, Y.; Ross, L. and Wachter M. (2000) manuscript, available online http://www.usc.edu/schools/sppd/lusk/research/pdf/wp_1999_108.pdf.
- Devisch O.; Arentze T.; Borgers A. and Timmermans H. (2005) An Agent-based Model of Residential Choice Dynamics In Non-Stationary Housing Markets, In: Proceedings of the Computers in Urban Planning and Urban Management Conference, London.
- Dieleman, F.; Clark, W. A. and Deurloo, M. C. (2000) The Geography of Residential Turnover in Twenty-seven Large US Metropolitan Housing Markets, 1985-95. *Urban Studies* 37(2), 223-245.

Doohee, N., and Fred M. (2000) An Exploratory Hazard-Based Analysis of Highway Incident Duration. *Transportation Research: Part A: Policy and Practice* 34(2), 85-102.

Downs, A. 1981. *Neighborhoods and Urban Development*. Washington D.C., Brookings Institution.

Dunphy, R.T. and Fisher, K. (1996) Transportation, congestion, and density: new insights. *Transportation Research Record* 1552, 89-96.

ÆRØ, T. (2006) Residential Choice from a Lifestyle Perspective. *Housing, Theory and Society* 23(2), 109-130.

Elder, G. and Shanahan, M. (2006) The Life Course and Human Development. *Handbook of Child Psychology*, Volume 1, 6th Edition: Theoretical Models of Human Development. Edited by William Damon and Richard M. Lerner. Wiley and Stone, New York, 665-715.

Ellen, I. and Hempstead, K. (2002) Telecommuting and the demand for urban living: a preliminary look at white-collar workers. *Urban Studies* 39(4), 749-66.

Ewing, R.; DeAnna, M. and Li, S-C. (1996) Land use impacts on trip generation rates. *Transportation Research Record* 1518, 1-6.

Farley, J. (1995) Race still Matters: The Minimal Role of Income and Housing Cost as Causes of Housing Segregation in St. Louis, 1990. *Urban Affairs Review* 31(2), 244-254.

Fasolo, B.; Misuraca, R.; and McClelland, G. H. (2003) Individual Differences in Adaptive Choice Strategies. *Research in Economics* 57(3), 219-233.

Freda C.; Sudipto B.; Bradley P. C. and Debajyoti S. (2007). Flexible Cure Rate Modeling under Latent Activation Schemes. *Journal of the American Statistical Association* 102(478), 560-72.

Fernandez L.; Brown D.; Marans R. and Nassauer J. (2005) Characterizing Location Preferences in an Exurban Population: Implications For Agent-based Modeling, *Environment and Planning B: Planning and design* 32, 799-820.

Freedman, O. and Kern, R. (1997) A model of workplace and residence choice in twoworker households. *Regional Science and Urban Economics* 27, 241-260.

Friedman, B.; Gordon, S. and Peers, J. (1994) Effect of neotraditional neighborhood design on travel characteristics. *Transportation Research Record* 1466, 63-70.

Galilea, P. and Juan de Dios Ortúza. (2005) Valuing Noise Level Reductions in a Residential Location Context, *Transportation Research Part D: Transport and Environment* 10(4), 305-322.

- Gabriel A. and Rosenthal S. (1989) Household Location and Race: Estimates of a Multinomial Logit Model. *Review of Economics and Statistics* 71(2), 240-249.
- Galster, C. (2001) On the Nature of Neighborhood, *Urban Studies* 38(12), 2111-2124.
- Genus K. (2006) Accessibility, Land Use and Transport, doctoral dissertation.
- Geoghegan, J. (2002) The Value of Open Spaces in Residential Land Use. *Land Use Policy* 19(1), 91-98.
- Gilbert, C. A Duration Model of Automobile Ownership. *Transportation Research Part B* 26(2), 222-242.
- Glasgow, N. (2000) Rural/urban patterns of aging and caregiving in the United States. *Journal of Family Issues* 21(5), 611-631.
- Golob, T. F. (1990) The Dynamics of Household Travel Time Expenditures and Car Ownership Decisions. The University of California Transportation Center, June.
- Golledge, R. (2003) Reflections on recent cognitive behavioral research with an emphasis on research in the United States of America. *Journal of the Institute of Australian Geographers* 41, 117-130.
- Goodwin P.B. (1989). Family Changes and Public Transport Use 1984-1987: a Dynamic Analysis Using Panel Data. *Transportation* 16, 121-154.
- Green, A. E. (1997) A Question of Compromise? Case Study Evidence on the Location and Mobility Strategies of Dual Career Households. *Regional Studies* 31(7), 641-657.
- Greene, W. H. (2003) Econometric Analysis. Prentice Hall. Upper Saddle River, New Jersey.
- Giuliano, G. and Small, K. (1993) Is the journey to work explained by urban structure? *Urban Studies* 30(9), 1485-1500.
- Goulias, K.G. (1999) Longitudinal analysis of activity and travel pattern dynamics using generalized mixed Markov latent class models. *Transportation Research B* 33(8), 535-558.
- Guo, J. and Bhat, C. (2007) Operationalizing the Concept of Neighborhood: Application to Residential Location Choice Analysis. *Journal of Transport Geography* 15(1), 31-45.
- Guo, J. and Chen, C. (2007) The Built Environment and Travel Behavior: Making the Connection. September, *Transportation* 5, 529-533.
- Hachen, S. D. (1988) The Competing Risks Model. A Method for Analyzing Processes with Multiple Types of Events. *Sociological Methods & Research* 17(1), 21-54.

- Hallman, W. (1984) *Neighborhoods: Their Place in Urban Life. Sage Library of Social Research.* 154.
- Hammond, D. (2005) Residential Location and Commute Mode Choice. Dissertation submitted in fulfillment of the MSC in Transport and Planning, University of Wales, Cardiff.
- Han, A. and Hausman, J. A. (1990) Flexible Parametric Estimation of Duration and Competing Risk Models. *Journal of Applied Econometrics* 5, 1-28.
- Hanaki, N. (2005) Individual and Social Learning. *Computational Economics* 26(3-4), 213-232.
- Handy S. and Niemeier D. (1997) Measuring accessibility: an exploration of issues and alternatives. *Environment and Planning A* 29, 1175-1194.
- Hansen, G. (1959) How Accessibility Shapes Land Use. *Journal of the American Institute of Planners* 25, 73-76.
- Hanson, S. and Pratt, G. (1995) *Gender, Work and Space*, New York: Routledge.
- Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, 24, 7-21.
- Hajivassiliou, V., McFadden, D., and Ruud, P. (1996). Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results, *Journal of Econometrics* 72, 85-134.
- Hensher, D. and Johnson, L. (1980) Application of Multinomial Probit to a Time Series of Cross-sections Data Set. Conference on Disaggregate Travel Demand Models, Leeds University.
- Hensher, D. and Mannering, F. (1994) Hazard-based Duration Models and their Application to Transport Analysis. *Transport Reviews* 14(1), 63-82.
- Humpel, N.; Owen, N.; Iverson, D.; Leslie, E. and Bauman, A. (2004) Perceived environment attributes, residential location, and walking for particular purposes. *American journal of preventive medicine* 26(2), 119-125.
- Hunt, J.D.; McMillan, J.D.P and Abraham, J.E. (1994) A stated preference investigation of influences on the attractiveness of residential locations, *Transportation Research Record* 1466, 79-87.
- Hwang E.; McGarrell E. F.; and Benson B. L. (2005) Public satisfaction of the South Korean Police: The Effect of Residential Location in A Rapidly Industrializing Nation, *Journal of Criminal Justice*, 33(6), 585-599.

- Iamond, P. and Hausman, J. (1984) The Retirement and Unemployment Behavior of Older Men. Retirement and Economic Behavior. Aaron H and Burtless G., Eds., Brookings Institute, Washington DC. 97-134.
- Ingram R. (1971) The Concept of Accessibility: A Search for an Operational Form. *Regional Studies* 5, 101-107.
- Iragael, J. (2006) Stability or Regularity of the Daily Travel Time in Lyon? Application of a Duration Model. *International Journal of Transport Economics* 33(3), 369-400.
- Jara-Díaz, R. and Martínez F. (1999) On the specification of indirect utility and willingness to pay for discrete residential location models. *Journal of Regional Science* 39, 675-688.
- Kahneman, D. and Tversky A. (1979) Prospect Theory: An Analysis of decision Under Risk. *Econometrica* 47(2), 263-292.
- Keller, S. (1968) The Urban Neighbourhood: A Sociological Perspective. New York, Random House.
- Kendig, H. (1984) Housing Careers, Life Cycle, and Residential Mobility: Implications for the Housing Market. *Urban Studies* 21(3), 271-283.
- Kertzner, D. I. and Keith, J. (1984). *Age and Anthropological Theory*. Ithaca, NY: Cornell University Press.
- Kiefer, N. (1988) Economic duration data and hazard functions. *Journal of Economic Literature* 26, 646-679.
- Kim, S.; Peter, O. and Daniel. O. (2001) The Effects of Housing Prices, Wages, and Commuting Time on Joint Residential and Job Location Choices. *American Journal of Agricultural Economics* 83, 1036-1048.
- Kim, J.; Pagliara F. and Preston J. (2005) The Intention to Move and Residential Location Choice Behaviour, *Urban Studies* 42(9), 1621-1636.
- Kim, M. and Kwan, M (2003) Space-time Accessibility Measures: A Geocomputational Algorithm With a Focus on the Feasible Opportunity Set and Possible Activity Duration. *Journal of Geographical Systems* 5, 71 -91.
- Kim, T.; Horner, M. and Marans, R. W. (2005) Life Cycle and Environmental Factors in Selecting Residential and Job Locations. *Housing Studies* 20(3), 457-473.
- Kitamura R. (1989) A Causal Analysis of Car Ownership and Transit Use. *Transportation* 16, 155-173.

- Kitamura, R. (1990) Panel Analysis in Transportation Planning: an Overview. *Transportation Research Part A* 24(6), 401-415.
- Kitamura, R. (2002) Time Use: Workshop Report. In *In Perpetual Motion – Travel Behavior Research Opportunities and Application Challenges*. Mahmassani, H. S. (Ed.). Elsevier Science, Limited: Oxford, UK, 333-338.
- Kitamura, R.; Yamamoto, T. and Fujii, S. (2003) The Effectiveness of Panels in Detecting Changes in Discrete Travel Behavior. *Transportation Research Part B*, 191-2006.
- Kitamura, R.; Mokhtarian, P. and Laidet, L. (1997) A Micro-analysis of Land Use and Travel in Five Neighbourhoods in the San Francisco Bay Area. *Transportation* 24, 125-158.
- Kockelman, K. (1997) Travel Behavior as a Function of Accessibility, Land Use Mixing and Land Use Balance: Evidence from the San Francisco Bay Area. *Transportation Research Board* 1607, 116-125.
- Krizek, J. (2003) Residential Relocation and Changes in Urban Travel: Does Neighborhood-scale Urban Form Matter? *Journal of the American Planning Association* 69(3), 265-281.
- Krizek, J. and Waddell, P. (2002) Analysis of Lifestyle Choices: Neighborhood Type, Travel Patterns, and Activity Participation, *Transportation Research Record* 1807, 119-128.
- Krout, A.; Holmes, H.; Erickson, A. and Wolle, S. (2003) Residential relocation. In J. A. Krout & E. Wethington (Eds.). *Residential choices and experiences of older adults. Pathways to life quality*. New York: Springer, 27-48.
- Kwan, M. and Hong, X. (1998) Network-based Constraints-oriented Choice Set Formation Using GIS. *Geographical Systems* 5, 139-162.
- Kwan, M. (1998) Space-time and Integral Measures of Individual Accessibility: a Comparative Analysis Using a Point-based Framework. *Geographical Analysis* 30(3), 191-217.
- Kwan, M. (1999) Gender and Individual Access to Urban Opportunities: A Study Using Space-time Measures. *Professional Geographer* 51(2), 210-227.
- Kwan, M. and Weber, J. (2003) Individual accessibility revisited: implications for geographical analysis in the twenty-first century. *Geographical Analysis* 35, 341 – 353.
- Lancaster, S. R. (1966) A New Approach to Consumer Theory. *Journal of Political Economy* 74, 132-157.
- Lerman, P. (1975) Community treatment and social control. Chicago: University of Chicago Press.

Lerman, S. R. (1977) Location, Housing, Automobile Ownership and Mode to Work, A Joint Choice Model. *Transportation Research Record* 610, 6-11.

Lenntorp, B. (1976) Paths in Time-Space Environments: A Time Geographic Study of Movement Possibilities of Individuals. Lund Studies in Geography B: Human Geography, Lund: Gleerup.

Linneman, P. (1982) Hedonic Prices and Residential Location. *The Economics of Urban Amenities*, edited by D. Diamond and G. Tolley. Academic Press, New York.

Li, Y. and Lin X. (2006) Semiparametric Normal Transformation Models for Spatially Correlated Survival Data. *Journal of the American Statistical Association* 101(474), 591-603.

Lin, J.; Chen, C. and Niemeier, D. (2008) An Analysis on Long Term Emission Benefits of a Government Vehicle Fleet Replacement Plan in Northern Illinois. *Transportation* 35, 219-235.

Louviere J. and Timmermans H. (1990) Hierarchical Information Integration Applied to Residential Choice Behavior. *Geographical Analysis* 22,127-145.

Luce D. (1959) Individual Choice Behavior, John Wiley and Sons, New York.

Ma, J. and Goulias, K. G. (1997) Multivariate Marginal Frequency Analysis of Activity and Travel Patterns in First Four Waves of Puget Sound. *Transportation Research Record* 1556, 67-76.

Ma, J. and Goulias, K. G. (1997) A Dynamic Analysis of Person and Household Activity and Travel Patterns Using Data From First Two Waves in the Puget Sound Transportation Panel. *Transportation* 24, 309-331.

Madden, J. F. (1981) Why Women Work Closer to Home? *Urban Studies* 18, 181-194.

Marschak, J. (1960) Binary Choice Constraints on Random Utility Indications, Stanford Symposium on Mathematical Methods in the Social Sciences, Stanford University Press, Stanford, CA, 312-329.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior, *Frontiers in Econometrics*, Academic Press, New York, 105-142.

McFadden, D. (1978) Modeling the Choice of Residential Location, Spatial Interaction Theory and Planning Models, edited by A. Karlqvist et al, North Holland Publishers, Amsterdam.

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

McGinnis, R. (1968). A Stochastic Model of Social Mobility. *American Sociological Review* 23, 712-722.

- Meloni, I.; Spissu, E. and Bez, M. (2007), A Model of the Dynamic Process of Time Allocation to Discretionary Activities. *Transportation Science*, Vol. 41, Issue 1, 15-28.
- Miller, H. (1991) Modelling Accessibility Using Space-time Prism Concepts Within Geographic Information Systems. *International Journal of Geographical Information Systems* 5(3), 287-301.
- Mokhtarian, P. and Cao, X. (2008) Examining the Impacts of Residential Self-selection on Travel Behavior: A Focus on Methodologies. *Transportation Research Part B* 43 (3), 204-228.
- Molin, E.; Oppewal, H. and Timmermans, H. (1999) Group-based Versus Individual-based Conjoint Preference Models of Residential Preferences: a Comparative test, *Environment and Planning A* 31, 1935-1947.
- Molin, E. and Timmermans H. (2003) Testing Hierarchical Information Integration Theory: the Causal Structure of Household Residential Satisfaction. *Environment and Planning A* 35(1), 43 – 58.
- Morris, D. and K. Hess (1975). *Neighborhood Power: The New Localism*. Boston, Beacon Press.
- Muth, R.F. (1969) *Cities and Housing*, University of Chicago Press, Chicago.
- Nechyba, J. and Strauss P. (1998) Community choice and local public services: a discrete choice approach. *Regional Science and Urban Economics* 28, 51-73.
- Nilles, J. (1976) *The Telecommunication Transportation Trade Off*, Wiley Interscience Pub, New York.
- Næss, P.; Roe, P.G. and Larsen, S. (1995) Travelling distances, modal split and transportation energy in thirty residential areas in Oslo. *Journal of Environmental Planning and Management* 38(3), 349-370.
- Omariba, D. Walter R; Beaujot, R; and Rajulton, F (2007). Determinants of Infant and Child Mortality in Kenya: An Analysis Controlling for Frailty Effects. *Population Research and Policy Review* 26(3), 299-321.
- Onaka, J. and Clark, W. A. (1983) A Disaggregate Model of Residential Mobility and Housing Choice. *Geographical Analysis* 15, 287-304.
- Oswald, F. and Rowles, D. (2006) Beyond the relocation trauma in old age: New trends in today's elders' residential decisions. In H.-W. Wahl, C. Tesch-Römer, & A. Hoff (Eds.). *New Dynamics in Old Age: Environmental and Societal Perspectives*. Amityville, New York, 127-152.
- O'Sullivan, A. (2000) *Urban Economics*, 4th edn. Boston, MA: Irwin McGraw-Hill.

Parkes, A.; Kearns, A.; and Atkinson, R. (2002) What Makes People Dissatisfied with their Neighborhood? *Urban Studies* 39, 2413-2438.

Parks, S. E., and Housemann R. A. (2003) Differential correlates of physical activity in urban and rural adults of various socioeconomic backgrounds in the United States. *Journal of Epidemiology and Community Health* 57, 29–35.

Pas, E. I. (1985). State of the Art and Research Opportunities in Travel Demand: Another Perspective. *Transportation Research A* 19(5/6) 460-464.

Pas, E. I. (1996). Recent Advances in Activity-Based Travel Demand Modeling. In *Activity-Based Travel Forecasting Conference, June 2-5, 1996 Summary, Recommendations, and Compendium of Papers*. Texas Transportation Institute (Eds.). Washington, D.C.: Travel Model Improvement Program, U.S. Department of Transportation and U.S. Environmental Protection Agency, 79-102.

Pendyala, R. M., and Goulias, K.G. (2002). Time Use and Activity Perspectives in Travel Behavior Research. *Transportation* 29(1) 1-4.

Pendyala R. (2003) Time Use and Travel Behavior in Space and Time (Chapter 2). In *Transportation Systems Planning: Methods and Applications*. Goulias, K.G. (Ed.). CRC Press, Boca Raton, FL, 2-1 to 2-37.

Perez, P.; Martinez, F. J. and de Dios Ortuzar, J. (2003) Microeconomic Formulation and Estimation of a Residential Location Choice Model: Implications for the Value of Time. *Journal of Regional Science* 43(4), 771-789.

Quigley, J. M. (1973) *Housing Demand in the Short Run: Analysis of Polytomous Choice*. Econometric Society, New York, unpublished paper.

Quigley, J. M. (1976) Housing Demand in the Short Run: An Analysis of Polytomous Choice. *Explorations in Economic Research* 3, 76-102.

Quigley, J. M. (1985) Consumer Choice of Dwelling, Neighborhood, and Public Services. *Regional Science and Urban Economics* 15, 41–63.

Rouwendal, J. and Meijer, E. (2001) Preferences for Housing, Jobs and Commuting: A Mixed Logit Analysis. *Journal of Regional Science* 41, 475 - 505.

Rouwendal, J and Nijkamp, P. (2004) Living in Two Worlds: A Review of Home-to-Work Decisions. *Growth & Change* 35(3), 287-303.

Rutten, A. and Abel T. (2001) Self reported physical activity, public health, and perceived environment: results from a comparative European study. *Journal of Epidemiology and Community Health* 55, 139–146.

Sam, A. and Ronald, S. W. (2003). The Duration of Homelessness: Evidence from a National Survey. *Journal of Housing Economics* 12(4), 273-90.

Sander, W. (2004) Education and urban household location. *Economics Letters* 82(3), 327-331.

Segal, D. (1979) A Quasi-log Linear Model of Neighborhood Choice. *The Economics of Neighborhood*, edited by David Segal, Academic Press, New York.

Sermons, W. and Koppelman, S. (2001) Representing the differences between female and male commute behavior in residential location choice models, *Journal of Transport Geography* 9, 101-110.

Simon, H. (1984) *Models of Bounded Rationality*. MIT Press, Cambridge.

Simpson, W. (1987) Workplace Location, Residential Location, and Urban Commuting. *Urban Studies* 24(2), 119-128.

So, K.; Orazem, P. and Otto, D. (2001) The Effect of Housing Prices, Wages, and Commuting Time on Joint Residential and Job Location Choices. *American Journal of Agricultural Economics* 83(4), 1036-1048.

Srour, M.; Kockelman, K. and Dunn, T.(2002) Accessibility Indices: Connection to Residential Land Prices and Location Choices, *Transportation Research Record*, 1805, 25-34.

Stead, D. (2001) Relationships between land use, socio-economic factors and travel patterns in Britain. *Environment and Planning B* 28(4), 499-528.

Stefano, M. (2003). Urban Environment and Travel Times to Work. *International Journal of Transport Economics* 30(2), 187-203.

Tammie, X. and Malpezzi, S. (2006). Household Mobility in New York City's Regulated Rental Housing Market Simmons-Mosley. *Journal of Housing Economics* 15(1), 38-62.

Thurstone, L. (1927). A Law of Comparative Judgement, *Psychological Review* 34, 273-286.

Timmermans, H.; Noortwijk, van L.; Oppewal, H.; and Waerden, van der P.(1996) Modeling Constrained Choice Behaviour in Regulated Housing Markets by Means of Discrete Choice Experiments and universal Logit Models: an Application to Residential Choice Behaviour of Divorcees, *Environment and Planning A* 28, 1095-1112.

Timmermans, H.J.P.; Heijden RECM van der and Westerveld H. (1982) Perception of Urban Retailing Environments: an Empirical Analysis of Consumer Information and Usage Fields. *Geoforum* 13, 27-37.

Tobler, W. (1970) A Computer Model Simulating Urban Growth in the Detroit Region, *Economic Geography* 46(2), 234-240.

Toussaint-Comeau, M. and Rhine. S. (2004) Tenure Choice with Location Selection: The Case of Hispanic Neighborhoods in Chicago. *Contemporary Economic Policy* 22(1), 95–110.

Train, K. (2003) Discrete Choice Methods with Simulation, Cambridge University Press.

Tu, Y. and Goldfinch, J. (1996) A Two-Stage Housing Choice Forecasting Model. *Urban Studies* 33(3), 517-37.

Turner, T. and Niemeier D. (1997) Travel to Work and Household Responsibility: New Evidence. *Transportation* 24, 397-419.

Vickerman, W. (1974) Accessibility, Attraction, and Potential: A Review of Some Concepts and Their Use in Determining Mobility. *Environment and Planning A* 6, 675-691.

Venti, S. and Wise, D. (1989) Aging, Moving, and Housing Wealth. NBER Working Papers 2324, National Bureau of Economic Research, Inc.

Wachs, M.; Taylor, B.; Levine, N. and Ong P. (1993) The Changing Commute: a Case-study of Jobs-housing Relationship Over Time. *Urban Studies* 21, 15-29.

Waddell P.(1996) Accessibility and Residential Location: The Interaction of workplace, Residential Mobility, Tenure, and Location Choices, Lincoln Land Institute TRED Conference paper.

Warren, D. (1981) Helping Networks: How People Cope with Problems in the Urban Community. Notre Dame, IN, University of Notre Dame Press.

Weiner E., (1997). Urban Transportation Planning In the United States: An Historical Overview, available online <http://tmip.fhwa.dot.gov/clearinghouse/docs/utp/>.

White, M. J. (1977) A Model of Residential Location Choice and Commuting by Men and Women Workers. *Journal of Regional Science* 17, 41-52.

Wilcox, S., C. and Castro et al. (2000) Determinants of leisure time physical activity in rural compared with urban older and ethnically diverse women in the United States. *Journal of Epidemiology and Community Health* 54, 667–672.

Williams, R. C. (1979) A Logit Model of Demand for Neighborhood. The Economics of Neighborhood, edited by David Segal. Academic Press, New York, 17-42.

Office of Washington Superintendent of Public Instruction, *Washington Assessment of Student Learning* 2002. <http://www.k12.wa.us/assessment/WASL/default.aspx>, Accessed March 15, 2007.

Wilson, G. (1971) A family of spatial interaction models, and associated developments. *Environment and Planning A* 3, 1-32.

Xie, Y. and Giles, E. D (2007). A Survival Analysis of the Approval of U.S. Patent Applications Department of Economics, University of Victoria, Econometrics, Working Papers.

Yamamoto, T.; Madre, J. and Kitamura, R. (2004) An Analysis of the Effects of French vehicle Inspection Program and Grant for Scrappage in Household Vehicle Transaction. *Transportation Research Part B* 38(10), 905-926.

Zhang, J.; Timmermans, H.J.P.; Borgers, A.W.J. and Wang, D. (2004), Modeling Traveler Choice Behavior Using The Concepts Of Relative Utility And Relative Interest. *Transportation Research B* 38, 215-234.

Zhang, M. (2004) The Role of Land Use in Travel Mode Choice. *Journal of the American Planning Association* 70(3), Summer, 344-360.