The Dissertation Committee for Abdul Rawoof Pinjari certifies that this is the approved version of the following dissertation:

Modeling Residential Self-Selection in Activity-Travel Behavior Models: Integrated Models of Multidimensional Choice Processes

Committee:

Chandra R. Bhat, Supervisor

Ram M. Pendyala

C. Michael Walton

S. Travis Waller

Stephen Donald

Modeling Residential Self-Selection in Activity-Travel Behavior Models: Integrated Models of Multidimensional Choice Processes

by

Abdul Rawoof Pinjari, B. Tech.; M.S.C.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2008

Dedication

To my mother, Murthuja bi and my father, Abdul Jaleel

Acknowledgments

I would like to express my sincere gratitude, first and foremost, to my advisor, Dr. Chandra Bhat, for his invaluable guidance, support and encouragement throughout my Ph.D. program. He has been a great mentor and is a person I would like to emulate in several ways. Many special thanks my former advisor Dr. Ram Pendyala for his encouragement and advice through the past six years. Thanks to Dr. Michael Walton, Dr. Travis Waller, and Dr. Stephen Donald for serving on my dissertation committee and for other helpful interactions.

Thanks to my colleagues at the University of Texas. Special thanks to Naveen on both professional and personal fronts. Thanks to Sudeshna also. Thanks to Siva and Jessica for their help and advice. Thanks to Ipek, Rachel, Erika, Bharath and Nazneen for several joint projects. Many thanks to Lisa for her help with the administrative work.

Special thanks to Laurel and other "mandamakara" friends, including Jack and Booze. Thanks to my friends at Austin: Vikas and Akansha, Saurabh, Ganesh, Hari and Ranju, Srikanth, Sneha, Aarti, Shreya, Nivi, Shalu and Ankur, and Jagadish. Thanks to my friends at Tampa: Granny, Chai, Ravi, Harish, Raghu, Uttam and Pallavi, and Vipan.

Lastly, but importantly, my deepest appreciation goes to my mother, my father, my brother, Samee and my sister, Roshina for their unconditional love, encouragement and unwavering support in all my endeavors. Indeed, they deserve the credit for this degree (and all of my achievements) more than I do. Very special thanks to my wife Sultana for her love, support and patience. Her presence made the last and difficult part of my doctoral study a joyous and memorable journey for me.

Modeling Residential Self-Selection in Activity-Travel Behavior Models: Integrated Models of Multidimensional Choice Processes

Publication No. ____________

Abdul Rawoof Pinjari, Ph.D.
The University of Texas at Austin, 2008

Supervisor: Chandra R. Bhat

The focus of transportation planning, until the past three decades or so, was to provide adequate transportation infrastructure supply to meet the mobility needs of the population. Over the past three decades, however, in view of increasing suburban sprawl and auto dependence, the focus of transportation planning has expanded to include the objective of sustainable development. Contemporary efforts toward sustainability include, for example, integrated land-use and transportation planning, travel demand management, congestion pricing, and transit and non-motorized travel oriented development. Consequently, in an effort to understand individuals' behavioral responses to (and to assess the effectiveness of) these policies, the travel demand modeling field evolved along three distinct directions: (a) Activity-based travel demand modeling, (b) Built environment and travel behavior modeling, and (c) Integrated land-use –

transportation modeling. The three fields of research, however, have progressed in a rather disjoint fashion.

The overarching goal of this dissertation is to contribute toward the research needs that are at the intersection of the three fields of research identified above, and to bring the three research areas together into a unified research stream. This is achieved by the simultaneous consideration of the following three aspects, each of which is of high importance in each direction of research identified above: (1) The activity-based and tour-based approaches to travel behavior analysis, (2) Residential self-selection effects, and (3) Integrated modeling of long-term land-use related choices and medium- and short-term travel-related choices. To this end, a series of integrated models of multidimensional choice processes are formulated to jointly analyze long-term residential location decisions and medium- and short-term activity-travel decisions (such as auto ownership, bicycle ownership, commute mode choice, and daily time-use). The models are estimated and applied using data from the 2000 San Francisco Bay Area Travel Survey to understand and disentangle the multitude of relationships between long-, medium-, and short-term choices.

This dissertation also formulates a multiple discrete-continuous nested extreme value model that can accommodate inter-alternative correlations and flexible substitution patterns across mutually exclusive subsets (or nests) of alternatives in multiple discrete-continuous choice models.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 BACKROUND

1.1.1 Objectives of Transportation Planning

The focus of transportation planning, until the past three decades or so, was primarily mobility-centric and supply-oriented. That is, the mobility needs (or the travel demand) of the population and the economic system were met by providing adequate transportation infrastructure supply. Over the past three decades, however, there has been an increasing realization that simply increasing the capacity (or supply) of transportation facilities is not a sustainable solution to meet the ever growing levels of travel demand. This is primarily due to the decreasing amount of available space for, and the escalating costs of, building additional transportation infrastructure. In addition, increased household income levels and easier availability of automobiles, and the resulting preferences to live in own and exclusive houses in the suburbs, have set a trend of sprawling suburbia and increasing auto dependency (see Ewing *et al.,* 2002; Litman, 2002; Newman and Kenworthy, 1998). Suburban sprawl and auto dependency, together, have been identified as causes of non-sustainable development and various adverse effects. These adverse effects include increased reliance on fossil fuel resources, traffic congestion, air quality non-attainment, health concerns (such as decreased physical activity levels and increased obesity problems), social inequity and/or segregation issues, and a reduction in the availability of transport alternatives such as transit/walk/bike modes of travel.

Over the years, the symptoms of non-sustainable development and other adverse impacts have become increasingly apparent, and the travel demand levels have increased higher than ever. Consequently, the mobility-centric and supply-oriented focus of transportation planning has expanded to include the objectives of (a) promoting sustainable and livable communities and urban areas by integrating transportation planning with land-use planning, and (b) addressing mobility needs and problems by managing travel demand within the available transportation supply. While the former objective focuses on coordinating transportation planning with land-use planning in an effort to control suburban sprawl, auto dependency, and the resulting adverse impacts, the latter objective focuses on reducing the need to add new transport infrastructure.

1.1.2 Role of Travel Demand Models

Until about three decades ago, when the transportation planning process was primarily mobility-centric and supply-oriented, and when the mobility needs were met by providing additional infrastructure supply, the main role of travel demand models was to predict the travel demand for future years to estimate the required amount of transportation supply. The travel demand prediction was carried out for various long-term socio-economic scenarios, and for alternative transportation system characteristics and land-use configurations (Bhat and Koppelman, 1999). For such a forecasting exercise, urban areas were divided into mutually exclusive spatial units labeled as traffic analysis zones (TAZs). Subsequently, a statistically oriented, "trip-based", four-step method was used to predict the aggregate number of "*trips*" between the TAZs (*i.e.,* the inter-zonal trips).

Over the past three decades, however, due to the expanded focus of transportation planning (integrated with land-use planning), there has been an increasing emphasis on the use of (a) travel demand management strategies, and (b) built environment policies as tools to mitigate traffic congestion, sprawl, and auto dependency, and the resulting adverse impacts. The travel demand management strategies modify the transportation system service characteristics to alter individual travel behavior and aggregate travel demand in an effort to accommodate the travel demand within the available transportation capacity and mitigate traffic congestion. Similarly, the built environment policies modify the land-use patterns to control sprawl, auto dependency and resulting adverse impacts. The interest in analyzing the potential of travel demand management and built environment policies, in turn, has led to a shift in the role of travel demand modeling from the statistical prediction of aggregate number of inter-zonal trips to understanding disaggregate-level (*i.e.,* individual-level) behavioral responses to travel demand management strategies and built environment policies. This is evidenced in the evolution of the travel demand modeling field along three directions: (a) Activity-based travel demand modeling, (b) Built environment and travel behavior modeling, and (3) Integrated land-use – transportation modeling. Each of these three research directions is discussed in turn in the next three sections.

1.2 ACTIVITY-BASED TRAVEL DEMAND MODELING

1.2.1 Rise of the Activity-based Approach

The interest in analyzing the potential of travel demand management and built environment policies has started placing higher demands on the policy evaluation abilities of travel demand models. For example, public policy mandates (such as the SAFETY-LU, ISTEA, TEA-21, and the CAAA) require travel demand models to be responsive to a host of transportation and land-use policies. At the same time, there has been a growing dissatisfaction in the field regarding the abilities of the traditionally used, statistically oriented, trip-based models in evaluating these policies, both from a predictive accuracy view point and a behavioral validity viewpoint (see Jones *et al.,* 1990, Axhausen and Garling 1992, and Bhat and Koppelman, 1999). These factors have resulted in the emergence of the behaviorally oriented activity-based approach to travel demand analysis.

The activity-based approach to travel demand analysis enables the evaluation of a wide range of travel demand management polices that cannot be analyzed, or can be analyzed only partially, using a traditional trip-based framework (Vovsha and Bradley, 2005). This is because of the conceptual superiority and behavioral realism of the activity-based approach that can be attributed to three salient features, which are (see Davidson *et al.,* 2007): (1) The recognition of *activities* as the underlying reason for travel, (2) The explicit treatment of *time* (see Bhat and Koppelman, 1999), and (3) The representation of travel patterns in the form of *trip chains* or *tours* (see Davidson *et al.,* 2007).

1.2.2 Recognition of Activities as the Underlying Reason for Travel

The trip-based approach to modeling travel demand directly focuses on travel, without explicit recognition of the motivation or reason for the travel. This is difficult to justify from a behavioral standpoint, since it is unlikely (in general) that travel occurs without a reason. Rather, the needs of the households and individuals are likely to be translated into a requirement that individuals be present at different places at different times. Thus, the activity-based approach views travel as derived from the need to participate in activitites at different points in space and time (see Jones 1979; Carpenter and Jones, 1983; Kitamura, 1988; Jones *et al.,* 1990; and Axhausen and Garling, 1992). This need for activity participation, to a large extent, drives travel decisions.

1.2.3 Explicit Treatment of Time

In the trip-based approach, "*time is reduced to being simply a "cost" of making a trip"* (Bhat and Koppelman, 1999) and a day is viewed as a combination of broadly defined peak and off-peak time periods. The activity-based approach, on the other hand, treats time as the main backdrop/setting within which the daily activity and travel (or activity-travel) related decision-making takes place (see Kurani and Lee-Gosselin, 1996). The central basis of the activity-based approach to travel demand modeling is that individuals' daily activity-travel patterns are a result of their time-use decisions (see Pas, 1996, Pas and Harvey, 1997, and Bhat and Koppelman, 1999). That is, individuals have a limited amount of time available (for example, 24 hours in a day), and must make decisions on how to allocate their time to various activities subject to their socio-demographic, spatial, temporal, transportation system, and other contextual constraints. These decisions determine the generation and scheduling of travel. Hence, determining the impact of travel demand management strategies and built environment policies on activity participation and time-use (or activity time-use) behavior is an important precursor step to assessing the impact of such polices on individual travel behavior. For example, one may analyze whether improving a neighborhood with walkways, bikeways, and recreational parks encourages individuals to invest more time in physically active recreation pursuits in the place of in-home passive recreation (such as watching television or playing computer games). The travel dimensions can then be "derived" from the changes in time-use and activity-scheduling patterns.

1.2.4 Representation of Travel Patterns: Tour-based Structure

The trip-based approach uses individual trips as the unit of travel demand analysis. Each trip is considered as independent of another trip, without considering the inter-relationship in the choice attributes (such as mode, destination, and time) of different trips within a single sojourn from home or other places such as work. The activity-based approach, on the other hand, uses *tours* (defined as a chain of trips beginning and ending at a same location, say, home or work) as the basic elements to represent and model travel patterns. By using a tour-based structure, of which the trips are a part, the activity-based approach captures the interdependency (and consistency) of the modeled choice attributes among the trips of a same tour.

1.3 BUILT ENVIRONMENT AND TRAVEL BEHAVIOR MODELING

1.3.1 Built Environment Policies

As a result of the expanded focus of transportation planning (integrated with land-use planning), there has been an increasing emphasis on the use of built environment policies as tools to mitigate sprawl and auto dependency, and resulting adverse impacts. Since sprawl is associated with such characteristics as: (1) widely spread population in low-density developments, (2) homes separated from other activity centers, including workplaces and shops (or segregated land-use), (3) poor accessibility to activity centers and the transportation network, and (4) "*lack of well defined and thriving activity centers, such as downtowns and town centers"*(Ewing *et al.,* 2002), several built environment polices aim to: (a) promote high-density (or compact) and mixed land-use developments through zonal ordinances, (b) bring residents closer to destinations, and (c) revitalize downtowns and Brown fields. Similarly, since auto dependency is associated with reduced consideration and provision of non-motorized and transit modes of travel (see Litman, 2002; and Newman and Kenworthy, 1998), another set of built environment polices aim at developing walk and bicycle friendly neighborhoods and transit-oriented developments. All of these strategies fall under what is known as the *smart growth* theory of urban and transportation planning or the *neotraditional*[1] movement of urban design. They attempt to modify the land-use configuration in an effort to influence the housing, employment, and transportation choices of households and individuals to control sprawl, auto dependency, and related adverse impacts. These strategies are of interest not only to transportation professionals and urban planners, but also to public health officials. The overall objective of these strategies, as indicated earlier, is to improve the livability of communities by promoting sustainable neighborhoods to live in, work at, and travel to, and to reduce the health, environmental, and societal impacts.

1.3.2 Built Environment and Travel Behavior: A Debate

As the popularity of built environment policies has increased, a number of agencies have started adopting or are considering smart growth strategies targeted toward reducing sprawl and auto dependency (see Transportation Research Board Conference Proceedings on Smart Growth and Transportation, 2005, for a review of agencies that have adopted such strategies). However, along with the rise in the popularity, there has also been a debate on the effectiveness of built environment policies in modifying travel behavior and improving overall quality of life. The debate is whether the built environment causally affects travel behavior or whether the relationship between the built environment and travel behavior is a mere statistical correlation. In this context, while a number of studies have shown a favorable and causal impact of the built environment policies on travel behavior (see Ewing and Cervero, 2001), there have also been studies that have not

[1] The terms smart growth, neotraditional design, traditional neighborhood design, and neo urbanism are all used to refer to a similar set of practices such as "*the planning, design, development, and revitalization of cities, towns, suburbs, and rural areas in order to create and promote social equity, a sense of place and community, and to preserve natural as well as cultural resources*" (The American Planning Association (APA), 2002), or the *set of development practices to create more attractive, efficient, and livable communities* (The TDM encyclopedia of the Victoria Transport Policy Institute website, accessed November 30, 2007)

found any discernible impact of the built environment on various dimensions of travel behavior (see Crane, 2000).[2]

The empirical evidence accumulated over the past 15 years indicates an association between residence in high density and mixed land-use neighborhoods and characteristics of "desirable" travel behavior, including lower auto ownership, lesser driving, and more walking and bicycling (see, for example, Cervero and Duncan, 2003; Krizek, 2003; Rajamani *et al.,* 2003; and Shay and Khattak, 2005). However, to date, there has not been a conclusive consensus on the underlying mechanism of this association. Specifically, there are still unanswered questions such as (1) is the relationship between the built environment and travel behavior causal in nature or merely a statistical correlation? and, (2) if there is a causal relationship, what is the extent of this causal relationship? To answer these questions one must consider the role of residential self-selection in the built environment – travel behavior relationship.

1.3.3 The Residential Self-Selection Phenomenon

In the typical approach to assessing the impact of built environment on travel behavior, the residential built environment attributes are considered pre-determined and exogenous, and are used as independent variables to explain travel behavior. Such an approach implicitly assumes a one-directional relationship between built environment and travel behavior, according to which the built environment impacts travel behavior. Assuming such a one-way causal relationship implies that households and individuals first locate themselves in neighborhoods based on market forces such as housing affordability, crime statistics, and school quality. Their travel behavior is then shaped by the neighborhood characteristics (*i.e.,* the built environment attributes). The above reasoning would imply that the neighborhood attributes can be modified to achieve a significant amount of desired shift in travel behavior. A potential fallacy in such a one-way cause-and-effect assumption, which implies a sequential nature of residential location and travel choice

---

[2] Opponents of smart growth and new urbanism policies (especially, the regulatory measures that encourage high density development, mixed land-use, and public transit) have also cautioned against a potential risk of increased housing prices and ignoring the housing and lifestyle preferences of consumers (see, for example O'Toole, 1996; and the Decomgraphia website accessed in November 2007).

decisions (in that order), is that it ignores the possible bi-directional nature of the decisions.

In a bi-directional relationship, both the residential location choice and the travel preferences may affect each other. In one direction, the residential built environment may affect travel behavior, and hence changes in the residential built environment "causes" changes in travel behavior. In another direction, the travel-related preferences may affect the residential location choice. That is, households and individuals may "self-select" to reside in neighborhoods that allow them to pursue activities and travel compatible with their socio-demographics (*e.g.*, income), attitudes (*e.g.*, auto-disinclination), and travel preferences (*e.g.*, preference for smaller commute time). This "residential self-selection" phenomenon leads to the possible associative nature of the relationship between built environment and travel behavior where the built environment and travel behavior may be merely correlated rather than any underlying causality. In such case, changes in the built environment may not cause any changes in travel behavior.

To understand this issue better, consider a specific built environment policy to improve bicycling facilities, with the objective of reducing automobile dependence and increasing physically active recreational pursuits. To assess the impacts of such a policy, assume that a data collection effort has been undertaken to examine the bicycling levels of individuals in neighborhoods with different levels of existing bicycling facilities. An analysis of this data may find that individuals residing in neighborhoods with good bicycling facilities pursue more bicycling-related activities. The question is whether this relationship between bicycling facilities and bicycling levels is "causal" or "associative". That is, whether this relationship implies that building neighborhoods with good bicycling facilities would result in higher bicycling levels in the overall population (*i.e.,* a "causal" relationship), or whether this relationship is an artifact of individuals who are bicycling-inclined (because of, say, physical fitness consciousness) self-selecting themselves to reside in neighborhoods with good bicycling facilities (*i.e.,* an "associative" relationship). If the latter "residential sorting (or self-selection) process" is at work, building neighborhoods with good bicycling facilities would not result in higher bicycling levels in the overall population, but simply lead to an alteration of spatial residence patterns of the population based on physical fitness consciousness.[3]

In reality, the nature of the relationship between built environment and travel behavior may be part causal and part associative (and hence bi-directional). Thus, any attempt to examine the built environment-travel behavior connection should disentangle the causal and associative elements of the relationship to inform and contribute to the credible assessment of the impact of built environment policies on travel behavior. Specifically, ignoring the residential self-selection effects (when present), can result in the identification of "spurious" causal effects of neighborhood attributes on travel behavior and lead to distorted policy implications.

Econometrically speaking, accommodating residential self-selection effects in models of travel behavior (and hence the bi-directional nature of the relationship between the residential built environment and travel behavior) calls for an endogenous treatment of the residential location choice. That is, the residential choice and travel-related decisions ought to be viewed, modeled, and analyzed simultaneously (*i.e.,* in an integrated fashion) so that the impacts of the travel-related preferences on residential location choices may be captured along with the impact of the residential locations on travel behavior.

1.4 INTEGRATED LAND-USE – TRANSPORTATION MODELING

Most of the travel demand models treat the longer-term choices concerning the housing decisions (such a residential tenure, housing type, and residential location), employment choices (such as entry into/exit from labor market, employment type and arrangement, and employment location), and vehicle ownership as exogenous inputs. In such cases, the possibility that households can adjust with combinations of short-, medium- and long-term behavioral responses to land-use and transportation policies is systematically

[3] As indicated in Bhat and Guo (2007), the assumption here is that there is an adequate supply of neighborhoods to choose from for individuals who are bicycling-oriented. If there is an undersupply of such neighborhoods, then enhancing bicycling facilities in some neighborhoods would increase overall bicycling activity across the population even if the only process at work is residential sorting.

ignored (Waddell, 2001). A significant increase in transport costs, for example, could result in a household adapting with any combination of daily activity and travel pattern changes, vehicle ownership changes, job location changes, and residential location changes (Waddell, 2001). As the focus of transportation planning has expanded to integrate transportation planning with land-use planning, it has become important to accurately analyze the impact of land-use and transportation policies on the short-, medium- and long-term choices that influence land-use and travel demand. Consequently, a stream of research under the label of "integrated land-use – transportation" modeling has evolved.

Integrated land-use – transportation modeling is primarily concerned with understanding and modeling the interactions between the processes that influence regional land-use patterns (*i.e.,* the spatial pattern of urban activities and development) and the processes that influence regional travel demand patterns on the transportation network. Modeling these interactions requires making connections between the long-, medium-, and short-term individual and household choices that influence land-use and travel demand. In other words, the integrated modeling of land use and transportation involves analyzing the entire choice continuum defining individual and household choices across different temporal scales. The choices include long term choices such as residential and work location choices, medium term choices such as vehicle and bicycle ownership, and short term choices such as mode choice, destination choice, and departure time choice. While the residential and work location choices influence land-use patterns, the vehicle/bicycle ownership, mode, destination, trip departure time and other travel choices influence travel demand.[4]

[4] This dissertation limits the discussion of integrated land-use transportation modeling to understanding (and modeling) the interactions between the individual and household choices that influence land-use patterns and travel demand patterns. In a much broader sense, integrated land-use travel demand modeling includes other important aspects such as the interactions between individuals and households, and other decision makers (or players) within the housing, labor, and transportation markets. The other players include real estate developers, employers, and production, manufacturing and service firms. The choices made by all the players together influence the land-use patterns and transportation system network and influence the evolution of urban systems.

1.5 GAPS IN THE LITERATURE

The preceding discussion has indicated the evolution of the travel demand modeling field along three directions: (1) Activity-based travel demand modeling, (2) Built environment and travel behavior modeling, and (3) Integrated land-use transportation modeling. While there has been a substantial amount of previous research in the three areas identified, there are still research needs in these areas, which are discussed in turn in the subsequent three sections.

1.5.1 Gaps in Activity-based Travel Demand Modeling Research

A limitation of activity-based travel modeling research is the inadequate consideration of the effects of the built environment (or land-use) on activity-travel behavior. While there have been major advances in addressing the various aspects of activity-travel behavior including activity participation and time-use (or activity time-use) analysis, and tour-based travel analysis, only a few recent studies have attempted to consider the impact of a comprehensive set of built environment (or land-use) characteristics on these aspects of activity-travel behavior. Even if some activity-based travel modeling research efforts have considered the impacts of the built environment, all activity-based travel demand model systems developed to date are limited in the following two ways:

(1) Activity-based travel demand models ignore residential self-selection effects while considering the impact of the built environment. For example, most existing activity participation and time-use studies ignore the impact of the residential self-selection effects while assessing the built environment impacts. Similarly, no tour-based models account for residential self-selection effects.

(2) The activity-based travel demand modeling methods use a relatively "dis-integrated" land-use – activity-travel demand modeling approach in which longer-term land-use and travel-related decisions (such as residential, employment, and auto ownership choices) are simply assumed as exogenous inputs.

Another limitation of the literature in this field is that the activity participation and time-use models have focused more on the activity generation aspect of activity-travel behavior. That is, the time-use studies to date focus on the types of activities undertaken by individuals within a given time frame, and do not adequately consider the settings (*i.e.,* the spatial, temporal, scheduling, sequencing, and travel contexts) within which the activities are pursued. On the other hand, it is important to understand the contexts of activity participation and time-use to answer the where (location), when (timing), and how (travel mode and route) of activity participation. Such an understanding can enable an integration of activity participation and time-use analysis and travel behavior analysis.

1.5.2 Gaps in Built Environment and Travel Behavior Modeling Research

A limitation of the built environment and travel behavior modeling research is the adoption of a "trip-based" approach to modeling travel behavior while assessing the impact of the built environment on travel behavior. More specifically, the first major limitation in this area of research is the use of a trip-based approach (rather than the activity-based approach) while accommodating residential self-selection effects on the impact of the built environment on travel behavior. Although several recent built environment and travel behavior studies have alluded to and/or accommodated residential self-selection in one of several ways, almost all earlier efforts attempt to study the built environment and travel behavior relationship by directly focusing on specific travel behavior dimensions, such as trip frequency or trip mileage for one or more trip purposes.[5] These studies ignore: (1) the intervening effects of activity participation and time-use on the impact of the built environment on travel behavior, and (2) the tour-based structure of travel patterns, while assessing the residential self-selection effects.

[5] See Chatman (2005), Kitamura *et al.* (1997), Schwanen and Mokhtarian (2003), Boarnet and Sarimento (1998), Greenwald and Boarnet (2001), Khattak and Rodriguez (2005), Handy *et al.* (2006), Handy and Clifton (2001), and Krizek (2000 and 2003) for analyses of residential self-selection in models of trip frequency by one or more modes and/or purposes; Schwanen and Mokhtarian (2005b), Khattak and Rodriguez (2005), Bagley and Mokhtarian (2002), Handy *et al.* (2005), and Krizek (2000, 2003) for analyses of residential self-selection in models of travel mileage by one or more modes; Cervero and Duncan (2002), Hammond (2005), Pinjari *et al.* (2007), Schwanen and Mokhtarian (2005a), Salon (2006), and Zhang (2006) for commute mode choice analyses that consider residential self-selection; and Cervero and Duncan (2003), Greenwald (2003), and Salon (2006) for non-commute mode choice analyses that accommodate residential self-selection.

From a methodological stand point, the second major limitation in this area of research is the lack of appropriate methodologies to incorporate residential self-selection in travel demand models. To be sure, several mathematical and econometric methods exist to account for residential self-selection in travel demand models (see Chapter 2, and/or Mokhtarian and Cao, 2008; and Bhat and Guo 2007, for a review of the methodologies). However, there is still a need to develop appropriate methods to accommodate residential self-selection in several mathematical structures used to model travel demand, including the most commonly used multinomial logit (MNL) model[6] (of, say, travel mode choice) as well as models of multiple choice making that are gaining prominence in recent years. Further, most of the methods do not adopt an "integrated" land-use travel demand modeling approach. Thus, the third major limitation in this area of research is that most studies adopt a "dis-integrated" land-use travel behavior modeling approach, and ignore the interconnections between the long-, medium-, and short-term behaviors of individuals and households while evaluating the impacts of the built environment policies. More specifically, although the endogeneity of residential location choice (which is a long-term choice) is considered while accommodating the residential self-selection effects (see Section 1.3.3), the residential location choice is not explicitly considered and modeled, and/or the intervening impacts of the medium-term choices such as auto ownership and bicycle ownership are ignored.

Other gaps in the literature include: (a) lack of studies addressing residential self-selection effects in models of activity location choice, and (b) lack of studies addressing the heterogeneity in the population of the extent of residential self-selection.

1.5.3 Gaps in Integrated Land-use – Transportation Modeling Research

The primary limitation of integrated land-use transportation models is that they are not "truly" integrated because they do not consider all the long-term, medium-term, and short-term choices of households and individuals within a unified integrated modeling

[6] It is very important to develop appropriate methods to control for residential self-selection in the commonly used mathematical models of travel demand, because of the widespread use of these models in practical regional travel demand modeling.

framework. For example, as indicated in the preceding section, the intervening impacts of the medium-term choices such as auto ownership are ignored when considering the interconnections between the long-term choices (e.g., residential location choice) and short-term choices (e.g., commute mode choice). It is important to develop appropriate integrated modeling methods to understand the impact of sociodemographics, built environment, and transport system characteristics on the various choice dimensions including residential location, auto and bicycle ownership, and activity travel decisions in an integrated fashion.

Another major limitation of integrated land-use – transportation modeling research is the adoption of a "trip-based" approach to represent and model travel behavior without considering the activity-based aspects such as activity participation and time-use aspects, and the tour-based structure of travel choices. It is important to consider travel behavior in the context of activity participation and time-use, and use a tour-based structure to represent travel choices, while integrating travel demand models with land-use models.

It is also interesting to note that the appropriate representation of land-use related choices and transportation choices, and the modeling of the interactions between these choices within a "truly" unified and integrated framework, can address the various objectives of the built environment and travel behavior research, including the incorporation of residential self-selection effects and an accurate assessment of the impact of the built environment on travel behavior. This is because, as indicated in Section 1.3.3, incorporating the residential self-selection effects requires a joint framework to model and analyze the residential location and travel choices in an integrated fashion. However, most of the integrated land-use travel behavior modeling research does not adequately consider the built environment impacts, and ignores residential self-selection effects.

1.6 OBJECTIVES OF THE DISSERTATION

The preceding discussion has indicated the evolution of the travel demand modeling field along three distinct directions – activity-based travel demand modeling, built environment and travel behavior modeling, and integrated land-use transportation modeling – and identified the research needs (or gaps) in the three directions. While the emergence of the activity-based approach may be attributed to the need to understand individual-level behavioral responses to travel demand management policies, the research on built environment and travel behavior has been driven by the need to accurately estimate the impact of built environment policies on travel behavior. The integrated land use-travel demand modeling has emerged because of efforts to examine the impacts of land-use and transportation planning policies and infrastructure investments on both land-use and transportation patterns within a unified behavioral and policy analysis framework.

An important observation that can be made from the discussion is that the major gaps in the literature are at the very intersection of the three streams of research. That is, the aspects that are important in one stream of research have not been considered in the other two research streams. In particular, while the activity-based travel modeling research has ignored the impact of residential self-selection effects when assessing built environment impacts, the built environment and travel behavior research has adopted a trip-based travel modeling approach when considering residential self-selection effects. Further, while both the activity-based travel modeling, and the built environment and travel behavior research, fields have adopted a "dis-integrated" land-use – travel demand modeling approach, the integrated land-use – travel demand models have ignored both the activity-based aspects (of travel demand) and the residential self-selection effects. In other words, the three streams of research have progressed in a rather disjoint fashion.

In view of this discussion, <u>the overarching theme and goal of this dissertation</u> is to contribute toward the research needs that are at the intersection of the three streams of research identified earlier, and to bring the three research areas together into a unified research framework.

The goal of the dissertation is achieved by the simultaneous consideration of the following three aspects, each of which is of high importance in each direction of research identified above: (1) Activity-based and tour-based approach to travel behavior analysis, (2) Consideration of a comprehensive set of built environment impacts and corresponding residential self-selection effects, and (3) Integrated modeling of long-term land-use related choices and medium- and short-term travel-related choices. To this end, a series of integrated models of multidimensional choice processes are formulated to jointly analyze long-term residential location decisions and several medium- and short-term activity-travel decisions. The medium-term decisions are the auto ownership and bicycle ownership decisions and the short-term decisions include commute mode choice and activity participation and time-use behavior. More specifically, the goal of the dissertation is pursued by considering the following objectives.

The first objective is to formulate an integrated residential location choice and commute mode choice model to accommodate residential self-selection effects in mode choice decisions, while assessing the impact of the built environment on both these decisions. This objective serves the following two purposes: (1) To develop an appropriate integrated land-use transport modeling methodology to accommodate residential self-selection effects in multinomial choice variables, and (2) To fill the gap at the intersection of integrated land-use transportation modeling and built environment and travel behavior modeling fields.

The second objective is to enhance the above developed modeling methodology to formulate a multidimensional econometric methodology to analyze residential location, auto ownership, bicycle ownership, and "tour-based" commute mode choices in an integrated fashion. Such a multidimensional choice model can be used to analyze the various interconnections between long-term, medium-term, and short-term choices, including residential self-selection effects. This effort represents a movement toward "truly" integrated land-use travel demand models (where long-, medium-, and short-term choices are modeled and analyzed in a unified integrated framework), as well as the bringing together of all the three streams of research identified earlier.

The third objective is to develop a heterogeneously-joint choice model and apply it to the context of neighborhood type choice and bicycle ownership decisions. This helps in understanding the heterogeneity in residential self-selection effects (in bicycle ownership decisions) across various socio-demographic segments of the population.

The fourth objective is to formulate an integrated model of residential location choice and activity time-use behavior that can be used to account for residential self-selection effects in activity participation and time-use behavior. This objective also serves to bring together all the three streams of research identified earlier.

The fifth and final objective of this dissertation is to use the afore-mentioned modeling methods for practical policy analysis purposes to: (a) quantify the built environment influences on activity-travel choices, and (b) assess the extent of residential self-selection effects on evaluating these influences.

1.7 ORGANIZATION OF THE DISSERTATION

The rest of this dissertation is organized in seven additional chapters as follows.

Chapter 2 contributes to the first objective by developing a joint residential location and commute mode choice model to accommodate residential self-selection effects, while assessing the impact of the built environment on both these decisions. From a methodological standpoint, the chapter presents a methodology for jointly modeling the relationship between two unordered multinomial discrete choice variables – residential location choice and commute mode choice. To our knowledge, this is the first instance of the development of a joint model system that can capture both causal and associative relationships between two unordered multinomial choice variables. Further, as opposed to the multidimensional multinomial (or nested) logit modeling approaches adopted in the literature, this chapter uses a mixed multidimensional choice modeling approach to model several choice dimensions in an integrated fashion. Specifically, this chapter extends the mixed joint multinomial residential location choice and ordinal auto ownership choice model developed by Bhat and Guo (2007) to develop a joint model of multinomial residential location choice and multinomial mode choice.

Chapter 3 contributes to the second objective by extending the modeling framework developed in Chapter 2 to formulate a multidimensional model of residential location, auto ownership, bicycle ownership, and tour-based commute mode choices. The framework accommodates the intervening impact of medium-term choices (such as auto ownership and bicycle ownership) while modeling the interconnections between long-term decisions (residential location choice in this case) and short-term decisions (travel model choice in this case). The model developed in this chapter, to our knowledge, represents the first attempt in the literature to model four important choice dimensions in an integrated framework. In addition, in this model, the residential self-selection effects are simultaneously accommodated in three different models of travel behavior – auto ownership, bicycle ownership, and commute mode choice. Finally, all of the above-identified aspects are accommodated in the context of a "tour-based" commute mode choice model as opposed to a "trip-based" mode choice model. In summary, the integrated model development and application in this chapter contributes to, and brings together, three different fields of research – integrated land-use travel demand modeling, built environment and travel behavior modeling, and tour-based travel demand modeling.

Chapter 4 contributes to the third objective by focusing on the heterogeneity in residential self-selection effects in bicycle ownership decisions.

Chapter 5 contributes to the fourth objective by developing an integrated model of residential location, and daily activity time-use decisions. Specifically, the model accommodates residential sorting effects in examining the impact of built environment variables on individual time-use in maintenance activity (grocery shopping, household chores, personal care, *etc.*) and several types of discretionary activity purposes. A salient feature of the model is that the self-selection effects are accommodated simultaneously in the activity participation and time-use behavior (or the activity time-use behavior) in multiple types of activities. That is, the model recognizes the possibility that an individual can participate in multiple types of activities (as opposed to a single activity engagement) in a given period of time. The multiple activity participation and time-use behavior (*i.e.,* the activity time-use behavior) comes under what has recently come to be labeled as multiple discrete-continuous choice making behavior (see Section 5.2 for further details). Hence, to accommodate and model residential self-selection effects in activity time-use behavior, the joint residential choice-activity time-use model system in the chapter takes the form of a joint mixed multinomial logit – multiple discrete-continuous extreme value (MNL–MDCEV) model. To our knowledge, this is the first instance in the econometric or other literature of the development of such a model to jointly analyze an unordered discrete variable (residential location choice in the current context) and multiple discrete-continuous variables (activity participations and time-use decisions in multiple activities in the current context). Further, from a substantive viewpoint, this chapter contributes to the objective of bringing together activity-based travel demand modeling research and built environment and travel behavior modeling research.

Chapter 6 proposes an econometric modeling methodology to accommodate the interdependencies (or similarity, and hence the correlations) among the various activity types of the multiple discrete-continuous activity time-use model component of the joint MNL–MDCEV model system developed in Chapter 5. More specifically, this chapter formulates a multiple discrete-continuous nested extreme value (MDCNEV) model. The model can accommodate the interdependencies (or the interalternative correlations) and resulting flexible substitution patterns across mutually exclusive subsets (or nests) of alternatives in multiple discrete-continuous choice models. The salient feature of the model is that it offers closed form probability expressions for any (and all) multiple discrete-continuous choice patterns while accommodating the interalternative correlations. Thus, the MDCNEV model offers a powerful alternative to the currently used mixed MDCEV modeling approach that necessitates simulation-based estimation to accommodate interalternative correlations. This extension of the MDCEV model should constitute an important state-of-the-art development in multiple discrete-continuous choice modeling field, which is analogous to the nested logit extension of the multinomial logit model in single discrete choice modeling research.

Chapter 7 contributes to the fifth objective by employing the models developed in the previous chapters for various policy simulation analyses to: (a) quantify the built environment influences on activity-travel choices, and (b) assess the extent of residential self-selection effects on evaluating these built environment influences.

Chapter 8 concludes the dissertation by recapitulating the findings from, and contributions of, this research, and identifying directions for further research.[7]

[7] The reader will note that the literature specific to the topic covered in each chapter is reviewed in the corresponding chapter.

CHAPTER 2

MODELING RESIDENTIAL SELF-SELECTION IN A MULTINOMIAL LOGIT MODEL OF COMMUTE MODE CHOICE

2.1 INTRODUCTION AND MOTIVATION

The importance and the complexity of the land use - travel behavior relationship has been recognized for several decades in the transportation planning practice and research communities. The complexity of the land use - travel behavior association arises due to (1) the multitude of dimensions that define land use (for example, land use mix, urban form, street block density, and local network features) and travel behavior (such as auto ownership, mode choice, and overall travel demand), and (2) the possibility of multiple causal and/or pure associative relationships between the dimensions that define land use and travel behavior (Bhat and Guo, 2007).

In conventional transportation planning practice, a one-way causal flow is often assumed in which land use patterns affect travel behavior. Assuming such a one-way causal relationship would mean that households and individuals first locate themselves in neighborhoods based on market forces such as housing affordability, crime statistics, and school quality. Their travel behavior is then shaped by neighborhood characteristics (or built environment attributes).

The above reasoning would imply, for example, that land use patterns and neighborhood attributes can be modified to achieve a desired shift in travel mode shares. The potential fallacy in such a one-way cause-and-effect assumption (as discussed in Section 1.3), which implies a sequential nature of residential location and mode choice decisions (in that order), is that it ignores the associative nature of the decisions. That is, the relationship between residential location and travel mode choice decisions may be a mix of partial cause-and-effect linkage and partial associative correlation. In reality, households and individuals may locate (or self-select) themselves into neighborhoods that allow them to pursue their activities using modes that are compatible with their socio-demographics (*e.g.*, income), attitudes (*e.g.*, auto-disinclination), and travel preferences (*e.g.*, preference for smaller commute time). Ignoring such residential self-selection effects (when present) can result in the identification of "spurious" causal effects of neighborhood attributes (or the built environment attributes) on travel behavior and lead to distorted policy implications. To correctly assess the impact of land-use patterns on mode choice, one must recognize and control for the associative correlations that may arise due to residential sorting. Further, if such residential self-selection phenomenon dominates the cause-and-effect linkages, then urban land-use policies aimed at modifying neighborhood attributes for inducing mode shifts would alter the spatial residential location patterns more than the mode choice patterns.

The issue of residential self-selection analysis and modeling has gained timely prominence over the past decade. However, as mentioned in Section 1.5.2, there is still a need to develop appropriate methods to account for residential self-selection in several mathematical structures used to model travel choices, including the most commonly used multinomial logit (MNL) model. To be sure, while current methods may be used to control for residential self-selection in MNL models of travel choices, most of these methods do not adopt an "integrated" land-use travel demand modeling approach. More specifically, although the endogeneity of residential location choice is considered (to accommodate residential self-selection effects), the residential location choice is not explicitly considered and modeled along with the travel behavior model of interest in an "integrated" fashion. While the nested logit modeling method can be used to model residential location choice and travel choices in an integrated fashion, the approach cannot clearly distinguish the residential self-selection effects with respect to different built environment attributes (see Section 2.2 for detailed explanation).

In light of the preceding discussion, the specific objectives of this Chapter are to:

1. Review the methods used in the literature to test the presence of, and control for, residential self-selection in models of travel behavior, and identify an appropriate

integrated land use travel demand modeling method to address residential self-selection in models of travel behavior.

2. Develop an appropriate integrated econometric modeling method to address residential self-selection in the most widely used multinomial logit (MNL) modeling structure.
3. Apply the integrated econometric modeling methodology to analyze the impact of sociodemographic, built environment and transportation system characteristics on multinomial discrete variables of residential location choice and a “trip-based” commute travel mode choice, while controlling for residential self-selection effects (along with a review of the studies that attempt to address residential self-selection in mode choice behavior).

2.2 ADDRESSING RESIDENTIAL SELF-SELECTION: A REVIEW OF MODELING METHODS

As the issue of residential self-selection has gained substantial attention in recent years, a variety of research and modeling approaches have been used to address residential self-selection effects in models of travel behavior. Two recently completed studies, Bhat and Guo (2007) and Mokhtarian and Cao (2008), provide a thorough review of the existing methodological approaches. Specifically, Mokhtarian and Cao (2008) distinguish the existing methodological approaches to control for residential self-selection into seven categories: direct questioning, statistical control, instrumental variables models, structural equations models, sample selection models, joint discrete choice models, and models with longitudinal data and designs[8]. This discussion uses the same classification labels to provide an overview of the methods and to discuss their suitability for the current context. Then, the subsequent discussion focuses on joint discrete choice models that are of primary interest to this research.

[8] See Cao et al. (2006a) for an excellent empirically-oriented review of the residential-self selection issue focusing on empirical studies and their findings

2.2.1 Overview of the Methods

The direct questioning approach, as described by Mokhtarian and Cao (2008), is a straight forward way of qualitatively assessing the presence of self-selection. It entails simply asking people if their travel preferences and attitudes influenced their choice of residential neighborhood. The method can be improved by asking the (qualitative)extent to which their preferences and attitudes (as well as the neighborhood attributes) influence the selection of residential neighborhood. The approach, however, can only provide a qualitative assessment, and does not provide a means to quantify the true impacts of the built environment and the extent of residential self-selection. Other disadvantages include the potentially high biases introduced by the interviewer and/or the interviewee during the direct questioning process, either in a face-to-face interview or in a non-human interface interview. Further, this is not a modeling method as such. That said, the approach can provide qualitative insights and guide the specification and development of integrated land-use travel demand models.

The statistical control approach explicitly accounts for the influence of decision-maker demographic attributes, attitudes, and perceptions that may impact the residential neighborhood choice. This is achieved by including the attributes as explanatory variables in models of travel behavior. The basic idea is to control for *all* demographic and attitudinal factors that can potentially contribute to residential self-selection, while estimating the impact of the built environment on travel behavior. The effect of BE measures in such a model would be closer to the true effect. Most studies in the literature include the demographic attributes, but only a selected number of studies include attitudinal factors and perceptions (see, for example, Kitamura *et al.,* 1997; Schwanen and Mokhtarian 2005a, and Handy *et al.,* 2005). While it is desirable (and, if possible, most preferable) to measure and include as many attributes as possible to control for self-selection, it is not straight forward and practically feasible to measure and include all possible attitudes. Further, most standard surveys do not collect information on attitudinal factors.

The instrumental variables approach, which adopts a two-stage modeling procedure, is a classic econometric method to control for endogeneity of explanatory variables (see Wooldridge 2002 for a text book explanation of this approach). In the first stage, the neighborhood attributes are regressed on *instruments* that are believed to be related to the neighborhood attributes but are not correlated (in an ideal case) with the travel behavior dimension of interest. In the second stage, the predicted values of the neighborhood attributes from the first-stage regression are introduced as explanatory variables in the travel behavior model of interest. The estimated impacts of these predicted neighborhood attribute values represent the cleansed and true impacts of the built environment on travel behavior. Examples of the application of this technique include Boarnet and Sarimento (1998), Boarnet and Crane (2001), and Greenwald and Boarnet (2001). The well-known drawback of the instrumental variables approach is its inapplicability to the case where the travel behavior model of interest has a non-linear structure such as the multinomial logit model. While more recent literature (Berry *et al.,* 1995; Lewbel, 2004; and Louviere *et al.,* 2005) consists of alternative methods to deal with endogenous explanatory variables in the context of discrete choice models, such methods require computationally tedious standard error corrections to be able to make appropriate statistical inferences (Bhat and Guo, 2007). Further, due to the "dis-integrated" modeling approach, one cannot use the method to understand (a) the impacts of built environment on both residential location choices and travel behavior choice, and (b) the intervening impacts of medium term decisions such as auto ownership and bicycle ownership.

The structural equations modeling (SEM), approach is used, in general, to model the multi-way interrelationships among multiple endogenous variables of interest. This offers the ability to test which directions of causality are statistically supported by the data (Mokhtarian and Cao, 2008). Further, and more importantly, the endogenous variables can be modeled as directly influencing each other, which enables the accommodation of multi-way (or bidirectional) causal relationships between the endogenous variables. The ability to accommodate simultaneous bidirectional causality between endogenous variables is a particular advantage of the structural equations modeling approach over all other methods discussed in this section. In the case of residential self-selection, the endogenous variables can potentially be the travel behavior variable(s) of interest, the built environment attributes, and the attitudinal attributes. Thus, structural equations modeling approach offers a powerful approach to model the interrelationships between travel behavior, built environment, and attitudinal variables and hence minimizing the potential self-selection effects on the estimated impact of the built environment. The major drawback of the structural equations methodology, however, is the inability to model multinomial discrete variables, which are of high interest in regional travel demand modeling. Applications of this methodology in the residential self-selection literature include, for example, Bagley and Mokhtarian (2002) on a cross-sectional data. Mokhtarian and Cao (2008) indicate the use of structural equations methodology to longitudinal data as a very useful extension to appropriately incorporate the time-precedence of cause and effect (*i.e., cause* occurring first and then *effect* occurring later in time) in analyzing the causal impacts of the built environment on travel behavior. Cao et al (2007) employ structural equations modeling (SEM) approach to quasi-longitudinal data based on a retrospective survey.[9, 10]

The sample selection modeling approach, which we label as the binary sample selection modeling, models the prior selection into binary *residential location(or neighborhood) types* (such as suburban or urban types, and traditional or neo-traditional types), and model the travel behavior variable of interest as conditional upon that prior selection. The original form of the method was introduced in a classical paper by Heckman (1979) with a consistent estimator to correct for sample selection bias due to endogenous binary explanatory variables in linear regression models. Such modes in the context of residential self-selection typically have a joint equation system with the latent

---

[9] Another longitudinal residential self-selection study in the literature, although not coupled with a structural equations modeling approach, is Krizek (2003), which uses before-and-after residential relocation data to study the impact of the built environment on travel behavior.

[10] The extension of any of the modeling methods described here for use with longitudinal and quasi-longitudinal data would be very useful.

index equation (*i.e.,* a binary choice equation) for self-selection into a particular type of neighborhood, and the equation for travel behavior outcome variable of interest. The presence of significant correlation across the stochastic terms of the latent index equation and that of the travel behavior model indicate the presence of self-selection. Thus the idea is to capture the correlation (if any) between the residential location choice equation and the travel behavior choice equation to account for residential self-selection. In a recent residential self-selection study using this approach, Zhao and Kockelman (2008) treat the residential location as a binary (urban/suburban) variable and model the residential self-selection effects in the context of total household vehicle miles traveled(a continuous variable). A limitation of the binary sample selection models is the oversimplified representation of the complex residential location choice as a binary variable. The approach, in its original form, cannot be used to incorporate multinomial endogenous variables. To overcome this limitation, Dubin and Mcfadden (1984) extended Heckman's approach to include multinomial logit endogenous variables, and Lee (1983) proposed an alternative approach (to Dubin and McFadden, 1984) of transforming non-normal stochastic terms of multinomial logit for facilitating the correlation with the normal stochastic terms of the regression equation. Bhat (1998), for example, used the Lee's transformation approach to incorporate the endogeneity of multinomial discrete activity type choice variables in multiple continuous models of activity-travel behavior. However, no study in the residential-self selection literature appears to have used these <u>multinomial sample selection modeling</u> approaches[11]. One possible reason is that the approaches cannot be extended in a straightforward manner to incorporate the endogeneity of a full-scale residential location choice model with a large number of alternatives. Further, the approaches cannot be used in a straight forward fashion in the case of a multinomial discrete travel behavior variable.

[11] Greenwald (2003) used a slightly different multinomial sample selection approach in which he used the predicted choice probability of the chosen neighborhood type (from six residential neighborhood type alternatives) as an explanatory variable in travel behavior models with ratios of walking or transit trip times to driving trip times for different purposes. Mokhtarian and Cao (2008) reported on this model as "*the ability of this model to correct for selectivity bias is unclear*", because of the use of only the probability of only the chosen alternative to correct for self-selection bias.

The preceding discussion leads us to a search for approaches that can deal with the endogeneity of multinomial discrete endogenous residential choice variable (with a large number of alternatives) in multinomial discrete choice models of travel behavior. The joint choice models, which we label as multidimensional choice models, can do exactly that. In this approach, the multinomial residential location choice and the multinomial travel behavior choice are modeled jointly as a package (or a bundle) of choices. The advantage is the treatment of the residential location and travel behavior variables as multinomial discrete variables, which is of particular interest to practical travel demand modeling. Further, it is an "integrated" modeling approach, which is of interest for integrated land-use travel demand modeling. In addition to these reasons why the joint discrete choice modeling technique is of interest to this chapter, there are several salient aspects of the methods that will be discussed in detail in the next section devoted to joint discrete choice models.

2.2.2 Joint Choice Models (or Multidimensional Choice Models)

Joint choice models treat the residential location and travel behavior choices as a bundle (or a package) of choices. There are, in general, three types of joint discrete choice models: (1) multidimensional multinomial logit models, (2) multidimensional nested logit models and (3) mixed multidimensional choice models. Each of these modeling approaches is discussed below.

2.2.2.1 Multidimensional multinomial logit models

Multidimensional multinomial logit models treat each combination of the choice alternatives of various dimensions as a potential composite choice alternative. For example, Lerman (1976) estimated a model of household location, auto ownership and the mode to work choices, in which each potential location-housing-auto ownership-mode to work combination is treated as a composite alternative for a multinomial logit model. This study represents the first empirical example of modeling multiple choice processes, including residential location, within an integrated framework. The modeling approach is advantageous over the afore-mentioned approaches due to: (1) the ability to treat both residential location and travel behavior variables as multinomial discrete variables, (2) the integrated nature of the model that can potentially include multiple choice dimensions. While this is a useful and simple (in concept) method to model the multiple choice processes in an integrated fashion, the treatment of each possible combination of these decisions as a single composite choice alternative does not clearly recognize the mechanism of interdependencies among the individual decisions. Thus, the model does not clearly distinguish true impacts of the built environment from residential self-selection effects. Further, from a modeling perspective, the number of such composite alternatives can quickly explode with the increase in the number of alternatives in any of the choice dimensions; especially so with the residential location alternatives. The computation burden of estimation increases quickly with such an explosion in the number of potential composite alternatives to choose from. In this context, methods to decrease the number of residential alternatives, such as sampling of alternatives, can be used only with a multinomial logit model which fails to capture unobserved variations in behavior and flexible correlation patterns among the residential choice decisions (*i.e.,* a mixed logit model with random coefficients cannot be used).

2.2.2.2 Multidimensional Nested logit models

Ben-Akiva and Bowman (1995) proposed the use of the multidimensional nested logit modeling approach (see Daly and Zachary, 1978; McFadden, 1978; and Williams, 1977 for classic papers on the nested logit model) to analyze residential location and travel behavior choices in an integrated fashion, as a part of integrating residential location model with an activity-based modeling system. In this approach, the model system is similar to the above-discussed multidimensional multinomial logit model, in that a single indirect utility function specific to each combination of residential and travel behavior choice alternatives is specified and the joint probability of the chosen combination of alternatives (*i.e.,* the chosen composite alterative) is estimated. However, the nested logit structure allows for capturing correlation across the unobserved utility of subsets of alternatives in the choice set (see Train, 2003). This aspect of the nested logit structure imparts several advantages to the multidimensional nested logit approach, which are discussed below.

Advantages of the multidimensional nested logit approach

The nested logit approach provides an intuitive way of interpreting the estimated joint probability. Specifically, the joint probability expression can be written as a product of two standard multinomial logit probabilities, one conditioned upon the other: (1) The marginal probability of choosing a residential location alternative, and (2) The conditional probability of choosing a travel behavior alternative given that a particular residential location alternative is chosen. The choice dimension on which the other choice dimension is conditioned upon is labeled as the upper level choice (residential location choice, in this case), and the other choice dimension is labeled as the lower level choice (travel behavior choice, in this case).

A particularly appealing feature of the nested logit model is the elegant way to integrating the lower level and upper level choice processes in a unified modeling frame work. The expected value of the maximum utility of lower level (travel behavior, in this case) choices is utilized as a (pseudo)explanatory variable in the marginal probability expression of the upper level (residential location choice, in this case) choice to achieve integration. Further, and most importantly, the nested logit approach accommodates the correlations in the utility expressions of all combinations of upper and lower level alternatives for a specific upper level choice. That is, the similarity of alternatives due to common unobserved factors across (and specific to) the lower level choices (*i.e.,* the travel behavior choices) can be accounted for.

Due to the afore-mentioned attractive features, several studies in the literature have utilized nested logit modeling methodology to model residential location choice with other choice dimensions of interest (see, for example, Abraham and Hunt, 1997; and Ben-Akiva and dePalma, 1986). The nested logit approach is at the heart of are at the heart of a series of papers by Anas and his colleages (Anas and Duann, 1985; Anas, 1995; Anas, 1981) that form the basis of an integrated land use – transportation model. In the context of residential self-selection, Cervero and Duncan (2002) used a two dimensional nested logit modeling method to control for, and assess the extent of, residential self-selection effects in travel behavior choices. More recently Salon (2006) estimated a multidimensional nested logit model of residential location, car ownership, and commute mode choice using commuters' data from the New York City. There are, however, several disadvantages of this approach as discussed below.

<u>Disadvantages of the nested logit approachs</u>

<u>First</u>, while one can rewrite the joint probability expression (of the choice of a combination of residential location and travel behavior alternatives) as a product of a marginal probability and a conditional probability, the nested logit modeling structure does not necessarily imply any particular sequence or hierarchy of decisions. The model represents merely a particular correlation structure (*i.e.,* correlations across the utility expressions of sets of alternatives). The correlation structure is specific to the model specification rather than any behavioral sequence of the choice processes at hand. That is, different model specifications with a same empirical dataset can yield different correlation (or nesting) structures. This contradicts the possibility of nested logit model representing an underlying behavioral sequence in which the decisions are made. To understand this better, consider, for example, a model specification that includes all explanatory variables that could (statistically)significantly affect the choice processes in the observed utility expression. Such a model specification (or, to be precise, the deterministic utility specification) does not support any nesting structure because of the lack of significant (common)unobserved factors that might have lead to a nesting structure. This does not necessarily mean that the choices are made independently. Similarly, a different deterministic utility specification that supports a particular nesting structure does not imply any particular sequence of the individual decisions. In summary, it is important to note that the nested logit model does not necessarily imply any behavioral sequence of choices. Behaviorally, the choices could take place in any order or simultaneously.

<u>Second</u>, the nested logit model imposes a restriction that the coefficient on the expected value of the maximum utility of lower level choices should be between 0 and 1.

This coefficient is labeled as the log-sum parameter. Models with negative log-sum parameters are treated as inconsistent with the utility maximization theory and discarded. However, it is not uncommon in nested logit specifications to obtain log-sum parameters of value greater than 1. A related restriction, in the case of multidimensional nested logit models with more than two levels (for example, residential location, auto ownership, and mode choice), is that the log-sum parameters associated with the nests across different levels have to be in the ascending order from bottom to top. Ben-Akiva and Bowman (1995), for example, suggest such a *deeply nested logit* model to integrate various choice dimensions. It is difficult to estimate nested logit models with such restrictions on the log-sum parameters. In general, these requirements may be violated. Further, the non-negativity restriction on the log-sum parameter is not believed to be theoretically sound.

Third, as the number of choice dimensions increases, the difficulty of estimation (and the model complexity due to the number of potential nesting possibilities) increases. Nested logit models with multiple nests are usually estimated sequentially because simultaneous estimation can be cumbersome when there are multiple choice dimensions. The sequential estimation method underestimates the actual standard errors, leading to incorrect conclusions on the impact of the built environment attributes.

Fourth, in the context of self-selection, the magnitudes of common unobserved factors (such as attitudes and travel preferences) specific to lower level (travel behavior) alternatives is treated as the extent to which the self-selection (due to attitudes and travel preferences) into a particular residential neighborhood has not been captured by the observed factors that are included in the systematic utility of the model specification (Mokhtarian and Cao, 2008).[12] However, the self-selection effect is only a part of the alternative correlations estimated in the nested logit model. Thus, the estimated self-selection effect may be confounded with several other unobserved factors, leading to potential overestimation of the self-selection. Further, a common self-selection effect is estimated for all built environment attributes, without disentangling the extent of the

[12] Cervero and Duncan (2002) and Salon (2006) use slightly different ways to assess the self-selection effect using nested logit model. However, the drawbacks discussed here can be generalized to any nested logit modeling effort to assess residential self-selection.

effect with respect to each attribute. That is, it is not possible to understand, for example, the difference in extent of self-selection with respect to population density and bicycling facilities. Similarly, it is not possible to understand travel behavior preferences of which particular lower level alternative are contributing to self-selection. To understand this better, consider for example, the case of mode choice as a lower-level attribute. It is possible that unobserved bicycling inclination levels may lead to residential self-selection into bicycle friendly neighborhoods. Nested logit model does not allow us to understand that the preference for the bicycle mode leads to self-selection in the context of bicycling density. This is because of the muddling effect of all unobserved factors into a single log-sum variable and coefficient that represent the impact of all lower-level choice preferences on all upper-level choices.

Fifth, the nested logit modeling approach requires the choice alternatives to be represented as nominal outcome variables. Hence, the model cannot be used when the travel behavior choice variable is either continuous or ordinal discrete. It is important to note that several travel behavior choices are either continuous (e.g., activity durations, travel time and mileage *etc.*) or ordinal discrete (e.g.: car ownership, bicycle ownership, number of trips, *etc.*). Further, several choices can be categorized as discrete continuous; for example, activity participation and activity duration. In all these cases, the nested logit structure cannot be used to model residential self-selection.

2.2.2.3 Mixed multidimensional choice models

Mixed multidimensional choice modeling is a more general approach to jointly modeling various decision processes. In this approach, separate sub-models are formulated for different choice dimensions, and the models are econometrically joined together by the use of common stochastic terms. The stochastic terms (or the random coefficients) are assumed to be from a certain distribution, the parameters (or the properties, in a more general case) of which will be estimated along with the other parameters of the joint modeling systems.

The mixed multidimensional modeling approach avoids several of the afore-mentioned difficulties associated with the nested logit modeling approach. First, the mixed modeling approach can be used to separately represent and model hierarchical decisions and simultaneous decisions. When an outcome variable of interest is used as an explanatory variable in the equation of another variable, the model represents a hierarchical decision process. On the other hand, when any of the outcome variables is not used as an explanatory variable, the model represents a simultaneous decisions process. Thus alternative hypothesis of different decision hierarchies as well as simultaneity can be tested using the mixed modeling approach. It is not possible to test such hypotheses using the multidimensional multinomial logit and nested logit models. Second, the mixed modeling approach is free of log-sum related issues of the nested logit model. Third, in the context of integrated land-use travel demand modeling, as the number of choice dimensions increase, the nested logit approach becomes increasingly complex due to an increase in the large number of nesting possibilities, and the composite choice alternatives of the multidimensional MNL model start quickly exploding in number. On the other hand, as will be shown in Chapter 3, the mixed modeling of multidimensional choice processes is more straightforward.

Fourth, in the context of residential self-selection, the mixed modeling approach has several advantages over the nested logit modeling approach. We discuss these advantages using the pioneering study by Bhat and Guo (2007) as an example. Bhat and Guo (2007) employed the mixed modeling approach to jointly model residential location and auto ownership. They specified a latent utility equation for residential location choice, and a latent propensity equation for auto ownership, and joined the two equations econometrically by using common stochastic terms across the equations. To account for the impact of the built environment, they used built environment attributes in both residential location choice equation and auto ownership equation. They allowed for unobserved variations to each of the built environment attributes by using three types of random coefficients: (1) that are specific to (*i.e.,* affect only the) residential location choice, (2) that are specific to (*i.e.,* affect only the) auto ownership, and (3) that affect both auto ownership and residential location choice. The random coefficients that affect both auto ownership and residential location choice represent the unobserved factors (such as attitudes and unknown auto ownership preferences that affect residential location choice) that contribute to residential self-selection. Unlike in the case of nested logit model with potential confounding of several unobserved factors other than self-selection effects, Bhat and Guo (2007) minimized the confounding effect by separating the common (across residential location and auto ownership choices) unobserved factors from those unobserved factors that affect each of the choices independently. Further, they also specified separate random coefficients for each built environment attribute. Thus, the self-selection effect was specified to be a function of built environment attributes (and thus different for each built environment attribute), rather than a common self-selection effect across all attributes. If the travel behavior variable of interest is a nominal choice variable (for example, mode choice), as shown in Chapter 2, one can also separate the self-selection effects due to the preferences for different choice alternatives. For example, the approach allows the analyst to specify that the preference to bicycle mode (and not other modes) may lead to self-selection in the context of bicycling density (and not other built environment attributes). No other modeling approach allows this ability.

Fifth, unlike the nested logit and other approaches discussed before, the mixed modeling approach is generic and can be used to jointly model any type of outcome variables – continuous, nominal discrete, ordered discrete, or continuous discrete. The joint model estimated by Bhat and Guo (2007) is a special case with a joint nominal discrete (for residential location choice) and ordered discrete (for auto ownership) choice model structure. We extend this approach to jointly model residential location choice with different types of activity-travel behavior variables; multinomial mode choice variable in this chapter (*i.e.,* Chapter 2), and multiple discrete continuous activity participation and time-use variables in Chapter 4.

The above-mentioned advantages of the mixed modeling approach come with several disadvantages too. First, the mixed modeling approach requires simulation methods for model estimation. This is because the probability expressions of models with random coefficients do not usually take closed form expression. Resorting to simulation based estimation, leaves the possibility of inaccuracy in probability calculations and potential bias in the estimates. To minimize such bias, the simulation has to be carried out over a large number of draws to sufficiently cover the assumed distribution of the random coefficients. This may lead to large estimation times; especially so with the presence of large number of residential location choice alternatives. Second, the assumed distribution of random parameters can have a significant impact on model estimates (Train, 2003). Thus an incorrect assumption may lead to incorrect model estimates and policy implications. Third, in the context of residential self-selection, the Bhat and Guo (2007) method requires the use of different random coefficients to clearly disentangle the self-selection effect for each built environment attribute from other unobserved factors. This may lead to potential empirical identification issues associated with the use of a large number of random coefficients (see Walker, 2002).

2.2.3 Discussion

The preceding discussion provided an overview of the various modeling methods used to control for residential self-selection effects in models of travel behavior, and a detailed review of the multidimensional choice modeling methods that can be used to model residential location and travel behavior choices in an integrated fashion.

Multidimensional choice modeling methods provide the means to control for residential self-selection while modeling residential location choice and travel behavior choices in an integrated fashion. Among the three multidimensional modeling methods, this dissertation employs the mixed modeling method, because of the following reasons:

(1) While all the multidimensional modeling methods allow for modeling residential self-selection in multinomial discrete models of travel behavior (such as the MNL model), the mixed modeling method allows for the incorporation of self-selection with several types of model structures of travel behavior. Chapter 4, for example, employs this method to incorporate residential self-selection effects in a multiple discrete-continuous model of activity participation and time-use.

(2) It is more straightforward to extend the joint residential location and travel choice models to include more choice dimensions. Chapter 3, for example, employs this

method to develop an integrated model of residential location choice, auto and bicycle ownership, and commute mode choice decisions.

(3) Further, within the context of integrated modeling of multiple choice dimensions, mixed modeling allows the analyst to test for alternative decision hierarchies including simultaneity with no hierarchy.

(4) Mixed modeling allows the analyst to specify residential self-selection effects separately for each built environment attribute under consideration. Further, in the case of a multinomial nominal (MNL) discrete choice travel behavior variable, mixed modeling allows the analyst to specify residential self-selection effects separately for each travel behavior alternative under consideration. Thus, unlike in the case of the nested logit modeling approach, the self-selection estimates of a mixed model are less likely to be confounded with other unobserved effects.

(5) Further, as will be discussed in chapter 7, the mixed modeling approach allows us to easily estimate: (1) The short-term impacts (that are free from residential self-selection) of the built environment changes on travel behavior choices, (2) The long-term impacts of the built environment changes on residential location choices and travel choices, and (3) The extent of residential self-selection effects.

2.3 ADDRESSING RESIDENTIAL SELF-SELECTION IN MODE CHOICE: A REVIEW OF EMPIRICAL STUDIES

Numerous studies in the past have examined the impact of neighborhood attributes on mode choice. Several of them (for example, see Friedman *et al.,* 1994, Frank and Pivo, 1994, Ewing *et al.,* 1994, Handy, 1996, Cervero and Wu, 1997, Cervero and Kockelman, 1997, Kockelman, 1997, Badoe and Miller, 2000, Crane, 2000, Ewing and Cervero, 2001, Rajamani *et al.,* 2003, and Rodriguez and Joo 2004, and Zhang, 2004) reported a significant impact of neighborhood attributes in mode choice decisions. However, not all earlier studies have found such significant impacts of neighborhood attributes. For instance, Crane and Crepeau (1998) and Hess (2001) found no evidence that land use affects travel mode choice patterns. Kitamura *et al.* (1997) examined the effects of land use, demographic, and attitudinal variables on the proportion and number of trips by various modes, and found that attitudinal and demographic variables dominate neighborhood attributes in their effects on travel mode choice. Cervero (2002) studied mode choice behavior in Montgomery County, Maryland and found that the influences of urban design tend to be more modest than those of intensities and mixtures of land use on mode choice decisions.

Most of the studies listed above ignore residential sorting effects when estimating the impact of neighborhood characteristics on travel mode choice. The studies that explicitly accommodated for self-selection in multinomial mode choice analysis include, Cervero and Duncan (2002), Schwanen and Mokhtarian (2005b) Zhang (2006), and Salon (2006). There are also several studies that account for residential self-selection in models of travel attributes (such as travel frequency, and travel mileage) by various modes (see, for example, Boarnet and Sarmiento, 1998)[13]. This review, however, focuses on those studies that account for residential self-selection in *multinomial choice* models of mode choice.

Cervero and Duncan (2002) accommodated for residential self-selection by estimating a nested logit model for the joint choices of residing near a rail station (a binary residential choice dimension) and commuting by rail transit (a binary mode choice dimension). They compared the average odds of choosing rail over auto for those residing near rail stations with the average odds for those not residing near rail stations to assess the contribution of residential self-selection to the excess rail mode shares of residents near rail stations. Their analysis with the 2000 San Francisco Bay Area data suggests that residential self-selection accounts for about 40 percent of the excess rail mode shares of those living near rail stations. Schwanen and Mokhtarian (2005b) controlled for the impact of the dissonance (or mismatch) between neighborhood type preferences and the

[13] Also, see Hammond (2005) for a descriptive analysis of the interactions between residential and commute mode choices using a direction questioning approach described earlier in Section 2.1. In particular, he asked 90 respondents to describe their decision sequence of residential and mode choices, and studied an eight-person focus group to understand if individuals incorporated commute mode choice into their residential location choice.

actual residential neighborhoods while assessing the impact of neighborhood type on commute mode choice. They found that the mismatched urban residents (*i.e.,* suburban-oriented urban residents) to be less auto-oriented than the matched suburban residents (*i.e.,* suburban-oriented suburban residents. Based on this finding, they concluded that the neighborhood type had an autonomous effect on shaping (or, at the least, constraining) commute mode choice after controlling for residential preferences. Zhang (2006) indicated that residential self-selection can potentially have an impact at the mode choice set *consideration level* itself, rather than at the mode *choice level*. For example, individuals who are highly dependent on their automobiles (due to their attitudes or habituation) may self-select to live in auto oriented neighborhoods and simply refuse to consider transit modes of travel. For such individuals non-auto modes may not even exist as a choice alternative. To accommodate such effects, Zhang (2006) fist regressed the vehicle ownership variable on sociodemographic variables and urban residential location variable. Then he used the predicted vehicle ownership obtained from the regression as an explanatory variable (*i.e.,* an instrument variable) in the mode choice model with a probabilistic mode choice set generation mechanism.

In the first two of the three studies discussed so far (*i.e.,* Cervero and Duncan, 2002 and Schwanen and Mokhtarian 2005b), the residential location choice is reduced to a binary choice (near/far from rail-station, and urban/suburban). Zhang's study controls for self-selection while providing insights on auto-dependency (the extent to which auto is the only considered modal option), an important precursor to auto mode choice, which is usually ignored in mode choice analyses. However, even this study does not address the relationships between auto-dependency and a full scale residential location choice. In fact, Zhang's study does not explicitly model residential location choice.

As opposed to the three studies discussed above, a recently completed residential self-selection study by Salon (2006) jointly modeled residential location choice with auto ownership choice and walking-level choice, in which the residential location choice is treated as the choice of a census tract from all the residential census tracts in the study area, and the walking-level is treated as a trinomial choice (with zero walking-level, low walking-level, and high walking-level as the alternatives). Salon (2006) adopted the multidimensional multinomial modeling approach (*i.e.,* composite alternative multinomial logit modeling approach similar to that of Lerman, 1976) to jointly analyze the three choices. She also tested various nested logit model specifications and found the composite alternative multinomial logit modeling approach as the most appropriate. In the context of residential self-selection, she proposed an elasticity computation-based methodology to quantify the extent of self-selected neighborhoods on travel behavior choices using the multidimensional choice model. According to this methodology, the extent of residential self-selection is equal to the difference between the unconditional elasticities of walking-levels (or, travel behavior, in a general case) and those calculated as conditional on the chosen residential locations. Her analysis suggested that, in the context of New York City, the residential self-selection accounts for between one-third and one-half of the total effect of population density on walking-levels.

Salon's study is a notable contribution to the literature because it appears to be the first methodologically appropriate quantification of the extent of residential self-selection using an elasticity computation-based technique. However, the downside of this study relates to the use of composite alternative multinomial logit (or nested logit) modeling approach, which quickly leads to an explosion in the number of composite alternatives as the number of residential location alternatives increase. To overcome this problem, a randomly chosen sample of 10 alternatives (out of 2216 tracts) was used in the model estimation. The MNL model estimation with a random sample of alternatives is asymptotically equivalent to estimating the model with a full choice set (McFadden, 1978). While this is feasible in the context of traditional logit modeling frameworks, such sampling approaches do not allow the adoption of newer mixed logit modeling methods that accommodate random taste variations and flexible substitution or correlation patterns. Further, the composite alternative multinomial logit (or nested logit) modeling approach is associated with all other disadvantages discussed in section 2.2.2.2.

2.4 CONTRIBUTION AND ORGANIZATION OF THE CHAPTER

In this chapter, as opposed to the multidimensional multinomial (or nested) logit modeling approaches, we propose the use of mixed multidimensional choice modeling approach for various reasons discussed in Section 2.2.3. Specifically, we extend the mixed joint multinomial residential location choice and ordinal auto ownership choice model developed by Bhat and Guo (2007) to develop a joint model of multinomial residential location choice and multinomial mode choice. The joint model can control for residential self-selection effects to obtain the "true" effect of neighborhood attributes on mode choice.

From a methodological standpoint, the chapter presents a methodology for jointly modeling the relationship between two unordered multinomial discrete choice variables, residential location choice and commute mode choice. Another contribution of this chapter is the explanation of how the mixed joint model can be specified in two different ways, with identical model estimation procedure and results: (1) a simultaneous model of residential location choice and commute mode choice, which treats the two choices as a bundle of choices made simultaneously, and (2) a joint model of residential location choice and commute mode choice, with an in-built hierarchy that residential location choice precedes commute mode choice. Alternative forms of the latter way of specification can be used to test alternative hypotheses of choice making hierarchy.

The remainder of the chapter is organized as follows. Section 2.5 presents and discusses the econometric modeling methodology of a mixed joint residential location choice and mode choice model. Section 2.6 describes the data used for the empirical study of this chapter, and the empirical estimation results are presented in Section 2.7 together with a discussion of the interpretation of the findings. Finally, conclusions are presented in Section 2.8, along with a discussion of the directions for future expansions of this chapter's work.

2.5 ECONOMETRIC MODELING FRAMEWORK

2.5.1 Mathematical Formulation

The equation system for the joint residential location choice and commute mode choice model may be written as follows:

$$u_{hi}^{*} = \gamma_h' x_i + \varepsilon_{hi}, \text{ spatial unit } i \text{ chosen if } u_{hi}^{*} > \max_{\substack{d=1,2,\ldots I \\ d \neq i}} u_{hd}^{*} \qquad (2.1)$$

$$\mu_{q_h ij}^{*} = \alpha_{q_h j}' y_{q_h} + \beta_{q_h}' z_{q_h ij} + \delta_{hj}' x_i + \xi_{q_h ij}, \; mode\, j \; chosen \; if \; \mu_{q_h j}^{*} > \max_{\substack{k=1,2,\ldots J \\ k \neq j}} \mu_{q_h k}^{*}$$

The utility expressions in the equation system (1) can be rewritten as the following equation system (the reader is referred to Table 2.1 for a quick reference of the terms used in Equations 1 and 2):

$$u_{hi}^{*} = \sum_l \left(\gamma_l + \Lambda_l' w_{hl} + v_{hl} \right) x_{il} + \left(\sum_l \omega_{h1l} x_{il} + \sum_l \omega_{h2l} x_{il} \ldots + \sum_l \omega_{hjl} x_{il} \ldots + \sum_l \omega_{hKl} x_{il} \right) + \varepsilon_{hi}$$

(2.2)

$$\mu_{q_h ij}^{*} = \alpha_{q_h j}' y_{q_h} + \beta_{q_h}' z_{q_h ij} + \sum_l \left(\delta_{jl} + \Delta'_{jl} s_{hl} + \eta_{hjl} \right) x_{il} + \left(\sum_l \pm \omega_{hjl} x_{il} + \zeta_{q_h ij} \right)$$

The first equation in the equation systems (2.1) and (2.2) is the utility function for the choice of residence in which u_{hi}^{*} is the indirect utility that the household h derives from locating itself in spatial unit i, x_i is a vector of attributes corresponding to spatial unit i (x_i can potentially include non-built environment (non-BE) attributes such as racial composition, commute time, *etc.* and built environment (BE) attributes such as land-use mix, density, transit-accessibility, *etc.*), and γ_h in equation system (2.1) is a household-specific coefficient vector capturing the sensitivity to attributes in vector x_i. Elements of γ_h, γ_{hl} are parameterized in the first equation of the equation system (2.2) as:

$\gamma_{hl} = \left(\gamma_l + \Lambda_l' w_{hl} + v_{hl} + \left(\sum_l \omega_{h1l} x_{il} + \sum_l \omega_{h2l} x_{il} \ldots + \sum_l \omega_{hjl} x_{il} \ldots + \sum_l \omega_{hKl} x_{il} \right) \right)$, where w_{hl} is a vector of observed household-specific factors affecting sensitivity to the l^{th} attribute in vector x_i, and v_{hl} and $\omega_{hkl}x_{il}$ $(k = 1, 2, ..., j, ...K)$ are household-specific unobserved factors impacting the sensitivity of household h to the l^{th} attribute in vector x_i.

TABLE 2.1. Description of Terms Used in Equation Systems 2.1 and 2.2

h	Subscript for household h
q_h	Subscript for individual q from household h
i	Subscript for any residential spatial unit
j,k	subscripts for any mode
l	Subscript for l^{th} attribute
x_{il}	l^{th} neighborhood attribute of spatial unit i, used in residential utility
w_{hl}	vector of socio-demographic attributes affecting sensitivity to l^{th} neighborhood attribute
y_{q_h}	vector of socio-demographic attributes affecting modal utility
$z_{q_h ij}$	vector of commute level-of-service (LOS) attributes by mode j between the chosen residential and work locations
s_{hl}	vector of socio-demographic attributes affecting sensitivity to l^{th} neighborhood attribute (x_{il})
γ_l	sensitivity to l^{th} neighborhood attribute (x_{il}) in residential utility
δ_{jl}	sensitivity to l^{th} neighborhood attribute (x_{il}) in modal utility
Λ_l'	vector of coefficients on w_{hl}, indicating heterogeneous sensitivity to l^{th} neighborhood
Δ'_{jl}	vector of coefficients on s_{hl}, indicating heterogeneous sensitivity to l^{th} neighborhood
$\alpha'_{q_h j}$	vector of coefficients on socio-demographics (y_{q_h}) in modal utility
β'_{q_h}	vector of coefficients on LOS attributes ($z_{q_h ij}$) in modal utility. This vector can be
v_{hl}	residential error component capturing unobserved factors affecting the sensitivity to l^{th} neighborhood attribute (x_{il}) in residential utility
η_{hjl}	modal error component capturing unobserved factors affecting the sensitivity to l^{th} neighborhood attribute (x_{il}) in modal utility
ω_{hjl}	common error component capturing common unobserved factors affecting the sensitivity to l^{th} neighborhood attribute
r	Subscript for the **chosen** residential spatial unit
x_{rl}	l^{th} neighborhood attribute of **chosen** residential spatial unit (or location) r, used in modal utility
$z_{q_h rj}$	vector of commute level-of-service (LOS) attributes by mode j between the **chosen** residential and work locations

v_{hl} includes only those household-specific unobserved factors that influence sensitivity to residential choice, while each ω_{hkl} (k = 1, 2,…j, …, K) includes only those household-specific unobserved factors that influence both residential choice and the choice of commute mode k. Finally, ε_{hi} is an idiosyncratic error term assumed to be identically and independently extreme-value distributed across spatial alternatives i and households h.

The second equation in equation systems (2.1) and (2.2) is the utility function for the choice of commute mode in which $\mu^*_{q_h ij}$ is the indirect utility that an individual q from household h residing in spatial unit i associates with commute mode j. In the explanatory variables, y_{q_h} is a vector of attributes that includes non-spatial determinants of modal utilities such as individual and household level socio-demographics (for example, household and personal income, age, gender, *etc.*), $z_{q_h ij}$ is a vector of level-of-service (LOS) attributes faced by the individual q of household h between the residential location i and employment location[14] by mode j (for example, travel time, travel cost, *etc.*), and x_i is a vector of attributes corresponding to the residential spatial unit i (for example, BE attributes such as land-use mix, density, *etc.*, and household level non-BE attributes such as the total commute time of all commuters in the household).

In the coefficient vectors in the second equation of the equation systems (2.1) and (2.2), $\alpha_{q_h j}$ represents the impact of socio-demographics on the utility of mode j, β_{q_h} is a vector of response sensitivities to the LOS attributes in $z_{q_h ij}$, and δ_{hj} is a household-specific coefficient vector capturing the impact of BE and non-BE attributes (in vector x_r) of residential spatial unit i on the utility of mode j. The elements (indexed by l) of δ_{hj} are parameterized in the second equation of the equation system (2.2) as: $\delta_{hjl} = (\delta_{jl} + \Delta'_{jl} s_{hl} + \eta_{hjl})$, where s_{hl} is a vector of observed household-specific factors influencing the sensitivity to l^{th} attribute in x_i, Δ_{jl} is the corresponding vector of

[14] It is assumed in this model formulation that the employment location is known *a priori*.

coefficients, and η_{hjl} is a term capturing the impact of household-specific unobserved factors on the sensitivity to l^{th} attribute in x_i. Finally, $\xi_{q_h ij}$ of the equation system (2.1) is an error term that is partitioned into two components in the equation system (2.2) as: $\sum_l (\pm\omega_{hjl})x_{il} + \zeta_{q_h ij}$. The $\pm\omega_{hjl}x_{il}$ terms are the common error components in residential choice and mode choice (for mode *j*), while $\zeta_{q_h ij}$ is an idiosyncratic term assumed to be identically and independently (IID) logistic distributed across individuals and modal alternatives.

2.5.2 Intuitive Discussion of Model Structure

In the equation system (2.2), the self-selection of households into certain neighborhoods (that explains the endogeneity in the effect of neighborhood specific BE and non-BE attributes on commute mode choice) is captured by controlling for both observed and unobserved factors that impact residential location and commute mode choice. The explanation is as follows.

First, the model formulation controls for the effect of systematic/observed socio-demographic differences among individuals in their mode choice decisions. Suppose households with high income avoid residing in high density neighborhoods. This can be reflected by including income as a variable in the w_{hl} vector in the residential choice equation. High income households are also likely to own more cars and the individuals belonging to those households are more likely to choose auto as their commute mode choice. The residential sorting based on income can then be controlled for when evaluating the effect of the BE attribute "density" on commute mode choice by including income as a variable in the y_{q_h} vector in the mode choice equation. Ignoring such residential sorting effects due to observed demographics can lead to an artificial inflation of the neighborhood attribute effects in mode choice decisions.

Second, the model formulation controls for unobserved attributes (such as attitudes/perceptions, and environmental considerations) that may influence both residential choice and commute mode choice. For example, households with individuals that are environment-conscious and auto-disinclined may locate themselves into neighborhoods that are conducive to the use of non-motorized forms of transport so that they may walk or bike to work. Such common unobserved preferences are captured in the ω_{hjl} terms of the residential choice utility equations and the non-motorized modal utility equations, respectively. These common unobserved factors cause the endogeneity in the effect of corresponding BE and non-BE attributes in the commute mode choice model, and give rise to correlation in the error components across the residential location and mode choice models leading to the joint nature of the model structure.

The '$\pm$' in front of the $\omega_{hjl}x_{il}$ terms in the mode choice equation indicates that the impact of common unobserved factors in moderating the influence of the characteristics represented by x_{il} across the residential choice and mode choice equations may be in the same or opposite directions, respectively (called as positive or negative correlation, respectively). If the sign is '+', it implies that the unobserved factors that increase (decrease) the individuals' (households) preference to the characteristic represented by x_{il} in residential location choice decisions also increase (decrease) their preference for commute mode j, while a '–' sign implies that the unobserved factors that increase (decrease) the individuals' preference to the characteristic captured by x_{il} in residential location choice decisions decrease (increase) their preference for commute mode j.

If the x_{il} measures are defined in the context of promoting smart growth and neo-urbanism concepts (such as high density and increased land use diversity) to promote non-motorized travel to work, then there may be an expectation that the appropriate sign in front of the $\omega_{hjl}x_{il}$ term in non-motorized modal utility equations should be positive. Through the model formulation adopted in this Chapter, it is possible to test which one of the two signs is appropriate. A positive sign suggests that households who have an intrinsic preference for neo-urbanist neighborhoods also have a higher preference for non-motorized modes of transport (due to unobserved attributes such as auto-disinclination). Ignoring these $\omega_{hjl}x_{il}$ terms while estimating the mode choice utility equations leads to an artificial inflation of the positive sign on the corresponding neo-urbanist BE attributes (*i.e.,* an artificial inflation of the positive sign on the δ_{jl} terms in the non-motorized modal utility equations).

If x_{il} represents an attribute such as total commute time of all individuals in the household, the anticipated sign in front of the $\omega_{hjl}x_{il}$ term in auto modal utility equations could be either positive or negative. A negative sign indicates that the unobserved factors (such as attitudes/perceptions towards traveling and spending time on the road) that increase (decrease) individuals' sensitivity to total commute time in residential location decisions also increase (decrease) their preference for the relatively faster auto modes. On the other hand, a positive sign indicates the presence of unobserved factors affecting residential location choice that contribute to individuals/households increasing their total commute time and therefore becoming more auto-oriented in their commute mode choice. For example, one may consider such factors as crime, school quality, aesthetic appeal of neighborhood, neighborhood amenities, and perceptions of the prestige associated with living in a certain neighborhood. Although individuals/households would like to minimize their total commute time index, simply doing so may result in their locating in less-desirable residential neighborhoods. These unobserved factors then lead to individuals/households living in neighborhoods that increase their total commute time index and make them more auto-oriented.

In summary, the model formulation explicitly considers residential sorting effects that may be traced to observed socio-demographics, and unobserved attitudinal variables and personal lifestyle preferences. Further, the residential sorting effects can be accommodated separately for each built environment attribute under consideration by estimating common (to residential location and mode choice decisions) unobserved factors that are specific to each built environment attribute under consideration. For example, as discussed above, one can estimate separate common unobserved factors that are specific to 'density' and 'total household level commute time', and any other attribute. Thus, the fact that residential self-selection effect is specific to the built environment attribute under consideration is explicitly recognized in this method. Other existing modeling methods, on the other hand, estimate a common self-selection effect common to all built environment attributes that could potentially lead to biased estimation of the self-selection effect with respect to individual built environment attributes as well as the impacts of the built environment on travel behavior.

2.5.3 Model Estimation

The parameters to be estimated in the equation system (2.2) include the α and β vectors, the $\gamma_l, \delta_l, \Lambda_l$, and Δ_l vectors, and the variances of v_{hl} (= σ^2_{vl}), η_{hjl} (= $\sigma^2_{\eta l}$), and ω_{hjl} (= $\sigma^2_{\omega jl}$) for those BE and non-BE attributes with random taste heterogeneity. In a general case, where $\sigma^2_{vl} \neq 0$, $\sigma^2_{\eta l} \neq 0$, and $\sigma^2_{\omega jl} \neq 0$ for each of the BE and non-BE attributes (*i.e.,* for each l), there may be unobserved factors that affect the sensitivity to each of the BE and non-BE attributes, which are specific to residential location choice, mode choice, as well as common to both residential location and mode choices. However, in specific empirical cases, it is to be noted that the random taste heterogeneity to a particular attribute l may occur only in residential choice ($\sigma^2_{vl} \neq 0$, $\sigma^2_{\eta l} = 0$, $\sigma^2_{\omega jl} = 0$), only in some of the modal utilities ($\sigma^2_{vl} = 0$, $\sigma^2_{\eta l} \neq 0$, $\sigma^2_{\omega jl} = 0$), independently in residential choice and mode choice ($\sigma^2_{vl} \neq 0$, $\sigma^2_{\eta l} \neq 0$, $\sigma^2_{\omega jl} = 0$), or as combinations of the above patterns with a common effect on both residential choice and mode choice ($\sigma^2_{\omega jl} \neq 0$). Also, there may not be any random heterogeneity for some or all of the attributes in either of the residential choice and mode choice models ($\sigma^2_{vl} = 0$, $\sigma^2_{\eta l} = 0$, $\sigma^2_{\omega jl} = 0$).

Let Ω represent a vector that includes all the parameters to be estimated, and let $\Omega_{-\sigma}$ represent a vector of all parameters except the variance terms. Also, let c_h be a vector that stacks the v_{hl}, η_{hjl}, and ω_{hjl} terms across all BE and non-BE attributes and let Σ be a corresponding vector of standard errors. Define $a_{hi} = 1$ if household *h* resides in spatial unit i and 0 otherwise. Similarly, define $b_{q_h j} = 1$ if an individual q_h chooses the commute mode j and 0 otherwise. Then, the likelihood function for a given value of $\Omega_{-\sigma}$ and c_h may be written for an individual q_h as:

$$L_{q_h}(\Omega_{-\sigma}) \mid c_h = \prod_{i=1}^{I}\prod_{j=1}^{K}\left\{\left[\frac{\exp(\gamma_h' x_i)}{\sum_k \exp(\gamma_h' x_k)}\right]\left[\frac{\exp(\alpha'_{q_h j} y_{q_h} + \beta'_{q_h} z_{q_h ij} + \delta'_{hj} x_i)}{\sum_k \exp(\alpha'_{q_h j} y_{q_h} + \beta'_{q_h} z_{q_h ij} + \delta'_{hj} x_i)}\right]\right\}^{a_{hi} b_{q_h j}} \tag{2.3}$$

Finally, the unconditional likelihood function can be computed for individual q_h as:

$$L_{q_h}(\Omega) = \int_{c_h} \left(L_{q_h}(\Omega_{-\sigma}) \mid c_h\right) dF(c_h \mid \Sigma), \tag{2.4}$$

where F is the multidimensional cumulative normal distribution. The log-likelihood function can be written as: $L(\Omega) = \sum_{q_h} \ln L_{q_h}(\Omega)$. Simulation techniques are applied to approximate the multidimensional integral in Equation (2.4), and maximize the resulting simulated log-likelihood function. Specifically, the scrambled Halton sequence (see Bhat, 2003) is used to draw realizations of c_h from its population normal distribution. In the current study, 125 realizations of c_h were used to obtain stable estimation results.

2.5.4 An Alternative Model Specification

The joint residential location and mode choice model can be specified in a different fashion, however, with the same log-likelihood function, model estimation procedure, and model parameters. This section presents the alternative specification and highlights the subtle differences in the interpretation of the two specifications.

The joint model can be specified in a different fashion, as below (the reader is referred to Table 2.1 for a quick reference of the terms used in Equations 2.5 and 2.6):

$$u^*_{hi} = \gamma'_h x_i + \varepsilon_{hi}, \text{ spatial unit } i \text{ chosen if } u^*_{hi} > \max_{\substack{d=1,2,\ldots I \\ d \neq i}} u^*_{hd} \tag{2.5}$$

$$\mu^*_{q_h rj} = \alpha'_{q_h j} y_{q_h} + \beta'_{q_h} z_{q_h rj} + \delta'_{hj} x_r + \xi_{q_h rj}, \ \textit{mode } j \textit{ chosen if } \mu^*_{q_h j} > \max_{\substack{k=1,2,\ldots J \\ k \neq j}} \mu^*_{q_h k}$$

The utility expressions in the equation system (2.5) can be rewritten as the following equation system:

$$u_{hi}^{*} = \sum_{l}\left(\gamma_l + \Lambda_l^{'} w_{hl} + v_{hl}\right) x_{il} + \left(\sum_{l}\omega_{h1l}x_{il} + \sum_{l}\omega_{h2l}x_{il} \ldots + \sum_{l}\omega_{hjl}x_{il} \ldots + \sum_{l}\omega_{hKl}x_{il}\right) + \varepsilon_{hi}$$

(2.6)

$$\mu_{q_h rj}^{*} = \alpha_{q_h j}^{'} y_{q_h} + \beta_{q_h}^{'} z_{q_h rj} + \sum_{l}\left(\delta_{jl} + \Delta_{jl}^{\prime} s_{hl} + \eta_{hjl}\right) x_{rl} + \left(\sum_{l} \pm \omega_{hjl} x_{rl} + \zeta_{q_h rj}\right)$$

The reader will note that the above Equation systems are slightly modified versions of the Equation systems (2.1) and (2.2) and discussed in Section 2.5.1. The modification is that the subscript '*i*' in the second Equations of the Equation systems (2.1) and (2.2) are replaced by the Subscript '*r*' for the "chosen" residential spatial unit. That is, the second equations (*i.e.,* the mode choice utility equations) are now conditional upon the "chosen" residential spatial unit. That is, the commute mode choice is modeled conditional upon the residential location decisions. In such a case, the log-likelihood of the joint model system for an individual q_h is:

$$L_{q_h}(\Omega_{-\sigma}) \mid c_h = \prod_{i=1}^{I}\prod_{j=1}^{K} \left\{ \left[\frac{\exp(\gamma_h^{\prime} x_i)}{\sum_{k}\exp(\gamma_h^{\prime} x_k)}\right]^{a_{hi}} \left[\frac{\exp(\alpha_{q_h j}^{'} y_{q_h} + \beta_{q_h}^{'} z_{q_h rj} + \delta_{hj}^{'} x_r)}{\sum_{k}\exp(\alpha_{q_h j}^{'} y_{q_h} + \beta_{q_h}^{'} z_{q_h rj} + \delta_{hj}^{'} x_r)}\right]^{b_{q_h j}} \right\} \quad (2.7)$$

The reader will note that the log-likelihood expression in Equation (2.6) can be simplified to, and hence is equivalent to, the above log-likelihood expression. In essence, the log-likelihood expressions evaluate the probabilities of each decision maker's "chosen" combination of residential location and commute mode choice decisions. That is, two different model specifications yield the same log-likelihood expressions. Consequently, both the model specifications have the same model estimation procedure as well as result in the same model estimates.[15] However, the two specifications can be interpreted differently. The interpretation issues are discussed in the next section.

---

[15] This discussion leads to another important advantage of this method over other multidimensional choice models. In this method, for model estimation purposes, unlike in the case of composite alternative models

2.5.5 Model Interpretation: Simultaneity and Hierarchy

The model specification in Equation systems (2.1) and (2.2) does not assume any particular hierarchy. As it can be observed from the corresponding log-likelihood expression in Equation (2.3), this model structure implies enumerating all possible combinations of residential location alternatives and mode choice alternatives. Thus, this model specification is similar in interpretation to the multidimensional MNL and NL models that no particular hierarchy of decisions is implied. This model simply implies a simultaneous choice behavior that the decisions are made jointly as a package (or a bundle) of choices.

The model specification in Equation systems (2.5) and (2.6) also implies that the residential location choice and mode choice decisions are made jointly as a package of choices. However, within this joint decision making behavior, there exists an in-built hierarchy that the residential location choice decisions affect commute mode choice decisions. In other words, the residential location choice decision is treated as the leading decision impacting commute mode choice decisions. This is especially the case when the chosen residential alternative specific indicators (*i.e.,* the endogenous outcome variable(s)[16]; for example, the County or Central business district dummy indicators) are used as explanatory variables in the mode choice model. In order to further clarify this point, consider another model structure in which the chosen mode choice is used as an explanatory variable in the residential location choice model, rather than using the characteristics of the chosen residential location in the mode choice model. That is, the residential location choice decisions are modeled conditional upon the mode choice decisions. This model structure would imply commute mode choice decisions as leading

---

(*i.e.,* the multidimensional MNL models), the analyst need not enumerate each combination of residential location and mode choice alternatives. Rather, the multidimensional choice process is treated as comprising separate choice processes, which are joined together by using common random parameters. This greatly reduces the computation burden that arises due to considering all possible composite choice alternatives.

[16] In a full-scale residential location model with a large number of spatial location alternatives, no labels are given to the alternatives. Hence, the alternative specific indicators are not generally used. In such cases, county and other location specific dummies represent groups of residential location alternatives. However, in the cases where the residential location choice is defined using a limited number of labeled alternatives (such as urban/sub-urban, traditional/neo-traditional neighborhoods), the alternative specific indicators (urban/suburban *etc.*) can be used in the mode choice model. Such a case is presented in Chapter 4.

and affecting the residential location choice. The true nature of the residential location and mode choice decisions can be expected to be any one of the above mentioned patterns. However, considering the long-term nature of the residential location choice decision (relative to a mode choice decision), it is reasonable to assume a hierarchy (*i.e.,* a causal structure) where residential location choice affects commute mode choice. Along with this hierarchy, households and individuals may locate (or self-select) themselves in built environments (*i.e.,* residential locations) that are consistent with their socio-demographics, lifestyle preferences, attitudes and values. This self-selection phenomenon leads to endogeneity of the residential location choice and suggests that residential location choice and mode choice decisions are made jointly (as a package), but with a certain decision hierarchy in which mode choice is conditional upon residential location choice. Thus, by including observed and unobserved factors that affect both residential location choice and mode choice decisions, the residential self-selection phenomenon (and hence the behaviorally joint nature of the decisions) is accounted for. Within the context of unobserved factors, the presence of common unobserved factors leads to an econometrically joint model structure.

As it can be observed from this discussion, the mixed multidimensional modeling approach can be used to test alternative hypotheses of hierarchy and pure simultaneity. The other multidimensional modeling approaches (such as the multidimensional MNL and NL approaches), on the other hand, cannot be used to test alternative hypotheses of hierarchy as they assume a pure simultaneous choice behavior.

2.6 DATA

2.6.1 Data Sources

The primary data source used in the analysis is the 2000 San Francisco Bay Area Travel Survey (BATS), designed and administered by MORPACE International, Inc. for the Bay Area Metropolitan Transportation Commission (see MORPACE International Inc., 2002 for details on survey design, sampling, and administration procedures). In addition to the activity survey, six other data sets associated with the San Francisco Bay area were used in the current analysis: land-use/demographic coverage data, zone-to-zone network level-of-service (LOS) data, a GIS layer of bicycle facilities, the Census 2000 Tiger files, census demographic data, and Public Use Microdata Sample (PUMS) data. The following section provides a description of the estimation sample.

2.6.2 Estimation Sample

The geographic area of study in this research is the Alameda County in the San Francisco bay area with 233 transport analysis zones. The residential choice of households and commute mode choice of individuals within this county constitute the focus of analysis for this paper. After extracting the Alameda County households from the survey sample and merging the various secondary data sources, the final sample for analysis comprised 1,878 individuals from 1,447 households. The household characteristics of the estimation sample are provided in Table 2.2 (see next page).

The average household size is about 2.5 persons per household with nearly a quarter of the households reporting household sizes of four or more persons. Nearly one-third of the households report having an individual less than 18 years of age in the household. The median household income is rather high with about 50 percent of the households falling into the fourth and highest income quartile. On average, households reported a little over two cars per household with less than two percent of the households having zero cars. On average, the ratio of vehicles to licensed drivers is greater than one, generally indicating a high level of auto availability. A little over one-third of the households have zero bicycles; at the same time, about one-quarter of the households have three or more bicycles. About 70 percent of the households own the home in which they reside, generally indicating a rather high level of home ownership within the survey sample used in this paper. With respect to the type of residential location, about one quarter live in potentially higher density CBD and urban environments while 70 percent live in potentially lower density suburban locations. A very small percentage lives in rural locations.

Table 2.2 Household Characteristics in the Alameda County Sample

Characteristic	Sample Shares
Household Size	2.53
1	22.8%
2	36.2%
3	16.3%
≥ 4	24.7%
Presence of Kids of age <18 years	32.6%
Household Income	$83,510
Low (in quartile 1 and 2)	18.1%
Medium (in quartile 3)	30.9%
High (in quartile 4)	51%
Vehicle ownership	2.14
0	1.7%
1	26.5%
2	43.2%
3	18.9%
≥ 4	9.7%
#vehicles/#licenses	1.13
Bicycle ownership	1.6
0	34.1%
1	21.1%
2	20.1%
≥ 3	24.7%
Household Tenure	
Own house	69.8%
Housing Type	
Single Detached House	69.4%
Duplex/Apartment	28.2%
Other	2.4%
Residential Location	
CBD	1.3
Urban	24.7
Suburban	71.6
Rural	2.4

Table 2.3 presents person characteristics for the estimation sample used in this paper. The person sample includes only commuters who are employed outside the home. The average age of the sample persons is 43 years and more than 50 percent of the persons are male. More than 85 percent of the individuals are employed full time. The degree of work flexibility is quite high with about 60 percent of the individuals indicating that they have at least some (moderate) flexibility with respect to their work arrangement. A vast majority (98 percent) is licensed to drive and commute by auto; more than 80 percent drive alone to work. Less than one percent use transit, about 11 percent carpool to work either as a driver or passenger, and about 6.5 percent use non-motorized modes (bike/walk) to commute to work.

Table 2.3 Person Characteristics in the Alameda County Sample

Characteristic	Sample Shares
Age (in years)	42.96
18-29	17.1%
30-65	79.60%
66	3.30%
Gender	
Male	56.3%
Employment Stats	
Full-time	86.2%
Part-time	13.8%
Work Flexibility	
Not Flexible	40.7%
Moderately Flexible	30.2%
Flexible	29.1%
Driver’s License	
Licensed	97.9%
Commute Mode Share	
Auto - Drive Alone	82.1%
Auto - Drive with Passenger	4.7%
Auto – Passenger	6.0%
Walk	3.8%
Bike	2.8%
Transit	0.7%

Table 2.4 offers a description of the land use characteristics associated with the Alameda county sample used for model estimation. About 65 percent of the 233 zones are suburban locations and another 30 percent are urban locations. The percent of zones that are CBD or rural in nature is quite small at about 3 percent each. As expected, the number of zones accessible by transit is greatest for CBD and urban locations and lowest for rural locations. Consistent with land use definitions, household density, employment density, shopping accessibility, and employment accessibility are greatest for CBD and urban locations and lowest for rural locations. On the other hand, the recreational accessibility is greatest for rural locations. Similarly, the street block density, bikeway density, and land use mix show a decreasing trend with the highest values associated with CBD and urban locations and lower values associated with suburban and rural locations.

Table 2.4 Land-use Characteristics in the Alameda County Sample

	Average value by Area type			
Characteristic	**CBD** (2.8%)	**Urban** (30.2%)	**Suburban** (64.2%)	**Rural** (2.8%)
Number of destination zones accessible by transit	774.50	671.20	551.85	272.50
Household density (no. households per acre)	8.86	8.13	3.48	0.60
Employment density in zone (no. jobs per acre)	107.63	38.91	19.91	21.68
Shopping accessibility	33.67	25.14	22.12	18.68
Recreational accessibility	5.85	6.43	7.75	8.15
Employment accessibility	10.14	7.10	6.41	5.64
Street block density (no. blocks per square mile)	243.47	172.03	78.70	24.67
Bikeway density (miles per square mile)	7.67	4.88	3.59	0.52

Table 2.5 offers a summary of home-to-work level of service characteristics for the estimation sample used in this paper. As expected, the average travel times for the auto modes are about 20-25 minutes and the auto mode is the fastest among all modes. The average travel time is about six hours for the walk mode, more than one hour for the bicycle mode, and about 45 minutes for the transit mode. On average, the drive alone mode and the transit mode are nearly equal in cost while the auto-shared ride mode is considerably less expensive. The reader will note here that the descriptive statistics of the travel time and travel cost represent the travel times and costs faced by all of the

individuals in the sample regardless of the actual mode chosen. The average walk and bike times for individuals who actually chose those modes are about one hour and 40 minutes, respectively.

Table 2.5 Home-to-Work Level-of-Service Characteristics

Characteristic	Sample Average	Standard Deviation
Travel Times (in minutes for 1878 persons)		
Auto - Drive alone	19.67	15.10
Auto - Shared ride	24.02	12.59
Walk	293.21	346.50
Bike	73.30	86.62
Transit	47.98	36.12
Travel Costs (in dollars for 1878 persons)		
Auto - Drive alone	1.16	1.06
Auto - Shared ride	0.30	0.25
Transit	1.23	0.99

Note: Walk and bike times were calculated assuming speeds of 12mph and 3mph, respectively. The walk and bike costs are assumed to be associated with commuting cost.

All of these variables and attributes were merged with the travel survey sample data to form a comprehensive land use – travel behavior – network level of service database suitable for analyzing the relationships between residential location choice and commute mode choice.

2.7 MODEL ESTIMATION RESULTS

This section provides a description of the model estimation results. The model system is estimated as a joint choice model including both residential location choice and commute mode choice dimensions. All 233 zones are considered to be alternatives in the residential location choice set. The commute mode choice set definition accounts for modal availability at the individual/household level. A household must own an automobile and an individual must have a driver's license for the auto drive (drive alone and drive with passenger) modes to be available in the choice set. The auto-passenger mode choice is available to all individuals as are the bike and walk modes. The transit mode is included in the choice set based on transit availability (between residential and work zones) as specified in the network level of service files.

Table 2.6 presents estimation results for the residential location choice component of the joint model. In general, the results are found to be plausible and consistent with expectations. The first variable in Table 2.6, logarithm of the number of households in a zone is a surrogate measure for the number of housing opportunities in a zone. As expected, a positive coefficient on this variable indicates that households are more likely to locate in zones with larger number of housing opportunities. Similarly, households are more likely to locate in zones with high household density. However, it is found that seniors are less likely to locate in zones of high density as evidenced by the negative coefficient associated with the interaction term. As expected high employment density zones are less likely to be chosen for residential location, except for lower income households who may be compelled to choose lower cost housing in such locations. Also, households desiring to live in single family detached housing units are more likely to locate in zones with a higher fraction of such a housing stock. The land use mix measure is negatively associated with residential location choice; this suggests that households are more prone to live in zones that are rather homogeneous in nature. This finding may also be an artifact of both zoning policies and zone definition strategies. Zoning policies may often dictate that land uses be segregated and traffic analysis zones themselves are often defined based on homogeneity of land uses. As a result, the likelihood of a household being located in a mixed land use zone is potentially going to be small simply because such zones are few and far between. Rather surprisingly (but consistent with the findings in Bhat and Guo, 2007), the fraction of residential land area is negatively associated with residential location choice. A higher recreational accessibility is associated with a greater likelihood of locating residence in a particular zone.

TABLE 2.6 Estimation Results of the Residential Location Choice Component of the Joint Residential Location and Mode Choice Model

Variables	Parameter	t-stat
Zonal size and density measures (including demographic interactions)		
Logarithm of number of households in zone (x10^{-1})	9.803	15.02
Household density (#households per acre x 10^{-1})	0.351	3.70
Interacted with presence of seniors in household	-0.652	-1.93
Employment density (#employment per acre x 10^{-1})	-0.211	-2.89
Interacted with household income in the lowest quartile	0.196	2.38
Zonal land-use structure variables (including demographic interactions)		
Fraction of residential land area	-0.813	-5.70
Fraction of single family housing interacted with household living in single family detached housing	2.298	13.03
Land-use mix	-0.305	-2.07
Regional accessibility measures (including demographic interactions)		
Recreation accessibility x 10^{-2} (by auto mode)	0.425	6.35
Commute-related variables (including demographic interactions)		
Total drive commute time of all commuters in household (minutes x 10^{-2})	-11.472	-24.28
Standard deviation of the error term in residential location model	5.809	11.82
Standard deviation of the error term common to residential location and mode choice models (negative correlation between the error terms)	0.859	1.53
Total drive commute cost of all commuters in household (dollars x 10^{-1})	0	fixed
Interacted with household income in the lowest quartile	-4.600	-2.47
Local transportation network measures (including demographic interactions)		
Street block density (number of block per square mile x 10^{-2})	0.163	1.47
Interacted with number of vehicles per number of licenses in household	-3.526	-3.34
Bicycle facility density (miles per square mile x 10^{-1})	0.251	2.54
Interacted with number of bicycles in the household	0.864	2.34
Availability of transit service to work zone	0.570	2.71
Transit access time to stop (minutes x 10^{-1})	-0.425	-5.25
Zonal demographics and housing cost (including demographic interactions)		
Absolute difference between zonal median income and household income (\$ x 10^{-5})	-2.077	-11.59
Absolute difference between zonal average household size and household size	-0.349	-5.05
Average of median housing value (\$ x 10^{-5})	-0.182	-7.01

TABLE 2.6 (Continued) Estimation Results of the Residential Location Choice Component of the Joint Residential Location and Mode Choice Model

Variables	Parameter	t-stat
Zonal ethnic composition measure		
Fraction of Caucasian population interacted with Caucasian dummy variable	2.836	13.82
Fraction of African-American population interacted with African-American dummy variable	2.736	5.18
Fraction of Hispanic population interacted with Hispanic dummy variable	2.199	4.47

The total drive commute time for the household serves as a surrogate measure of the overall location of the household *vis-à-vis* the work locations of the commuters in the household (assuming work locations are exogenous). Thus, this variable may be treated as an overall commute time index for the household. As expected, households attempt to locate such that this commute time index is reduced as evidenced by the negative coefficient associated with this variable. The total drive commute cost variable is found to be significant for households in the lowest quartile suggesting that lower income households are more sensitive to commuting costs than other households.

Within the context of the commute time index, the standard deviation of its random coefficient specific to the residential location model is highly significant with a test statistic value of 11.82, indicating significant population heterogeneity in the sensitivity to commute time index in residential location decisions. It is also found that there are common unobserved factors affecting both residential location choice and auto mode (all auto modes) choice in the context of commute time index; the corresponding error components are found to be negatively correlated. The standard error of this negative error correlation is found to be marginally significant with a test statistic value of 1.53. The presence of this correlation suggests that it is very important to model residential location choice and mode choice in a simultaneous equations framework because there are unobserved factors related to commute time that affect both of these choice dimensions simultaneously. In this particular instance, the interpretation of the negative sign on the correlation is as follows. The unobserved factors that increase (decrease) the sensitivity of individuals/households to total commute time index in residential location decisions, also make them more (less) oriented towards the relatively faster auto modes. For example, one may consider such factors as individuals' attitudes/perceptions towards traveling and spending time on the road that could contribute to higher (lower) sensitivity to total commute time index in residential location decisions, as well as higher (lower) preference to auto modes. Not accounting for such endogeneity could potentially lead to biased estimates of the impact of total commute time index in the commute mode choice model.

Within the context of common unobserved factors, only the total drive commute time variable has common random coefficients representing residential self-selection effects due to unobserved factors. It is possible that there may be important but omitted neighborhood variables (due to unavailability in the data) that might have resulted in significant unobserved residential self-selection effects associated with them. Further, an analysis in a different context may indicate the presence of unobserved residential self-selection effects (and hence an econometrically joint nature of the residential location and mode choice model) and/or random heterogeneity in sensitivity with respect to several neighborhood attributes. In any case, even with a comprehensive set of neighborhood attributes, it is important to estimate the joint model to test for the presence of unobserved residential sorting effects.

The remaining variables in Table 2.6 offer plausible interpretations consistent with expectations. Among the network level of service measures, street block density, bicycle facility density, availability of transit service to work zone, and the ease of access to a transit stop are desirable attributes with respect to residential location choice. However, as expected, households with higher vehicle availability are likely to be those located in suburban zones with lower street block density. This is supported by the negative coefficient associated with the interaction term between street block density and household vehicle availability. Similarly, the positive coefficient associated with the interaction term between bicycle facility density and bicycle ownership indicates that households with higher bicycle ownership are likely to be located in zones with higher bicycle facility density. Although transit availability is itself positively influencing residential location choice, transit stop access time negatively impacts residential location choice. This finding is not surprising in that while most zones are served by transit, most households are living in suburban locations where the access time to a stop is likely to be greater.

The demographic, housing cost, and ethnic composition variables all indicate that there is a natural self-selection process that occurs in the housing market. Similar income groups, similar ethnic groups, and households of similar size tend to cluster together. The median housing value has a negative impact on residential location choice suggesting that, as housing prices increase, the likelihood of locating in a zone decreases.

Results of the mode choice model estimation are presented in Table 2.7 (in next page). All of the results are plausible and consistent with expectations. Relative to the auto mode, all other modes are less preferred as evidenced by the negative alternative specific constants. Higher vehicle availability is associated with auto mode usage while higher bicycle ownership is positively associated with bicycle mode usage. Higher household sizes are associated with the use of shared-ride modes consistent with the greater opportunity and/or need for sharing a ride when there are multiple individuals in a household. Both travel time and travel cost have negative coefficients, with an added negative effect in the absence of work arrangement flexibility. Presumably, sensitivity to travel time becomes more pronounced in the absence of work flexibility.

The total drive commute time for the household serves as a surrogate for the location of the household *vis-à-vis* the work locations of the workers in the household. The positive coefficient here is consistent with the notion that as households locate themselves such that their overall distance to the workplace increases, then the likelihood of becoming auto-oriented with respect to commute mode choice increases as well. The standard error of the negative error correlation term in the context of the total drive commute time index variable is suggestive of the influence of common unobserved factors that affect residential location choice and choice of auto modes. The interpretation and explanation of this finding was presented earlier in the context of the description of the results of Table 2.6.

TABLE 2.7 Estimation Results of the Mode Choice Component of the Joint Residential Location and Mode Choice Model

Variables	Parameter	t-stat
Alternative specific constants		
Auto – Drive alone	0	Fixed
Auto – Drive with passenger	-3.418	-16.88
Auto – Passenger	-1.397	-3.00
Walk	-1.020	-1.64
Bike	-3.021	-5.20
Transit	-3.825	-4.23
Socio-demographics		
Number of vehicles per number of licenses – Drive modes	1.918	4.32
Number of bicycles – Bike mode	0.419	7.70
Household size – Passenger and drive passenger modes	0.170	3.04
Individual level LOS variables (including demographic interactions)		
Travel time (in minutes)	-0.011	-1.57
interacted with inflexible work schedule	-0.008	-1.55
Travel cost (in dollars)	-0.144	-1.82
Household level commute-related variables		
Total drive commute time of all workers (minutes x 10^{-1}) – Auto modes	1.336	1.60
Standard deviation of the error term common to residential location and mode choice models – Auto modes (negative correlation)	0.859	1.53
Zonal size and density measures (including demographic interactions)		
Population density (#households per acre x 10^{-1}) – Non auto modes	0.019	2.25
Employment density (#employment per acre x 10^{-1}) – Non auto modes	0.004	2.16
interacted with household income in lowest quartile – Non auto modes	0.268	1.39
Zonal land-use structure variables		
Land-use mix – Transit mode	2.418	1.60
Local transportation network measures (including demographic interactions)		
Street block density (#blocks/square mile x 10^{-1}) – Non motorized modes	0.367	2.64
Total length of bikeways within one mile radius (meters x 10^{-5}) – Bike mode	1.267	1.22

Higher population and employment density contribute positively to bicycle and walk mode usage while a higher degree of land use mix contributes positively to transit usage. Similarly, a higher street block density and bicycle facility presence contribute positively to the use of non-motorized modes of transportation. It is to be noted here that the current model specification allows for the process of households *self selecting* themselves into neighborhoods with street block density (and bicycle facility density) compatible with their vehicle availability (and bicycle ownership). The control for such residential sorting is achieved by including vehicle availability and bicycle ownership variables in the mode choice model. These findings are consistent with those in the literature and suggest that, even when controlling for residential sorting effects, the built environment attributes (street block density and bicycle facility presence in this case) have non-negligible effects on commute mode choice.

Log-likelihood ratio tests were performed to assess the significance and contribution of observed factors and unobserved residential sorting (joint correlation) effects. The log-likelihood value at convergence for the final joint model is -9384.7. The corresponding value for the model with no allowance for unobserved variations in sensitivity to the built environment and commute attributes is -9430.94. Then, the likelihood ratio test for testing the presence of unobserved variations in sensitivity is 92.47, which is larger than the critical chi-square value with 2 degrees of freedom at any reasonable level of significance (the 2 degrees of freedom correspond to the standard deviations on the drive commute time coefficient in the residential location model, and on the common error component, related to drive commute time coefficient, between the residential location and mode choice models). Further, the log-likelihood value corresponding to equal probability for each of the 233 zonal alternatives in the residential location model and sample shares in the car ownership model (corresponding to the presence of only the threshold parameters) is -11494.3. Therefore, the likelihood ratio index for testing the presence of exogenous variable effects and unobserved taste variations is 4219, which is substantially larger than the critical chi-square value with 38 degrees of freedom at any level of significance. Overall, these test results indicate that residential sorting effects are significant as are observed and unobserved taste variations in explaining commute mode choice behavior.

2.8 SUMMARY AND CONCLUSIONS

This chapter addresses the key role of residential sorting effects in studying the impact of built environment attributes on travel mode choice. In the current land use – transportation planning context where the merits of altering the structure of the built environment to bring about changes in travel behavior are being debated, this study makes an important contribution to the field by presenting a joint model of residential location choice and commute mode choice that accounts for both observed and unobserved self-selection processes.

In previous studies of land use – travel behavior relationships, the residential location choice dimension is treated as exogenous and travel characteristics are often assumed to be affected by the attributes of the residential location. These studies often ignore the residential self-selection process that may be taking place in the housing market. Households/individuals may be locating in certain neighborhoods due to their lifestyle preferences, attitudes, values, and other unobserved factors. In the presence of such residential sorting effects, one may erroneously overestimate the impacts of built environment attributes on travel choices. In reality, individuals and households may simply be locating in neighborhoods that offer attributes consistent with their intrinsic preferences, attitudes, and values.

Recent work in the field, as described in the review of the literature in this chapter, has recognized this important concept and begun to attempt to account for residential sorting effects in evaluating the impacts of the built environment on travel behavior. However, most of the existing methods to account for residential self-selection effects either do not explicitly consider residential location choice in an integrated modeling framework. While multidimensional multinomial logit and nested logit models adopt an integrated modeling approach, the estimation of such models is computationally tedious due to the need to enumerate the composite residential location and mode choice alternative combinations. Further, several other disadvantages (as discussed in Section 2.2.2) motivate the need to adopt the mixed multidimensional modeling approach of Bhat and Guo (2007), which is further extended in this chapter to accommodate residential self-selection effects in a multinomial logit model of mode choice.

From a methodological standpoint, the chapter presents a methodology for jointly modeling the relationship between two unordered multinomial discrete choice variables, residential location choice and commute mode choice. The mixed joint model can be specified in two different ways, with identical model estimation procedure and results: (1) a simultaneous model of residential location choice and commute mode choice, which treats the two choices as a bundle of choices made simultaneously, and (2) a joint model of residential location choice and commute mode choice, with an in-built hierarchy that residential location choice precedes commute mode choice. Alternative forms of the latter way of specification can be used to test alternative hypotheses of choice making hierarchy.

The mixed joint model system developed in this chapter is estimated on a sample of households and individuals residing in Alameda County who responded to the activity-based household travel survey conducted in the San Francisco Bay Area in 2000. The model estimation results offer some key conclusions that shed additional light on the debate surrounding the land use – travel behavior relationship. First, it is found that there are significant observed factors contributing to residential self selection. It is found that households self select their residential location based on demographic characteristics such as auto and bicycle ownership, income, household size, and race. Second, and more importantly, the common error component on the total drive commute time variable supports the endogenous treatment of residential location choice in a simultaneous equations modeling framework. The negative error correlation associated with this variable suggests that there are unobserved factors that may increase (decrease) the sensitivity of households and individuals to overall commute time in their residential location decisions and also make them more (less) auto-oriented in their commute mode choice decisions. Third, the study explicitly recognizes and accommodates the fact that residential self-selection effects are specific to each built environment policy (and attribute) under consideration. For example, in this study the self-selection effect due to unobserved factors was found only in the context of total drive commute time. Unlike in this study, the previous studies have estimated a single self-selection effect with respect all built environment attributes in the model specification. Fourth, the built environment attributes such as accessibility, density, and land use mix have significant impacts on commute mode choice even after controlling for residential sorting effects and unobserved taste variations that contribute to such effects.

From a policy perspective, the results suggest that built environment attributes are not truly exogenous in travel choice decisions made by individuals. Households and individuals are locating themselves in built (transportation) environments that are consistent with their lifestyle preferences, attitudes, and values. In other words, households and individuals are making residential location and travel choice decisions jointly as part of an overall lifestyle package. Nevertheless, the findings in this paper suggest that modifying the built environment can bring about changes in mode choice behavior as evidenced by the significance of these attributes in the commute mode choice model even after controlling for residential sorting effects.

This research can be extended in at least three directions. First, use of rich data sets with attitudinal variables may enhance the understanding of the built environment – commute mode choice relationship. Third, the study relies upon statistical association between revealed choices as a means to assess the cause-and-effect relationship between the corresponding decisions. While such revealed choice data provides information on the observed decisions of decision-makers, it does not provide insights into the underlying behavioral processes that lead to those decisions (Ye *et al.,* 2007). In order to clearly understand the underlying behavior, detailed data on behavioral processes and decision sequences is needed.

CHAPTER 3

MODELING THE CHOICE CONTINUUM: AN INTEGRATED MODEL OF RESIDENTIAL LOCATION, AUTO OWNERSHIP, BICYCLE OWNERSHIP, AND COMMUTE TOUR MODE CHOICE DECISIONS

3.1 INTRODUCTION AND MOTIVATION

Integrated land-use – transportation modeling is concerned with understanding and modeling the interactions between the processes that influence regional land-use patterns (*i.e.,* the spatial pattern of urban activities and development) and the processes that influence regional travel demand patterns on the transportation network. An important aspect of modeling these interactions requires making connections between the long-, medium-, and short-term individual and household choices that influence land-use and travel demand. Making such connections includes analyzing the entire choice continuum defining individual and household choices across different temporal scales. The choices include long-term choices such as residential and work location choices, medium-term choices such as vehicle and bicycle ownership, and short-term choices such as mode choice, destination choice, and departure time choice. While the residential and work location choices influence land-use patterns, the vehicle/bicycle ownership, mode, destination, trip departure time and other travel choices influence travel demand.

Most of the travel demand models treat the longer-term choices concerning the housing decisions (such a residential tenure, housing type, and residential location), employment choices (such as entry into/exit from labor market, employment type and arrangement, and employment location), and vehicle ownership as exogenous (or predetermined) inputs. For example, most travel choice models assume variables describing residential and work locations (land use) and vehicle and bicycle ownership to be exogenous variables. Such travel choice models ignore the following two possibilities:

(1) It is possible that individuals choose to reside and work in locations that suit their lifestyle and mobility preferences. For example, those who prefer active urban lifestyles may choose to live in high density mixed land use environments, while those who prefer a more quiet sedentary lifestyle may choose to live in a suburban or rural community. This suggests that there may be substantial self-selection effects in residential location choice decisions. In other words, mobility preferences influence land use decisions, thus making location choices endogenous to the study of activity-travel demand. In a broader sense, given the cyclical or bi-directional nature of the land use – transportation relationship (*i.e.,* land use affects transportation and transportation affects land use), it is clearly inappropriate to model travel choices as a function of land use variables while assuming the latter to be predetermined or exogenous.

(2) It is possible that households can adjust with combinations of short-, medium- and long-term behavioral responses to land-use and transportation policies. A significant increase in transport costs, for example, could result in a household adapting with any combination of daily activity and travel pattern changes, vehicle ownership changes, job location changes, and residential location changes (Waddell, 2001). The travel choice models that do not consider the erogeneity of long-term and medium-term choices systematically ignore this possibility (Waddell, 2001). As the focus of transportation planning has expanded to integrate transportation planning with land-use planning, it has become important to accurately analyze the impact of land-use and transportation policies on the short-, medium- and long-term choices that influence land-use and travel demand.

To address the former issue of self-selection, several travel choice modeling studies considered the endogeneity of residential location choices. Chapter 2 presented a review of the literature on this issue, and developed a joint modeling methodology to analyze residential location and auto mode choice decisions in an integrated fashion. The residential-self selection modeling methodology developed in Chapter 2 provides a means to understand the interactions (both causal and associative) between residential location and commute mode choices, as well as estimate the impact of land-use and transportation policies on both the choice dimensions. However, incorporation of residential self-selection effects alone, although in an integrated modeling framework, does not accommodate for the possibility that individuals and households can adjust with combinations of short-, medium- and long-term behavioral responses to land-use and transportation policies. This is because the intervening impacts of the medium-term choices such as auto ownership are ignored when considering the interconnections between the long-term choices (e.g., residential location choice) and short-term choices (e.g., commute mode choice). Hence, it is important to develop truly integrated models that consider the long-term location choices, medium-term auto and bicycle ownership choices, and short-term activity-travel related choices in a simultaneous fashion.

3.2 A FRAMEWORK FOR INTEGRATED MODELING OF CHOICES

Waddell (2001) presented a framework for integrating land use and travel choices; a simplified version of this framework is presented below in Figure 3.1.

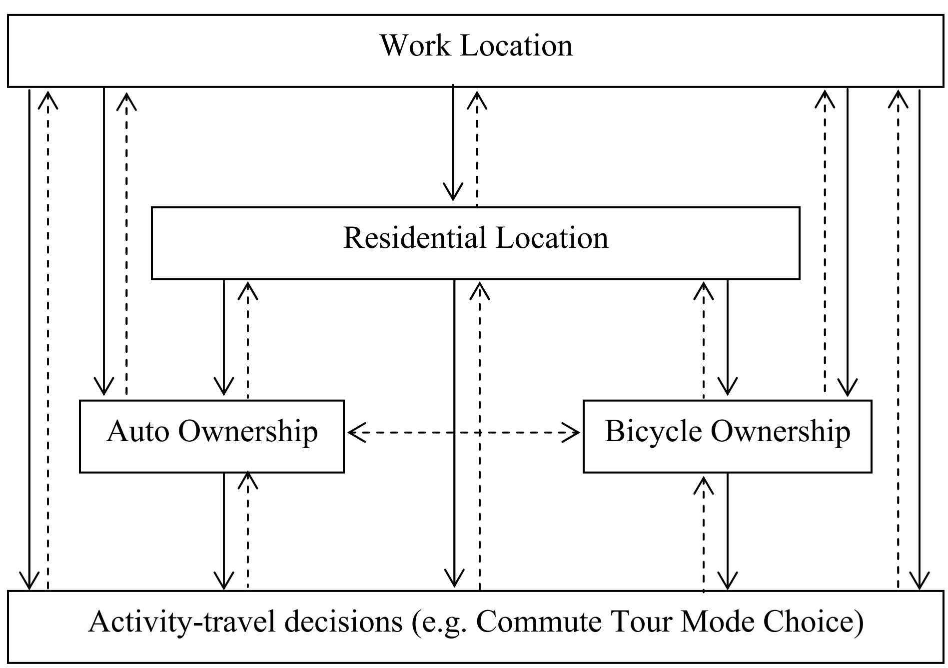

Figure 3.1. Interdependencies between Work and Home Locations, Auto and Bicycle Ownership, and Commute Mode Choice

In this framework, it is useful to consider the multitude of choices made by an individual as a lifestyle package. People choose locations to live and work, choose how many vehicles and bicycles to own/lease, and choose their mobility characteristics such as mode choice, destination choice, and trip chaining, as an overall package of choices. By developing integrated models that link these choice dimensions together, it is possible to explicitly consider the linkages between long-term, medium-term, and short-term choices. This is because one can consider location choices to be longer term choices (people locate and relocate residence or work places once every few years, on average), vehicle and bicycle ownership choices to be medium term choices (people acquire and dispose of vehicles and bicycles every few years or months), and travel choices to be short term choices (people choose mode, destination, route, departure time, and tour composition on a monthly, weekly, or daily basis).

Among the choices across such disparate temporal scales, there can be multiple multi-way interdependencies. The solid and dashed lines in the figure illustrate the two-way relationships that may exist among variables and the endogeneity of long-term and medium-term choices that results. The solid arrows indicate that each decision affects the subsequent decisions in a hierarchical manner. However, the interactions are not purely sequential in nature. The dashed arrows highlight that the interactions can be circular (or bi-directional) in nature. Each dashed arrow represents the endogeneity of a longer term decision with respect to a shorter-term decision. For example, in the context of residential location, auto ownership, bicycle ownership, and activity-travel behavior, the dashed arrows represent the possibility that individuals with particular travel preferences and needs, choose to maintain certain levels of auto/bicycle ownership and reside in neighborhoods that allow them to maintain the required levels of auto/bicycle ownership and carryout activities and travel according to their needs and preferences. That is, activity-travel needs and preferences may inform the auto/bike ownership levels (*i.e.,* the endogeneity of auto/bike ownership with respect to activity-travel behavior), and these preferences together with the auto/bike ownership levels and preferences may inform the residential location choices (*i.e.,* the endogeneity of residential location decisions or the residential self-selection).

Accommodating the above-identified relationships to link the choice dimensions across disparate temporal scales is a challenge, from a methodological standpoint and an empirical standpoint. From a methodological standpoint, the formulation and estimation of multi-dimensional choice models where all choice processes are estimated *simultaneously* is analytically and computationally complex. Of course, multiple choice processes can be estimated *sequentially* (one choice dimension at a time), but such sequential limited-information model systems are inefficient and ignore the presence of correlated unobserved heterogeneity that often pervade multi-dimensional choice processes. There may be common unobserved factors that simultaneously influence multiple choice dimensions. From an empirical standpoint, it is very rare to have a data set that simultaneously integrates information about long term location choices (built environment attributes of the residential or work place location), medium term choices (vehicle and bicycle ownership), and short term travel choices (mode choice, destination choice, departure time choice) with transportation supply (network level-of-service) attributes. Thus, integrated modeling of multi-dimensional choice processes has hitherto been rarely undertaken.

3.3 CONTRIBUTION AND ORGANIZATION OF THE CHAPTER

This chapter aims to make a substantive and methodological contribution in the integrated modeling of multi-dimensional choice processes across varying temporal scales. To this end, this chapter presents a modeling methodology and model estimation results for an integrated model of residential location choice, vehicle ownership choice, bicycle ownership choice, and commute tour mode choice. While residential location choice is a long-term choice phenomenon, vehicle and bicycle ownership choices are medium-term choice dimensions and tour mode choice is a short-term travel choice. The joint model combines a series of discrete choice models together to formulate the system – a multinomial logit (MNL) model of residential location, ordered logit models of vehicle ownership and bicycle ownership, and a multinomial logit model of commute tour mode choice.

The four dimensions considered in this joint model system are very important and of much interest to transportation researchers. Residential location choice directly impacts land use development patterns and defines the set of built environment attributes available to a household or individual. Vehicle ownership has long been considered an important determinant of mobility; although vehicle ownership is approaching saturation in the United States and one could argue that its influence as a determinant of trip making behavior is diminishing over time, this is not necessarily true in many countries around the world that are just beginning to experience a surge in vehicle ownership. Moreover, vehicle availability (defined as the ratio of number of vehicles to number of adults or drivers in a household) remains a key factor influencing levels of household and individual mobility. Bicycle ownership is a key determinant of bicycle use and active lifestyles. Finally, this chapter considers commute tour mode choice as the travel dimension of interest. Commute tours are of key interest to the profession for various reasons. Commute travel largely occurs in and contributes to congestion in the peak period.

The integrated model development and application in this chapter is associated with several salient features. First, in comparison with the model developed in the previous chapter (*i.e.,* Chapter 2), the integrated model developed in this chapter provides an enhanced modeling methodology to accommodate the intervening impact of medium-term choices (such as auto ownership and bicycle ownership) while modeling the interconnections between long-term decisions (residential location choice in this case) and short-term decisions (travel model choice in this case). Linking the four choice dimensions in a joint model system that explicitly incorporates endogeneity (self-selection effects) and flexible error correlation structures (presence of common unobserved factors affecting multiple choice dimensions) minimizes the bias in the estimation of the direct and indirect (through the intervening auto and bicycle ownership decisions) impacts of the built environment policies on commute mode choice decisions.

Second, this effort realizes the vision of modeling long-term, medium-term, and short-term choices of individuals and households as a lifestyle package (or a bundle of decisions) and represents a movement toward "truly" integrated land-use travel demand models. Third, the residential self-selection effects are simultaneously accommodated in three different models of travel behavior – auto ownership, bicycle ownership, and commute mode choice – while estimating the impact of the built environment attributes on each of these travel behavior choices. Thus, this effort makes a substantial contribution to the filed of built environment and travel behavior modeling. Fourth, all of the above-identified aspects are accommodated in the context of a "tour-based" commute mode choice model as opposed to a "trip-based" mode choice model. Commuters often link non-work activities with the commute to create commute tours; such trip linking influences mode choice (transit may not be conducive to trip linking), and contributes to additional activities or trips taking place in and around the peak period. In the context of recent developments in tour-based modeling, it is considered appropriate to treat mode choice as a tour-level decision as opposed to a trip-level decision (which has been repeatedly noted as a serious limitation of mode choice models in traditional trip-based four-step models of travel demand). In summary, the integrated model development and application in this chapter contributes to, and brings together, three different fields of research – integrated land-use travel demand modeling, built environment and travel behavior modeling, and tour-based travel demand modeling.

The rest of the chapter is organized as follows. Section 3.4 provides a review of the literature on simultaneously modeling multiple choice dimensions, including residential location choice, auto ownership and bicycle ownership choice, and travel behavior choices. Section 3.5 presents and discusses the integrated modeling methodology. Subsequently, following a description of the San Francisco Bay Area data used for the empirical analysis in Section 3.6, model estimation results are presented and discussed in Section 3.7. Finally, conclusions and directions for further research are noted in Section 3.8.

3.4 INTEGRATED MODELING OF MUTIPLE DIMESNIONS OF LOCATION AND TRAVEL CHOICES: A REVIEW OF THE LITERATURE

Integrated modeling of land use and transportation explicitly recognizes the cyclical nature of the relationship between these two entities. On one hand, land use development contributes to travel demand and the provision of transportation supply. On the other hand, travel demand (choices) and the availability of transportation supply (often measured in terms of modal accessibility and level-of-service) influence where people and businesses locate – and therefore, land use development patterns. Therefore, it is not surprising that the profession is keenly interested in integrated land use – travel behavior model development; with the advent of microsimulation modeling methodologies capable of simulating behaviors of disaggregate behavioral units or agents (decision makers), the development and implementation of such models is clearly taking center stage. Waddell (2001) and Bhat *et al*. (2004) present frameworks toward the behavioral integration of land use and transportation models, and Waddell *et al.* (2002) and Eluru *et al.* (2007) present the design of an integrated land use and activity-based travel model system for the Puget Sound Region and Dallas Fort-Worth regions, respectively. The proposed model design corresponds to a behavioral integration of various choice processes at the disaggregate level across appropriate time frames. Wegener *et al.* (2001) uses the lifestyle concept to describe the choice bundle that includes household choices of residential location, individual labor force activity, and auto ownership. It is this lifestyle concept that motivates the treatment of the choice dimensions considered in this chapter as a lifestyle package or bundle that is chosen by the individual or household and calls for the joint modeling of these dimensions in a unified framework.

The interest in and attempts at developing integrated models of location choices (land use decisions) and travel choices is not new. About 30 years ago, Lerman (1976) published a paper that presented a joint choice model of location, housing, automobile ownership, and mode choice to work. In this paper, all of these choice dimensions are treated as a jointly determined mobility bundle. Each potential combination of choices is treated as a composite alternative for a multinomial logit model of a sample of skilled, single-worker households in the Washington, D.C. area. In this modeling framework, however, as indicated in the previous chapter, the treatment of each possible combination of these decisions as a single composite choice alternative does not clearly recognize the interdependencies (*i.e.,* the causal relationships and feedback effects or associative correlations) among the individual decisions. Further, from a modeling perspective, the number of such composite alternatives can quickly explode with the increase in the number of alternatives along any choice dimension (especially in the context of residential location choice). In this context, methods to decrease the number of alternatives, such as sampling of alternatives, can be used only with a multinomial logit model which fails to capture unobserved variations in behavior and flexible correlation patterns among the decisions (*i.e.,* a mixed logit model with random coefficients cannot be used).

Since then, there has been a plethora of research efforts aimed at jointly modeling location choices (such as residential location and work locations) and mobility choices (mode choice, for example). Many of these efforts have taken the form of a multidimensional nested logit modeling framework. The nested logit modeling framework is adopted to accommodate correlations among alternatives and to provide a computationally feasible methodology for dealing with multi-dimensional choice phenomena of interest. In virtually all of these applications, location choice alternatives are sampled to form the residential (or work) location choice set; while this is feasible in the context of traditional logit modeling frameworks, such sampling approaches do not allow the adoption of newer mixed logit modeling methods that accommodate random taste variations and flexible substitution or correlation patterns. Examples of the nested logit modeling approaches to modeling multi-dimensional location and mobility choices include Abraham and Hunt (1997), Waddell (1993a), Ben-Akiva *et al.* (1980), Weisbrod *et al.* (1980), Ben-Akiva and de Palma (1986), Ben-Akiva and Bowman (1995, 1998), Rich and Nielsen (2001), and Clark *et al.* (1986). The multinomial logit and nested logit approaches are at the heart of a series of papers by Anas and his colleages (Anas and Duann, 1985; Anas, 1995; Anas, 1981) that form the basis of an integrated land use –

transportation model that includes joint choice of residential location, work location, commute mode, and shopping trip frequencies by mode and destination. These papers link the land use, travel demand (choices), and transportation supply through a series of economic-theory based market equilibrium concepts. Although many of the papers cited here constitute major contributions to the integrated modeling of location choices and mobility choices, they use multinomial and nested logit approaches that have several limitations[17] and/or treat certain choice dimensions (such as vehicle ownership) as exogenous to the model system. Another such model is presented by Eliasson and Mattsson (2000).

More recently, Salon (2006) completed a doctoral dissertation that explores the relationship between the transportation and land use system in New York City by modeling the transportation and residential location choices of New Yorkers. The research involves the development of a multidimensional nested logit model of residential location, car ownership, and commute mode choice. In another joint modeling effort, Salon (2006) addressed the issue of residential self-selection in understanding the impact of the built environment on walking levels. This is accomplished by developing and estimating a multidimensional nested model of residential location, auto ownership, and walking levels. This study comes closest to what is trying to be achieved in this chapter. However, in this chapter, both of the concepts – joint choice modeling and addressing self-selection or endogeneity – are combined into a unified multi-dimensional choice model of residential location choice, auto ownership, bicycle ownership, and commute tour mode choice. The addition of the bicycle ownership dimension is unique and has virtually never been treated as an endogenous choice dimension in past work. Further, this chapter adopts the mixed multidimensional choice modeling methodology that is associated with several advantages identified in Chapter 2.

[17] See Chatper 2 for a detailed discussion on the limitations of the multidimensional multinomial logit and nested logit modeling approaches.

There are a host of studies that address the issue of residential self-selection in modeling and analyzing the impacts of built environment attributes on travel choices[18]. This chapter attempts to make a substantive contribution to this area of research by considering residential self-selection in a tour-based modeling context as opposed to a trip-based modeling context when examining mode choice. Also, this study accounts for potential self-selection in auto ownership, bicycle ownership, and commute tour mode choice dimensions simultaneously.

Early studies that incorporated endogeneity of car ownership in mode choice models include those by Train (1980) and Ben-Akiva and Lerman (1974). However, this study is the first that simultaneously incorporates endogeneity of both car ownership and bicycle ownership in a mode choice model. Thus, in this chapter, mode choice preferences that could potentially impact car ownership and bicycle ownership levels are considered simultaneously.

A recent study by Lawrence *et al.* (2007) considers the impact of the built environment (of residential and work locations) in tour-based mode choice models, but ignores the residential self-selection effects and endogeneity of auto ownership and bicycle ownership. Maat and Timmermans (2007), in a car ownership analysis, and Lawrence *et al.* (2007), in a mode choice analysis, are rare studies that have considered the impact of the built environment associated with work locations on mobility choices. These studies, however, have limited the analysis to a limited number of attributes. In addition, self-selection effects and endogeneity effects are not considered in these studies.

In summary, the limited literature review presented here shows that that there has been much work in the area of simultaneously modeling multiple choice dimensions

[18] See Chatman (2005), Kitamura *et al.* (1997), Schwanen and Mokhtarian (2003), Boarnet and Sarimento (1998), Greenwald and Boarnet (2001), Khattak and Rodriguez (2005), Handy *et al.* (2006), Handy and Clifton (2001), and Krizek (2000 and 2003) for analyses of residential self-selection in models of trip frequency by one or more modes and/or purposes; Schwanen and Mokhtarian (2005b), Khattak and Rodriguez (2005), Bagley and Mokhtarian (2002), Handy *et al.* (2005), and Krizek (2000, 2003) for analyses of residential self-selection in models of travel mileage by one or more modes; Cervero and Duncan (2002), Hammond (2005), Pinjari *et al.* (2007), Schwanen and Mokhtarian (2005a), Salon (2006), and Zhang (2006) for commute mode choice analyses that consider residential self-selection; and Cervero and Duncan (2003), Greenwald (2003), and Salon (2006) for non-commute mode choice analyses that accommodate residential self-selection.

covering different time scales (long term location choices, medium term auto ownership choices, and short term mode choices). There has also been work in the area of modeling impacts of built environment attributes on mobility choices while accounting for endogeneity or residential self-selection effects. This study constitutes a quantum leap forward in simultaneously modeling multiple choice dimensions across varying time scales while accommodating endogeneity effects, self-selection effects, and unobserved heterogeneity, all at a time. In other words, this chapter combines all of the different aspects that researchers have attempted to address in a piecemeal fashion in the past. Moreover, the modeling methodology proposed in this chapter overcomes the limitations of the multinomial and nested logit modeling approaches that have dominated past research.

3.5 DATA

This section provides a description of the data and the sample of observations used in this study.

3.5.1 Data Sources

The primary source of data used for this analysis is the 2000 San Francisco Bay Area Travel Survey (BATS) designed and administered by MORPACE International Inc. for the Bay Area Metropolitan Transportation Commission (MTC). The survey collected information on all activity episodes (in-home and out-of-home) undertaken by individuals from over 15,000 households in the Bay Area for a two-day period (see MORPACE International Inc., 2002, for details on survey, sampling and administration procedures). Information characterizing the context (activity type, start and end times of the activity, and location of participation) of each activity episode was collected. Furthermore, data on individual and household socio-demographics was also obtained.

In addition to the 2000 BATS data, several other secondary data sources were used to derive spatial variables characterizing the activity-travel environment in the region. These include: (1) Land-use/demographic coverage data, obtained from the MTC, (2) The zone-to-zone motorized travel level of service files, obtained from MTC, (3) The

Census 2000 population and Housing data summary files (SF1), (4) GIS layers of businesses (automotive businesses, shopping and grocery stores, medical facilities and personal services, food stores, sports and fitness centers, parks and gardens, restaurants, recreational businesses, and schools), obtained from the InfoUSA business directory, (5) GIS layers of bicycling facilities and bicycle network, also obtained from MTC, and (6) GIS layers of highway (interstate, national, state and county highways) network and local roadways (local, neighborhood, and rural roads) network, extracted from the Census 2000 Tiger files.

3.5.2 Sample Formation

3.5.2.1 Sample Extraction

The analysis was confined to individuals and households residing in five Counties (San Francisco, San Mateo, Santa Clara, Alameda, and Contra Costa) of the southern San Francisco Bay area. Several steps were undertaken to generate the tour mode choice data for the current analysis. First, only employed adults (of age 16 years or older) who traveled to work on the travel day were selected. Second, only weekday data was selected from the sample, since the focus of our analysis is on the mode choice for commute travel, which usually takes place on weekdays. Third, of the two-day activity-travel diary data available for each individual, only one randomly chosen day was picked. In addition, the commute tour mode choice analysis was restricted to a single randomly chosen individual from each household. These decisions were made to keep the sample size manageable in the estimations, and also to avoid the problem of repeated data measurement from the same individual and/or household. Fourth, the activity-travel information was converted into an activity-based database in which each record was an activity (not travel to the activity), and a tour-based database with each record as a tour. In the activity-based database, the mode of travel (and any change in the mode of travel) to each activity stop was recorded. Subsequently, only commute tours (and corresponding activity-travel database) were selected. Fifth, the work-to-home commute tours were discarded, since in most of the cases the home-to-work mode was the same as the work-to-home mode.

The final estimation sample consists of 5147 individuals from 5147 households residing in five Counties of southern San Francisco Bay area. Subsequent to the extraction of the tour mode choice data, the household residential location, auto ownership, and bicycle ownership, and other socio economic information was used to build a comprehensive residential location – auto ownership – bicycle ownership – tour mode choice database with several socioeconomic and built environment attributes ready for the integrated analysis.

3.5.2.2 Residential Location Alternatives

In the residential location choice model, each individual/household was assumed to have had the choice of residing in any one of the zones (traffic analysis zones) in the County in which (s)he resides. A sampling scheme was not implemented to reduce the number of alternatives in the residential choice model because the overall residential choice model does not correspond to a simple multinomial logit model; instead, a mixed logit modeling approach was adopted.

3.5.2.3 Tour Mode (and Modal Availability) Definitions

The modes of travel reported in the BATS data include auto (drive alone, and shared ride), transit (bus transit, rail transit, and a small share of ferry boats), bicycle and walk[19]. Several steps were undertaken to define the tour mode based on a descriptive analysis of the data. First, in the activity-based database, the primary mode of travel (*i.e.,* the mode used for the longest segment of travel, among the trip segments) was defined for each activity stop in the tour. For example, if a person accessed the transit stop nearby his/her home by bicycle/walk, traveled in a bus or rail, and reached an activity stop by bicycle/walk, the primary mode of travel for this stop was deemed as transit. Second, the tour mode was defined based on the primary mode used for each activity-stop in the tour. A descriptive analysis indicated that for most of the commute tours other than those involving pick-up/drop-off/other serve passenger stops, the primary mode of travel remained the same for all stops within a tour. For the tours involving pick-up/drop-

[19] While other modes such as taxi were reported in the survey, such commute tours were very small in number.

off/other serve passenger stops, the mode of travel was a combination of drive alone and shared ride modes. For simplicity, the drive alone and shared ride modes were combined into a single auto mode. Thus the final commute tour database included four tour modes: (1) Auto, (2) Transit, (3) Bicycle, and (4) Walk.

Subsequent to the tour mode definitions, the modal availability was defined as follows. The auto mode was assumed to be available for those individuals with at least one car available in the household. The transit mode was assumed to be available for an individual if his/her home and work zones were connected by transit (using the transit level-of-service data obtained from MTC). The bicycle mode was assumed to be available for an individual if his/her household owned at least one bicycle and if his/her home and work zones were connected by a bicycle route within 2 hours. The walk mode was assumed to be available for an individual if his/her home and work zones were within two hours of each other.

3.5.2.4 Built Environment (or Activity-Travel Environment or ATE)[20] *Measures*

The data from the secondary sources identified in Section 3.5.1 were used to compute a host of built environment (or ATE) measures for each traffic analysis zone (TAZ), including:

1. Zonal size and density measures, such as total population, population density, household density, density of employment by each of several employment categories, and dummy variables for central business district (CBD), urban, suburban, and rural areas (computed based on employment density). These attributes were obtained from the zonal land-use data file.

[20] In the rest of this chapter, we will use the more general term activity-travel environment (ATE) instead of built environment. The ATE attributes of a traffic analysis zone (TAZ) include the sociodemographic, natural, and built environment characteristics of the TAZ, as well as the transportation system facility characteristics in and around the TAZ. The sociodemographic environment may include such characteristics as the ethnic composition, age distribution, and income level distribution in the TAZ; the natural environment comprises vegetation, gardens, parks, and water bodies; the built environment consists of the urban form design, land-use structure (such as land-use mix, employment and residential density), and spatial distribution of the activity centers; and the transportation system consists of the elements of transportation infrastructure, such as highways, bikeways, transit systems, local streets, and side walks *etc.*, and the accessibility and level-of-service offered by the transportation infrastructure for various activities and travel.

2. Zonal land-use structure variables, such as housing type measures (fractions of single family, and multiple family dwelling units), fractions of zonal area in residential and commercial land-uses, and land-use mix (see Bhat and Gossen, 2004 for a description of the land-use mix variable; the variable takes a value of zero for zones with only one type of land-use and a value of 1 for zones with equal distributions in area among residential, commercial, and other land-uses). The zonal land-use structure variables were constructed from the zonal land-use data file.
3. Regional accessibility measures, such as shopping accessibility, recreational accessibility, and employment accessibility. These are Hansen-type (see Fotheringham, 1983) accessibility measures computed from the zonal land-use and level-of-service data.
4. Zonal demographics, such as average household size and median household income. These demographic measures were derived from the Census 2000 population and housing data summary file (SF1).
5. Zonal ethnic composition measures, constructed as fractions of Caucasian, African-American, Hispanic, Asian, and other ethnicity populations, also derived from the Census 2000 population and housing data summary file.
6. Zonal activity opportunity variables, such as activity center intensity (*i.e.,* the number of business establishments per square mile) and density (*i.e.,* the number of business establishments per square mile) for each of the following activity types (extracted from the InfoUSA business establishments data): (a) maintenance (grocery stores, gas stations, food stores, car wash, automotive businesses, banks, medical facilities, *etc.*), (b) physically active recreation (fitness centers, sports centers, dance and yoga studios, parks, gardens, *etc.*), (c) Physically inactive recreation (theatres, amusement centers, arcades, *etc.*), and (d) eat-out (restaurants and eateries).

Such zonal-level activity-travel environment measures were created not only for the potential residential location alternatives for each household, but also for the employment locations of all employed individuals residing in the household. The zonal-level characteristics of the employment locations were averaged over all the employed individuals in the household to obtain household level aggregate measures of the activity-travel environment in the work locations. In addition to the zonal-level activity-travel environment variables, the following variables were extracted and appended to the database: (1) several transportation network level of service (LOS) characteristics by each mode, in and around the residential locations and work locations (such as highway density, bikeway density, and street block density, number of zones accessible by transit from the residential and work locations, number of zones accessible by bike and walk within 6 miles, *etc.*); (2) individual and household level commute variables (such as travel times and travel costs by each mode, total auto commute travel time for all employed members in the household, travel distance by bike and walk modes, number of commuters in the household that have transit available between home and work zones, number of commuters in the household that have bike route available between home and work zones, *etc.*).

All of the above mentioned activity-travel environment attributes of residential and work locations and home-to-work level of service characteristics at the individual and household level were merged with the analysis database. We am not aware of any other study in the literature that considers the impact of such a comprehensive set of activity-travel environment attributes of both residential and work locations, and commute level of service characteristics.

3.5.3 Sample Description

A descriptive analysis of the estimation data sample (5147 individuals from 5147 households) indicated the following characteristics of the dependent variables (residential location, auto and bicycle ownership, and commute tour mode choice) in this study.

The residential locations belonged to one of the 127 TAZs of the San Francisco (SF) County, 115 TAZs of the San Mateo (SM) County, 269 TAZs of the Santa Clara (SC) County, 236 TAZs of the Alameda (AL) County, or 154 TAZs of the Contra Costa (CC) County. Out of the 5147 households, 12.3% belonged to SF, 13.6% belonged to SM, 32.7% belonged to SC, 25.7% belonged to AL, and 15.6% belonged to CC Counties.

The auto ownership descriptives indicate an average ownership of 1.85 autos per household. It was found that 3.5% of the households did not own any automobiles, 33.9% owned one car, 43.1% owned two cars, and 19.4% owned three or more cars.

The bicycle ownership descriptives indicate an average bicycle ownership of 1.42 bicycles per household. Of all households, 36.8% of the households did not own a bicycle, 22.5% owned one bicycle, 20.9% owned two bicycles, and 19.8% three or more bicycles.

Descriptive analysis of the commute tour mode choice variable indicate the following mode shares: 83.83% auto, 11.47% transit, 1.22% bicycle, and 3.48% walk. The mode shares matched well with the Census journey to work mode shares for each of the 5 counties. Among all commute tours, 12.7% involved at least one stop between the origin (home) and destination (work). In a trip-based mode choice analysis, such complex tours would be ignored and the mode shares used for analysis would be skewed.

3.6 ECONOMETRIC MODELING METHODOLOGY

This section presents and discusses the econometric modeling methodology developed for the integrated modeling of multidimensional individual and household-level land-use related and travel related choices.

3.6.1 Model Structure

In the following presentation of the model structure, let the indices q (q = 1, 2, …, Q) , i (i = 1, 2, …, I), and k (k = 1, 2, …, K) represent the decision-maker, the spatial unit of residence, and the modal alternative, respectively, and the terms n (n = 1, 2, …, N), and m (m = 1, 2, …, M) represent the auto ownership level (*i.e.,* the number of cars) and the bicycle ownership level (*i.e.,* the number of bicycles), respectively.

Given the above notational preliminaries, the equation system for the joint residential location choice, auto ownership, bicycle ownership, and commute tour mode choice is as follows:

$$s^*_{qi} = \varphi'_q z_i + \varepsilon_{qi}, \text{ spatial unit } i \text{ chosen if } s^*_{qi} > \max_{\substack{j=1,2,\ldots,J \\ j \neq i}} s^*_{qj}.$$

$$c^*_{qi} = \alpha' x_q + \delta'_q z_i + \xi_{qi}, \;\; c_{qi} = n \text{ if } \psi_{n-1} < c^*_{qi} < \psi_n.$$

$$b^*_{qi} = \beta' x_q + \theta'_q z_i + \zeta_{qi}, \;\; b_{qi} = m \text{ if } \tau_{m-1} < b^*_{qi} < \tau_m. \quad (3.1)$$

$$u^*_{qkinm} = \phi'_k x_q + \hbar_k 1[c_{qi} = n] + \lambda_k 1[b_{qi} = m] + \chi'_{qk} z_i + \varsigma_{qkinm}, \text{ mode } k \text{ chosen if}$$

$$u^*_{qkinm} > \max_{\substack{d=1,2,\ldots,K \\ d \neq k}} u^*_{qdinm}$$

The discussion that follows presents the structure of the residential location choice model component (Section 3.6.1.1), the auto ownership and bicycle ownership model components (Section 3.6.1.2), and the commute tour mode choice model component (Section 3.6.1.3) of the above equation system, and then highlight the interdependencies among the four components (Section 3.6.1.4).

3.6.1.1. The Residential Location Choice Component of the Joint Model System

The residential location component, represented by the first equation of the Equation system (3.1), takes the familiar discrete choice formulation, as discussed below.

The first equation of the Equation system (3.1) is the indirect utility function for the choice of residence. Specifically, s^*_{qi} is the indirect (latent) utility that the q^{th} individual (as part of her/his household) obtains from locating in spatial unit *i*, z_i is a vector of activity-travel environment (ATE) attributes corresponding to spatial unit *i* (such as residential density and land-use mix), and φ_q is a coefficient vector capturing individual *q*'s sensitivity to attributes in z_i.

Each element *l* of φ_q is parameterized as $\varphi_{ql} = \varphi_l + \Gamma'_l x_q + v_{ql} + o_{ql} + \pi_{ql} + \sum_k \omega_{qkl}$.

In such a parameterization, x_q is a vector of observed individual-specific factors (such as income and/or household size of individual *q*'s household) affecting sensitivity to the l^{th} attribute in vector z_i, and v_{ql}, o_{ql}, π_{ql}, and ω_{qkl} (k = 1, 2, 3, …, K) are individual-specific unobserved factors impacting individual *q*'s sensitivity to the l^{th} attribute in vector z_i. v_{ql}

includes only those individual-specific unobserved factors that influence sensitivity to residential choice, o_{ql} includes the individual-specific unobserved factors that influence both residential choice and auto ownership, π_{ql} includes the individual-specific unobserved factors that influence both residential choice and bicycle ownership, while each ω_{qkl} ($k = 1, 2, 3, \ldots, K$) includes only those individual-specific unobserved factors that influence both residential choice and the choice of modal alternative k.

For example, consider the sensitivity of individuals from a household to residential density. The household may have a higher sensitivity (than its observationally equivalent peer group) to residential density because they are social extroverts and perceive higher residential density as providing a more socially vibrant setting. The socially extroverted nature, however, may not have an impact on car ownership, bicycle ownership, and commute mode choice. This would be captured in v_{ql}. Now, another unobserved individual factor may be the social status. This is likely to impact the sensitivity to residential density in residential choice (because locations with lower residential densities may be preferred by individuals with an inclination to live in exclusive enclaves to satisfy their social prestige desires) and also influence auto ownership propensity (because higher auto ownership is a symbol of social status). This would be included in o_{ql}. Similarly, there may be unobserved factors that affect both residential choice and bicycle ownership propensity (included in π_{ql}), and/or that affect both residential choice and mode choice (included in ω_{qkl} for mode k). Additional discussion on the unobserved factors is furnished later.

Finally, in the first equation of equation system (3.1), ε_{qi} is an idiosyncratic error term assumed to be identically and independently extreme-value distributed across individuals and spatial alternatives.

3.6.1.2. The Auto Ownership and Bicycle Ownership Components of the Joint Model System

The second and third equations in equation system (3.1) correspond to the ordered-response structure for auto ownership and bicycle ownership decisions, respectively. Specifically, c^*_{qi} and b^*_{qi} represent the latent car ownership and bicycle ownership propensities, respectively, of the q^{th} individual (as part of his/her household) should the household choose to locate in spatial unit *i*, x_q is a set of individual sociodemographic characteristics (such as income and number of children of his/her household) and z_i is the vector of activity-travel environment (ATE) attributes corresponding to spatial unit *i*.[21] α in the car ownership equation and β in the bicycle ownership equation are coefficient vectors representing the impact of socio-demographics on car ownership and bicycle ownership propensities, respectively. δ_q and θ_q are individual-specific coefficient vectors capturing the impact of ATE attributes on car ownership and bicycle ownership decisions, respectively.

Each element *l* of δ_q and θ_q is parameterized as follows: $\delta_{ql} = \delta_l + \Lambda_l^{'} x_q + \mu_{ql}$, and $\theta_{ql} = \theta_l + \Delta_l^{'} x_q + \eta_{ql}$. In such parameterizations, x_q is a vector of observed individual-specific factors influencing sensitivity to the ATE z_{il}, Λ_l and Δ_l are corresponding vectors of coefficients in car ownership and bicycle ownership equations, respectively, μ_{ql} and η_{ql} are the terms capturing the impact of individual-specific unobserved terms associated with different sensitivities to ATE attributes on car ownership and bicycle ownership propensities, respectively.

Finally, in the second and third equations of equation system (3.1), ξ_{qi} and ζ_{qi} are error terms in the car ownership and bicycle ownership propensity equations,

---

[21] Note that we are introducing the full vector z_i of ATE attributes, and the full vector x_q of individual-specific characteristics in the residential choice, car ownership, and bicycle ownership equations (as well as in mode choice equations, as will be discussed later), for notational ease. In general, it is not necessary that each of the ATE attributes in z_i and individual-specific characteristics in x_q influence each of the residential location, auto ownership, bicycle ownership, and mode choice decisions. Additionally, it is possible that some of the ATE attributes have a mean effect of zero across decision-makers for residential choice and/or car ownership and/or bicycle ownership and/or mode choice, but have a statistically significant distribution around the zero mean.

respectively, each of which is partitioned into four components as follows: $\xi_{qi} = \sum_l (\pm o_{ql}) z_{il} + \sum_k \vartheta_{qk} + \iota_q + \partial_{qi}$, and $\zeta_{qi} = \sum_l (\pm \pi_{ql}) z_{il} + \sum_k \upsilon_{qk} \pm \iota_q + \rho_{qi}$. The $\pm o_{ql}$ (and $\pm \pi_{ql}$) terms are the common error components relating to the sensitivity of the ATE attribute of z_{il} in residential choice and car ownership propensity (and residential choice and bicycle ownership propensity), the $\sum_k \vartheta_{qk}$ (and $\sum_k \upsilon_{qk}$) terms include the ϑ_{qk} (and υ_{qk}) elements that are common error components capturing the unobserved factors that affect car ownership propensity (and bicycle ownership propensity) and the utility of mode *k*, the ι_q terms are the common error components capturing the unobserved factors that affect car ownership propensity as well as bicycle ownership propensity, while ∂_{qi} and ρ_{qi} are idiosyncratic terms assumed to be identically and independently standard logistic distributed across individuals and spatial units, in the car ownership and bicycle ownership propensity equations, respectively.

The car ownership propensity c_{qi}^* and the bicycle ownership propensity b_{qi}^* are mapped to the observed car ownership level c_{qi} and the observed bicycle ownership level b_{qi}, using the ψ and τ thresholds, respectively, in the usual ordered-response fashion.

3.6.1.3. The Commute Tour Mode Choice Component of the Joint Model System

The fourth equation in the equation system (3.1) corresponds to the unordered response structure for the mode choice decisions. Specifically, u_{qkinm}^* is the indirect (latent) utility that the q^{th} individual obtains from choosing k^{th} mode for his/her commute should his/her household choose to locate in spatial unit *i* and own *n* number of vehicles and *m* number of bicycles. x_q is a set of household and individual socio-demographic characteristics (such as income and number of children of his/her household) that influences commute mode choice, and ϕ_k is the corresponding coefficient vector in the utility of mode *k*. $1[c_{qi} = n]$ ($1[b_{qi} = m]$) are scalars that take a value of 1 if the household is actually observed to choose a car ownership level (bicycle ownership level) of n (m) and 0 otherwise, while $\hbar_k$ ($\lambda\!\!\!^{-}_k$) are scalars of corresponding coefficient vectors capturing the impact of auto ownership (bicycle ownership) on the utility of mode k. z_i is the vector of ATE attributes corresponding to spatial unit i, and χ_{qk} is an individual-specific coefficient vector capturing the impact of ATE attributes on the utility for mode k.

Each element l of χ_{qk} is parameterized as follows: $\chi_{qkl} = \chi_{kl} + \Upsilon'_{kl} x_q + \gamma_{qkl}$, where x_q is a vector of observed individual- or household-specific factors influencing sensitivity to the l^{th} ATE attribute of z_i, Υ_{kl} is the corresponding vector of coefficients in the utility of mode k, and γ_{qkl} captures the impact of individual-specific unobserved terms associated with different sensitivities to ATE attributes in the utility of mode k.

Finally, in the fourth equation of equation system (3.1), ς_{qkinm} is an error term which is partitioned into four components as follows: $\varsigma_{qkinm} = \sum_l (\pm\omega_{qkl}) z_{il} \pm \vartheta_{qk} \pm \upsilon_{qk} + \wp_{qkinm}$. The $\pm\omega_{qkl} z_{il}$ terms are the common error components relating to the sensitivity to ATE attribute z_{il} in residential choice and mode choice for mode k, the $\pm\vartheta_{qk}$ term is the common error component capturing the unobserved factors that affect car ownership propensity and the utility of mode k, and the $\pm\upsilon_{qk}$ term is the common error component capturing the unobserved factors that affect bicycle ownership propensity and the utility of mode k, while the $\wp_{qkinm}$ is an idiosyncratic error term assumed to be identically and independently extreme-value distributed across modal individuals, alternatives, spatial units, auto ownership levels, and bicycle ownership levels.[22]

[22] In our empirical analysis, we further partitioned $\wp_{qki}$ into error components that generate covariance across modal alternatives (for example, individuals who are generically inclined toward walking and bicycling may associate a higher utility than their observationally equivalent peers for both walk and bicycle modes). However, in our final specification, we did not find any such statistically significant covariance terms. Thus, for simplicity in presentation, we are imposing the restriction of zero covariances across modal alternatives in the model structure presentation.

3.6.1.4. Interdependencies among the Components of the Joint Model System

The Equation system (3.1), the individual components of which were explained in the previous three sections, may be rewritten as the following equation system:

$$\begin{aligned}
s_{qi}^{*} &= \sum_{l} (\varphi_l + \Gamma_l' x_q + v_{ql}) z_{il} + \sum_{l} o_{ql} z_{il} + \sum_{l} \pi_{ql} z_{il} + \sum_{k} \sum_{l} \omega_{qkl} z_{il} + \varepsilon_{qi} \\
c_{qi}^{*} &= \alpha' x_q + \sum_{l} (\delta_l + \Lambda_l' x_q + \mu_{ql}) z_{il} + \sum_{l} (\pm o_{ql} z_{il}) + \sum_{k} \vartheta_{qk} + \iota_q + \partial_{qi} \\
b_{qi}^{*} &= \beta' x_q + \sum_{l} (\theta_l + \Delta_l' x_q + \eta_{ql}) z_{il} + \sum_{l} (\pm \pi_{ql} z_{il}) + \sum_{k} \upsilon_{qk} \pm \iota_q + \rho_{qi} \\
u_{qkinm}^{*} &= \phi_k' x_q + \hbar_k 1[c_{qi} = n] + \lambda_k 1[b_{qi} = m] + \sum_{l} (\chi_{kl} + \Upsilon_{kl}' x_q + \gamma_{qkl}) z_{il} + \sum_{l} \pm \omega_{qkl} z_{il} \pm \vartheta_{qk} \pm \upsilon_{qk} + \wp_{qkinm}
\end{aligned} \tag{3.2}$$

The Equation system (3.2) shows the joint nature and the interdependencies among the different components of the model system represented by Equation system (3.1). The joint nature of the model system arises because of the presence of the common unobserved terms, $o_{ql} z_{il}$, $\pi_{ql} z_{il}$, $\omega_{qkl} z_{il}$, ϑ_{qk}, υ_{qk}, and ι_q, between the different components of the model system. Each of these common unobserved terms captures the jointness between two specific components of the model system, and represents a particular type of interdependency between those two components. Each of the interdependencies is discussed below.

<u>Residential self-selection</u>

The $o_{ql} z_{il}$, $\pi_{ql} z_{il}$, and $\omega_{qkl} z_{il}$ terms capture the jointness of the residential location choice component with the auto ownership, bicycle ownership, and commute mode choice components, respectively. These terms allow self-selection of individuals into neighborhoods based on their unobserved (to the analyst) preferences of auto ownership, bicycle ownership, and commute mode choice, respectively. This can be explained by the following example.

Consider unobserved individual factors such as auto disinclination, fitness consciousness and environmental friendliness that make some individuals associate higher utility to bicycle/walk modes, and locate in neighborhoods that allow them to bicycle or walk to work (*i.e.,* locate in neighborhoods that are closer to work places) and/or own more bicycles and/or own less cars. In such case, the $\omega_{qkl} z_{il}$ terms in the first and fourth equations of the Equation system (3.2) capture the unobserved factors that affect both residential location choice and mode choice preferences, where z_{il} can, for example, be the commute time, and the index k can correspond to bicycle and walk modes. That is, the unobserved factors that affect the sensitivity of commute time in the bicycle/walk modal utilities (or preferences) also affect the sensitivity of commute time in residential location preferences, due to which individuals self select into neighborhoods that are closer to work locations. These common unobserved factors give rise to correlations between residential and modal utilities. The $\pm$ sign in front of the $\omega_{qkl} z_{il}$ terms in the modal utility equations indicates that the correlation in the unobserved factor *l* may be positive or negative (though for any attribute *l* in z_{il}, one can have '+' or ' – ' correlations between the residential and modal utilities; the $\pm$ specification allows us to test the most appropriate signs of correlations). Similarly, the $\pi_{ql} z_{il}$ terms capture the unobserved factors due to which individuals may have, for example, higher bicycle ownership propensity and live closer to work locations, and the $o_{ql} z_{il}$ terms capture the unobserved factors due to which individuals may have, for example, lower auto ownership propensity and live closer to work locations. Such residential self-selection effects are represented by the dashed arrows to the residential location box in Figure 3.1.

It is important to note here that the model system allows for residential self-selection due to observed individual attributes also. Consider, for example, that high income households stay away from high density neighborhoods. This can be reflected by including income as a variable in the x_q vector in the residential choice equation. High income households are also likely to own more cars than low income households. The residential self-selection based on income can then be controlled for when evaluating the effect of density on car ownership by including income as a variable in the x_q vector in the car ownership propensity equation.

Endogeneity of auto ownership and bicycle ownership in mode choice

The ϑ_{qk} and υ_{qk} terms capture the jointness of the auto ownership and bicycle ownership components, respectively, with the mode choice component. These terms allow the endogeneity of auto ownership and bicycle ownership in mode choice equations. Consider, for example, unobserved factors such as social status. Individuals with high social status may prefer to associate higher car ownership propensity as well as use auto mode for commuting to work. The ϑ_{qk} term captures such unobserved factors affecting both auto mode choice and auto ownership and gives rise to correlations between the utility of auto mode and the auto ownership propensity. Similarly, the υ_{qk} term captures the common unobserved factors, such as physical fitness orientation that may increase the bicycle ownership propensity and influence mode and increase the bicycle modal utility. The correlation in such cases can be expected to be positive, though one can again test the directionality of correlation by experimenting with both the $\pm$ signs in the mode utility equation. Such endogeneity effects discussed here are represented by the dashed arrows to the auto ownership and bicycle ownership boxes in Figure 3.1.

Ignoring the endogeneity of auto ownership in auto mode choice utility, and that of bicycle ownership in bicycle mode choice utility can lead to an inflated influence of auto ownership on auto mode choice and of bicycle ownership on bicycle mode choice.

Jointness of auto ownership and bicycle ownership propensities

The ι_q term captures the unobserved factors that affect both auto ownership and bicycle ownership, but not residential location choice or mode choice. These factors include, for example, recreational activity preferences due to which individuals/households may own larger number of vehicles (especially vehicles such as SUVs and vans for recreational purposes) and use them to carry their bicycles for recreational activities. In such case the correlation between auto and bicycle ownership propensities can be expected to be positive. Ignoring such factors by ignoring the ι_q term, can lead to inflated estimates of the other common unobserved terms and misinform about residential self-selection and endogeneity of auto ownership and bicycle ownership. Also, other coefficient estimates may be biased due to the neglect of the ι_q term. The jointness discussed here is represented by the dashed arrow between the auto ownership and bicycle ownership boxes in Figure 3.1.

<u>Other interdependencies in the model system</u>

The above three sections discussed the interdependencies among the various decisions (*i.e.,* the residential location choice, auto ownership, bicycle ownership, and mode choice decisions) due to unobserved factors. These interdependencies can also be termed as associative correlations. That is, the interdependencies such as residential self-selection and endogeneity are mere statistical correlations that do not imply a causal relationship.

In addition to the associative correlations (represented by the dashed arrows in Figure 3.1), the various components of the joint system are connected by causal relationships (represented by the solid arrows in Figure 3.1). These relationships are captured by the coefficients of ATE attributes (*i.e.,* the coefficients on the z_{il} terms) in the auto ownership, bicycle ownership and mode choice equations, and by the coefficients of auto ownership and bicycle ownership variables (*i.e.,* the n_{qi} and m_{qi} terms) in the mode choice equations. The purpose of including the common unobserved factors in the model system is to estimate the "true" impact of the z_{il}, n_{qi} and m_{qi} terms, in order to be able to carryout accurate predictions under various policy scenarios.

3.6.2 Model Estimation

The parameters to be estimated in the Equation system (3.2) include the vectors α, β, ϕ_k, φ_l, δ_l, θ_l, χ_{kl}, Γ_l, Λ_l, Δ_l, and Υ_{kl}, and scalars $\hbar_k$ and λ_k, and the variances of the stochastic components ν_{ql}, μ_{ql}, η_{ql}, o_{ql}, π_{ql}, ω_{ql}, ϑ_{qk}, υ_{qk}, and ι_q (all assumed to be normally distributed with variances $\sigma^2_{\nu ql}$, $\sigma^2_{\mu ql}$, $\sigma^2_{\eta ql}$, σ^2_{oql}, $\sigma^2_{\pi ql}$, $\sigma^2_{\omega ql}$, $\sigma^2_{\vartheta qk}$, $\sigma^2_{\upsilon qk}$, and $\sigma^2_{\iota q}$, respectively).

Let Ω represent a vector that includes all of the parameters to be estimated, and let $\Omega_{-\sigma}$ represent a vector of all parameters except the variance terms. Also, let g_q be a vector that stacks the ν_{ql}, μ_{ql}, η_{ql}, o_{ql}, π_{ql}, ω_{ql}, ϑ_{qk}, υ_{qk}, and ι_q terms and let Σ be a corresponding vector of standard errors. Define $r_{qj}=1$ if individual (or household) q resides in spatial unit j and 0 otherwise, $a_{qn}=1$ if individual (or household) q owns n cars and 0 otherwise ($a_{qn}=1[c_{qi}=n]$), $e_{qm}=1$ if individual (or household) q owns m bicycles and 0 otherwise ($e_{qm}=1[b_{qi}=m]$), and $p_{qk}=1$ if individual q commutes by mode k and 0 otherwise. Then the likelihood function for a given value of $\Omega_{-\sigma}$ and g_q may be written for an individual q as:

$$
\begin{aligned}
L_q(\Omega_{-\sigma})\,|\,g_q = \prod_{j=1}^{J}\prod_{n=1}^{N}\prod_{m=1}^{M}\prod_{k=1}^{K} \Bigg\{ & \left[\frac{\exp\left[\sum_l \left(\varphi_l+\Gamma_l' x_q+\nu_{ql}+o_{ql}+\pi_{ql}+\sum_k \omega_{qkl}\right) z_{qil}\right]}{\sum_j \exp\left[\sum_l \left(\varphi_l+\Gamma_l' x_q+\nu_{ql}+o_{ql}+\pi_{ql}+\sum_k \omega_{qkl}\right) z_{qjl}\right]}\right]\times \\
& \left[G\left(\psi_n-\alpha' x_q-\sum_l(\delta_l+\Lambda_l' x_q+\mu_{ql}\pm o_{ql})z_{il}-\sum_k \vartheta_{qk}-\iota_q\right)-G\left(\psi_{n-1}-\alpha' x_q-\sum_l(\delta_l+\Lambda_l' x_q+\mu_{ql}\pm o_{ql})z_{il}-\sum_k \vartheta_{qk}-\iota_q\right)\right]\times \\
& \left[G\left(\tau_m-\beta' x_q-\sum_l(\theta_l+\Delta_l' x_q+\eta_{ql}\pm \pi_{ql})z_{il}-\sum_k \upsilon_{qk}-\iota_q\right)-G\left(\tau_{m-1}-\beta' x_q-\sum_l(\theta_l+\Delta_l' x_q+\eta_{ql}\pm \pi_{ql})z_{il}-\sum_k \upsilon_{qk}-\iota_q\right)\right]\times \\
& \left[\frac{\exp\left[\phi_k' x_q+\hbar_k n_{qi}+\bar{\lambda}_k m_{qi}+\sum_l \left(\chi_{kl}+\Upsilon_{kl}' x_q+\gamma_{qkl}+\sum_k \pm\omega_{qkl}\right) z_{qil}\pm\vartheta_{qk}\pm\upsilon_{qk}\right]}{\sum_k \exp\left[\phi_k' x_q+\hbar_k n_{qi}+\bar{\lambda}_k m_{qi}+\sum_l \left(\chi_{kl}+\Upsilon_{kl}' x_q+\gamma_{qkl}+\sum_k \pm\omega_{qkl}\right) z_{qil}\pm\vartheta_{qk}\pm\upsilon_{qk}\right]}\right]\Bigg\}^{r_{qj}\times a_{qn}\times e_{qm}\times p_{qk}},
\end{aligned}
\tag{3.3}
$$

Finally, the unconditional likelihood function can be computed for individual q as:

$$L_q(\Omega)=\int_{g_q} \left(L_q(\Omega_{-\sigma})\,|\,g_q\right)\, d\boldsymbol{F}(g_q\,|\,\Sigma), \tag{3.4}$$

where $\boldsymbol{F}$ is the multidimensional cumulative normal distribution. The log-likelihood function can be written as: $L(\Omega)=\sum_q \ln L_q(\Omega)$. Simulation techniques are applied to approximate the multidimensional integral in Equation (3.4), and the resulting simulated log-likelihood function is maximized. Specifically, the scrambled Halton sequence (see Bhat, 2003) is used to draw realizations from the population normal distribution. In the current chapter, the sensitivity of parameter estimates was tested with different numbers of scrambled Halton draws per observation, and results were found to be stable with 100 draws.

3.7 MODEL ESTIMATION RESULTS

Estimation results are presented in Tables 3.1 through 3.4, one table for each choice dimension modeled in this chapter. Each of the Sections 3.7.1 through 3.7.4 discuss the estimation results for each choice dimension.

3.7.1 Residential Location Choice Model Component Results

The residential location choice component (whose estimation results are presented in Table 3.1 in next page) of the joint model takes the form of a mixed logit model that accommodates random coefficients (random taste variations).

Virtually all of the findings are consistent with expectations and shed considerable light on the factors influencing residential location, including the presence of residential self-selection, endogeneity, and varying sensitivity to commute time in residential location choice decisions. With respect to zonal size and density measures, it is found that, after including transportation network variables and activity opportunity variables in the model, household density and employment density had no significant impact on residential location choice except for certain demographic segments. For example, households with seniors and children are less likely to choose higher density residential locations (see the negative signs on the interaction variables related to household density in the third and fourth rows of Table 3.1). Similarly, households with higher income levels, belonging to the Caucasian race, and having children are less likely to choose locations of high employment density as their residential location (see the negative signs on interaction variables related to employment density in the last three rows of the block "Residence-end zonal size and density measures" in Table 3.1). Similarly, having a mixed land use configuration including commercial and other non-residential land uses reduces the likelihood of choosing a location as a residential location. All of these variables have negative coefficients associated with them. These findings are generally consistent with the desire of households to live in suburban locations that are virtually exclusively residential neighborhoods of lower density.

Table 3.1. Estimation Results of the Residential Location Choice Component of the Integrated Model

Variables	Parameter	t-stat
Residence-end zonal size and density measures		
Logarithm of number of households in zone	0.838	29.67
Household density (#households per acre x 10^{-1})	0.000	fixed
Interacted with presence of seniors in household	-0.479	-5.82
Interacted with presence of children (of age 5 to 15 years) in HH	-0.217	-2.86
Employment density (#jobs per acre x 10^{-1})	-0.002	-0.80
Interacted with household income greater than $ 90,000 per annum	-0.077	-1.94
Interacted with household belonging to the Caucasian race	-0.034	-1.77
Interacted with presence of children (age 15 years or younger) in HH	-0.011	-1.65
Residence-end zonal land-use structure variables		
Fraction of commercial land area	-0.683	-5.31
Land-use mix	-0.395	-4.47
Residence-end zonal demographics		
Absolute (zonal median income – household income) ($ x 10^{-3})	-0.016	-15.96
Absolute (zonal average household size – household size)	-0.345	-8.90
Average of the median housing value in the zone	-0.133	-10.48
Residence-end zonal race composition measures		
Fraction of African-American population * African-American dummy	3.239	6.83
Fraction of Asian population * Asian dummy variable	2.476	8.26
Fraction of Caucasian population * Caucasian dummy variable	1.713	13.86
Fraction of Hispanic population * Hispanic dummy variable	1.666	3.72
Residence-end zonal activity opportunity variables		
Number of physically active recreation centers	0.027	5.25
Number of natural recreational centers * Number of bicycles in HH	0.046	2.27
Residence-end zonal transportation network measures		
Street block density (number of blocks per square mile x 10^{-1})	-0.031	-5.56
Interacted with household income greater than $ 90,000 per annum	-0.039	-3.89
Bicycling facility density (miles of bike lanes per square mile)	0.018	2.97

Table 3.1 (Continued) Estimation Results of the Residential Location Choice Component of the Integrated Model

Variables	Parameter	t-stat
Household level commute variables		
Total commute time (by auto) of all commuters in the household (minutes)	-0.064	-23.69
Interacted with household income less than $ 35,000 per annum	-0.034	-5.10
Interacted with household income greater than $ 90,000 per annum	0.007	2.31
Standard deviation	0.041	16.29
Standard deviation of the random coefficient capturing common unobserved factors in residential location and bicycle ownership choices (negative correlation)	0.006	1.69
Number of commuters in the household with home and work zones connected by transit within 30 minutes	0.306	6.96
Number of commuters in the household with home and work zones connected by bike route of less than 6 miles length	0.272	7.00

The zonal activity opportunity variables reveal that the availability of destination opportunities contributes significantly to choosing to locate in a zone. The zonal transportation network measures show that street block density is negatively associated with residential location choice, particularly for higher income groups. On the other hand, bicycle facility density has a positive impact on residential location choice. All of these findings are consistent with expectations.

As expected, among the household level commute variables, commute time is negatively associated with residential location choice suggesting that individuals generally try to locate within close proximity of their workplace. However, there is a differential effect according to income levels. This tendency is more (less) pronounced for lower (higher) income households. High income households may be less sensitive to transportation costs because of their ability to afford higher transportation costs, thus making it possible to locate farther away from work, when compared to lower income households. It is found that there is a significant degree of variation in the sensitivity of residential location choice to commute time as evidenced by the significant standard deviation of the random coefficient. This indicates that there are significant unobserved factors influencing household sensitivity of residential location choice to commute time. In addition, the standard deviation of the random coefficient capturing common unobserved factors in residential location choice and bicycle ownership is found to be significant at the 0.09 level with a negative correlation. This suggests that people who like to bicycle (bicycle enthusiasts) and prefer a bicycling-oriented lifestyle are likely to own more bicycles and prefer residential locations close to the work place (reduce commute time). This contributes to the negative correlation in the common unobserved factors (higher bicycle ownership and lower commute time). These lifestyle preferences constitute unobserved variables or factors influencing both bicycle ownership and residential location. It was found that the inclusion of these random coefficients in the model decreased the overall influence of the commute time variable on residential location choice. Finally, modal accessibility for commuters in the household positively impacts residential location choice.

3.7.2 Auto Ownership Choice Model Component Results

The auto ownership component of the model (whose results are presented in Table 3.2 in next page) takes the form of an ordered logit with mixing to account for random taste variations.

Residing in a zone with higher housing or employment density is associated with lower levels of auto ownership. The standard deviation on the household density random coefficient is statistically significant indicating that there is a significant variation in the population in sensitivity of household density in auto ownership choices. Households residing in zones with higher fraction of single family dwelling units are likely to own more cars. With respect to commute time variables, as expected, as the total commute time of commuters in the household increases, auto ownership increases. However, for lower income households, higher auto commute costs are associated with lower auto ownership levels. Higher modal accessibility provided by transit and bicycle is associated with lower levels of auto ownership as evidenced by the negative coefficients on these variables.

Table 3.2. Estimation Results of the Auto Ownership Component of the Integrated Model

Variables	Parameter	t-stat
Household density (#households per acre x 10^{-1})	-0.031	-2.44
Standard deviation	0.056	3.84
Employment density (#employment per acre x 10^{-1})	-0.091	-1.80
Interacted with household income less than $ 35,000 per annum	-0.045	-3.42
Fraction of single family housing units	0.382	1.79
Total commute time (by auto) of all commutes in household (minutes)	0.003	1.67
Total commute cost by (auto) of all commutes in household ($) interacted with household income less than $ 35,000 per annum	-0.260	-2.58
Number of commuters in the household with home and work zones connected by transit within 30 minutes	-0.207	-2.07
Number of commuters in the household with home and work zones connected by bike route of less than 6 miles length	-0.099	-1.33
Density of highways (miles per square mile)	0.063	1.38
Street block density (number of block per square mile x 10^{-1})	-0.023	-1.53
Standard deviation	-0.008	-5.65
Number of zones accessible within 30 minutes by transit	-0.011	-1.97
Standard deviation	0.025	-3.34
Number of zones accessible within 6 miles by bicycle	-0.003	-1.21
Average bicycling facility density (miles of bike lanes per square mile) of the employment locations of all commuters in the household	-0.028	-1.88
Average street block density (number of street blocks per square mile) of the employment locations of all commuters in the household	-0.003	-4.56
Standard deviation	0.002	1.24
Number of active adults	1.688	14.90
Number of senior adults	1.783	11.45
Number of children (of age 5 to 15 years) in household	0.120	2.15
Number of employed individuals	0.772	8.22
Number of physically challenged individuals	-0.877	-5.16
Household income ($ x 10^{-5})	0.548	5.96
Single parent household	-1.168	-4.69
Single individual household	-1.374	-9.40
Age of householder is less than 30 years	-0.252	-1.99
Householder is male	0.279	3.68
Caucasian household	0.143	1.66
Residing in a single-family housing unit	0.782	7.11
Own household dwelling	0.831	7.90
Residing in San Francisco County	-0.366	-1.68

Transportation network measures at the residence end significantly impact auto ownership. Whereas the density of highways positively impacts auto ownership (presumably because it is an auto-oriented transportation network in the area), street block density is negatively associated with auto ownership. Street block density is more indicative of land use density and is likely to signify the potential for undertaking trips by walk or bicycle. The standard deviation on the random coefficient is significant suggesting that there is a significant variance in the population in sensitivity to street block density due to unobserved factors. Higher levels of accessibility by alternative modes contribute to lower levels of auto ownership, although again, there appears to be considerable heterogeneity in population auto ownership sensitivity to transit accessibility as evidenced by the significant standard deviation on the random coefficient associated with that variable. It was found that after including the modal accessibility attributes, the explanatory power of zonal density variables decreased, suggesting that the zonal-density variables may be acting as proxies in explaining auto ownership levels (see Bhat and Guo, 2007 for a similar finding).

The ATE attributes of employment zones such as average bicycling facility density and average street block density negatively impact auto ownership. Presumably these variables signify a greater ability to walk and bike, thus leading to lower car ownership levels for households whose individuals work in such locations. An important finding in the context of the ATE attributes of employment zones is that the explanatory power of transportation network related variables of the residential locations decreased after including the transportation network related variables of work locations. Thus, it is important to consider the ATE attributes of employment location attributes to avoid over estimation of the impact of the ATE of residential location attributes in auto ownership modeling.

A host of socio-economic and demographic variables appear significant in the model. In the absence of such variables in the model, the magnitudes of the coefficients associated with ATE attributes were considerably higher – thus indicating that the impacts of ATE attributes may be overstated when residential self-selection effects due to observed factors are omitted. Number of adults, children, and employed individuals positively impact auto ownership while the number of physically challenged individuals negatively impacts auto ownership. Auto ownership rises with income. Auto ownership is also found to be higher in households where the householder is a male or the household belongs to the Caucasian race. Those residing in a single family dwelling unit or who own their dwelling unit (presumably suggesting higher income suburban households) exhibit higher levels of car ownership even after controlling for other factors. Households residing in San Francisco County show lower levels of car ownership, possibly due to higher levels of transit service, higher densities, and higher parking costs.

The results clearly show the importance of incorporating the influence of observed and unobserved factors (note the significant standard deviations on random coefficients of selected variables including density variables) on auto ownership. In addition to the above findings, there are two significant error correlations that were found to be significant. They are as follows:

- *Common unobserved factors affecting auto ownership propensity and auto mode choice* (standard deviation = 0.4808, t-stat = 3.24, positive correlation, not shown in tables): This error correlation captures the unobserved auto mode choice preferences that can potentially impact auto ownership levels. For example, a person who likes to drive or enjoys a more auto-oriented lifestyle may choose to commute by auto and own more cars. This auto-inclined lifestyle preference constitutes unobserved factors that affect both auto mode utility and auto ownership levels. The explanatory power of the auto ownership variable in the auto modal utility equation (see Table 3.4) decreased after including this error component.
- *Common unobserved factors affecting auto ownership propensity and bicycle ownership propensity* (standard deviation = 0.5821, t-stat = 8.60, positive correlation, not shown in tables): This error component is found to be highly statistically significant. The common unobserved factors could potentially include recreational activity preferences due to which households may own a larger number of vehicles (especially vehicles such as SUVs and vans for recreational purposes) and use them to haul their bicycles for recreational activities. The coefficient (and the t-statistic) on

the bicycle ownership variable in the bicycle modal utility decreased in magnitude after including this error component in the model system.

These findings clearly point to the need to model these choice dimensions simultaneously while accounting for unobserved factors that affect multiple choice processes (error covariances). For example, it is important to accommodate the endogeneity of auto ownership to be able to estimate its "true" impact on mode choices.

3.7.3 Bicycle Ownership Choice Model Component Results

The bicycle ownership model is presented in Table 3.3.

Table 3.3. Estimation Results of the Bicycle Ownership Component of the Integrated Model

Variables	Parameter	t-statistic
Residence-end zonal demographics		
Average household income in the zone	0.438	2.77
Commute-related variables		
Total auto commute time of all commutes in HH (minutes x 10^{-1})	-0.023	-1.59
Standard deviation	0.009	2.42
Standard deviation of the random coefficient capturing common unobserved factors in residential location and bicycle ownership choices (negative correlation)	0.006	1.69
Residence-end local transportation network measures		
Bicycling facility density (miles of bike lanes per square mile)	0.040	3.27
Residence-end zonal activity opportunity variables		
Number of physically active recreation centers	0.016	1.51
Number of natural recreational centers	0.053	1.34
Household demographic variables		
Number of active adults	0.371	5.58
Number of children (of age less than 5 years) in household	0.491	6.63
Number of children (of age 5 to 15 years) in household	1.153	13.44
Number of students in the household	0.332	5.04
Household income (\$ x 10^{-5})	0.464	6.20
Single individual household	-0.386	-3.56
Age of householder is less than 30 years	-0.762	-6.41
Householder is male	0.144	2.35
Caucasian household	0.666	8.86
Residing in a single-family housing unit	0.470	5.68
Own household dwelling	0.302	3.76

Bicycle ownership is found to increase with average household income of the residence-end zone. The commute related variables show that, as commute time of commuters in the household increases, bicycle ownership decreases. This is consistent with expectation; as household members become more auto-dependent, it is unlikely that they will own many bicycles. However, two findings are noteworthy in the context of bicycle ownership sensitivity to commute time. First, there is significant heterogeneity in the population with respect to sensitivity to this variable as evidenced by the significant standard deviation on the random coefficient associated with total commute time. Second, as discussed in the context of the results presented in Table 3.1, the standard deviation of the random coefficient capturing common unobserved factors in residential location and bicycle ownership is significant with a negative correlation. The presence of common unobserved factors affecting sensitivity of residential location choice and bicycle ownership to commute travel time should not be ignored in models of multi-dimensional choice processes. In fact, the inclusion of this error component resulted in a decrease in the magnitude of the coefficient on the commute time variable, and also reduced the statistical significance of the variable. In the absence of this error component, the negative impact of commute time on bicycle ownership propensity would have been over-estimated.

Bicycle facility density and the availability of activity opportunities at the residence-end are positively associated with bicycle ownership. A host of socio-economic and demographic variables influence bicycle ownership suggesting the presence of residential self-selection effects due to observed factors. The number of adults, students, and children positively impact bicycle ownership. Single and young households tend to own fewer bicycles, as expected. Caucasian households, households where the householder is a male, higher income households, households residing in a single-family dwelling unit, and households owning their own housing unit have higher bicycle ownership levels.

The unobserved factors affecting the preferences for bicycling to work and the propensity for bicycle ownership are captured in the common error component between the bicycle ownership propensity and the bicycle mode choice utility. The common unobserved factors affecting bicycle ownership propensity and bicycle mode choice (standard deviation = 0.5165, t-stat = 2.49, positive correlation, not shown in tables) account for bicycle mode choice preferences that can potentially impact bicycle ownership levels. That is, the common error component serves to capture the influence of such common unobserved factors as environmental consciousness and physical fitness orientation, due to which individuals prefer to bicycle to work as well as own more bicycles. In the absence of this error component, the impact of bicycle ownership on bicycle mode choice would have been over-estimated.

3.7.4 Mode Choice Model Component Results

Mode choice model estimation results are presented in Table 3.4 (see next page). Each variable is denoted with the modal utility equation in which it is included.

In general, it is found that several ATE attributes of the residential locations and employment locations (the zonal density measures, land-use structure variables, and transportation system related variables) have a statistically significant influence on mode choice. For example, in the context of residential location-related attributes, it is found that employment density is positively associated with non-motorized modes, and land-use mix is positively associated with transit mode choice. Further, the standard deviation associated with the random coefficient on the employment density variable is statistically significant suggesting that there are unobserved factors that contribute to significant variation in individual sensitivity to this variable for mode choice to work. Among the transportation system variables, the positive impacts of street block density on walk mode, bicycle density on walk and bicycle mode, and transit accessibility on transit mode are all intuitive and expected.

An important finding from this study is the statistically significant impact of the ATE attributes (zonal density and transportation system related variables) of employment zones. For example, household density, and employment density at employment zones contribute positively to the choice of transit and walk modes, and bicycle facility density at the employment zone positively contributes to bicycle mode choice.

Table 3.4. Estimation Results of the Mode Choice Component of the Integrated Model

Variables	Parameter	t-stat
Residence-end zonal size and density measures (including demographic interactions)		
Employment density (#jobs per acre x 10^{-1}) – Bicycle and Walk modes	0.017	1.67
Standard deviation	0.046	2.39
Residence-end zonal land-use structure variables		
Land-use mix – Transit mode	0.621	1.91
Residence-end zonal transportation network measures		
Street block density (number of blocks per square mile x 10^{-1}) – Walk mode	0.041	1.71
Bicycling facility density (miles of bike lanes per square mile) – Bicycle	0.042	1.00
Bicycling facility density (miles of bike lanes per square mile) – Walk mode	0.070	1.85
Number of zones accessible within 30 minutes by transit – Transit mode	0.011	2.67
ATE attributes of employment zones		
Household density (#households per acre x 10^{-1}) – Transit mode	0.015	2.18
Household density (#households per acre x 10^{-1}) – Walk mode	0.021	1.50
Employment density (#jobs per acre x 10^{-1}) – Transit mode	0.017	3.83
Employment density (#jobs per acre x 10^{-1}) – Walk mode	0.012	1.49
Bicycle facility density (miles of bike lanes per square mile) – Bicycle	0.101	2.70
Number of zones accessible within 30 minutes by transit	0.032	9.07
Level of service variables (including demographic interactions)		
Travel time between home and work zones – All modes	-0.031	-8.39
Travel cost between home and work zones – All modes	-0.342	-9.29
Interacted with household income less than $ 35K per annum – All modes	-0.508	-2.15
Home and work zones connected by a bicycle route of length less than 6 miles – Bicycle mode	1.132	3.99
Home and work zones connected by walk route of length less than 3 miles – Walk mode	1.616	3.63
Activity participation and tour-level variables		
Individual undertakes pick-up/drop-off activity – Auto mode	0.842	4.42
Individual undertakes maintenance activity – Auto mode	0.248	1.62
Individual undertakes personal business activity – Auto mode	0.553	3.49
Tour involves at least one non-work stop – Auto mode	1.725	4.92
Household vehicle ownership and bicycle ownership		
Number of vehicles owned by the household – Auto mode	0.435	5.27
Number of bicycles owned by the household – Bicycle mode	0.292	3.79
Individual and household sociodemographic and economic variables		
Individual is a female – Bicycle mode	-1.417	-4.61
Individual has a drivers license – Auto mode	1.681	4.30
Individual has an inflexible work schedule – Auto mode	0.219	1.65

As expected, travel time and cost show negative coefficients for all travel modes suggesting that greater travel times and costs result in a negative utility being associated with any travel mode. The level of sensitivity to travel cost is greater for individuals living in households with low income, presumably due to their higher level of monetary constraints. Where the home and work zones are connected by bicycle or walk routes, individuals are more likely to choose the bicycle and walk modes (as evidence by the statistically significant positive coefficients on these variables).

Activity participation and tour-level variables significantly impact mode choice. Consistent with the findings of Ye *et al.* (2007), the inclusion of non-work stops in a tour contributes positively to the likelihood of choosing the automobile for traveling to work. Similarly, if an individual undertakes maintenance-type activities (serve passenger, maintenance activity such as shopping, and personal business activity) on the travel day, the likelihood of choosing the automobile increases. The personal automobile provides the individual the flexibility needed to undertake these types of activities in a commute tour and hence the positive coefficients associated with these variables are consistent with expectations. In this context, we would like to note that participation in discretionary type of non-work activities such as socializing and recreation was not statistically significant in explaining commute mode choice.

The endogeneity of bicycle ownership and auto ownership to mode choice was discussed earlier in the context of the bicycle and auto ownership model estimation results. The explanatory power of (*i.e.,* the coefficients and t-statistics) of the auto ownership and bicycle ownership variables decreased in magnitude after incorporating the endogeneity of these two variables through common unobserved components between auto mode choice utility and auto ownership propensity, and between bicycle mode choice utility and bicycle ownership propensity. Accounting for such endogeneity allows a more accurate estimation of the impacts of ATE attributes on mode choice, moderated through the influences of such attributes on automobile and bicycle ownership. Finally, several socio-economic/demographic variables are found to be significant in explaining mode choice. Females are less likely to commute by bicycle, while those with a drivers license are more likely to commute by auto. Individuals with inflexible (fixed) work schedules are more likely to use the automobile mode for commuting, presumably due to the need to use a reliable mode of transportation that is generally faster than other modes of transport. The alternative specific constants (not shown in the table) do not have a substantive interpretation in this context because they partially account for the sample range of continuous variables. However, they suggest an overall higher preference for the auto mode.

One of the key findings in the context of the mode choice model is that zonal density variables became either insignificant or less significant (with lower t-statistics) after adding transportation network variables (various modal accessibility variables). For example, the zonal household density and employment density variables in the utility of the transit mode became insignificant after adding these types of transportation (modal) accessibility variables. Similarly, household density in bike and walk modal utilities, and land use mix in walk utility became insignificant after adding street block density to the modal utility equations. Street block density of work zones in the walk and bike utility equations became insignificant after adding the home-work zone connectivity variables (within 6 miles by bike or within 3 miles by walk). These findings appear to suggest that land use density measures serve as proxies for transportation (modal) availability and accessibility, at least to some extent. Once modal availability/accessibility is accounted for, the land use variables themselves offer more modest additional explanatory power. If the modal accessibility variables had been omitted (self-selection effects due to observed factors), then the impacts of land use measures on mode choice could have been potentially grossly over-estimated.

Another key finding is that the explanatory power of transportation network related variables of the residential locations reduced after including the transportation network related variables of work locations. For example, the number of zones accessible by transit within 30 minutes from the residential zone became less statistically significant, and the number of zones accessible by transit within 30 to 60 minutes from the residential zone became insignificant in transit modal utility after adding the number of zones accessible by transit within 30 minutes from the work zone. Similar effects were found with the bicycle facility density variable. Thus, it is important to consider the ATE attributes of employment locations attributes to avoid over estimation of the impact of the ATE of residential location attributes in commute mode choice modeling. From a policy standpoint, this result indicates the importance of improving the accessibility of work locations in combination with that of residential locations, to be able to achieve higher non-auto mode shares.

Tests of the model system suggest that the incorporation of endogeneity and common unobserved heterogeneity across multiple choice dimensions significantly improve model fit. The log-likelihood value at convergence for an independent model system of residential location, auto ownership, bicycle ownership, and mode choice that includes 108 parameters is found to be -37001.06. The corresponding log-likelihood value for the joint model system that includes 119 parameters is -36871.87. The log-likelihood ratio statistic computed based on these figures is found to be statistically significant, suggesting that accounting for the simultaneity and unobserved heterogeneity in these decisions adds significant explanatory power.

3.8 SUMMARY AND CONCLUSIONS

This chapter presents an integrated model system of residential location choice, auto ownership, bicycle ownership, and commute tour mode choice. First, in comparison with the model developed in the previous chapter (*i.e.,* Chapter 2), the integrated model developed in this chapter provides an enhanced modeling methodology to accommodate the intervening impact of medium-term choices (such as auto ownership and bicycle ownership) while modeling the interconnections between long-term decisions (residential location choice in this case) and short-term decisions (travel model choice in this case).

Second, although there is a rich body of literature devoted to the integrated modeling of location choices and travel choices, much of the work in the past has been limited by the analytical and computational complexity associated with estimating truly simultaneous equations systems that include a multitude of discrete choice variables. In addition, there are data limitations that make it difficult to simultaneously model location choice and travel choice decisions within the same framework. This chapter overcomes both of these challenges, realizes the vision of modeling long term, medium term, and short term choices of individuals and households as a true lifestyle package or bundle, and represents a movement toward "truly" integrated land-use travel demand models. The multi-dimensional choice modeling methodology developed and presented in this chapter treats residential location as a multinomial logit, auto and bicycle ownership as ordered logits, and commute tour mode choice as a multinomial logit. The methodology explicitly incorporates a series of behavioral aspects that are critical to simultaneously modeling multiple choice dimensions. These include endogeneity effects (where any one dimension is not exogenous to another, but is really endogenous to the system as a whole), self-selection effects due to observed and unobserved factors (where households locate based on lifestyle and mobility preferences), unobserved heterogeneity (where households and individuals show significant variation in sensitivity to explanatory variables), and correlated error structures (where common unobserved factors significantly and simultaneously impact multiple choice dimensions). The development and estimation of such a model system constitutes an important step forward in the integrated modeling of land use and transportation choices.

Third, all of the above-identified aspects are accommodated in the context of a "tour-based" commute mode choice model as opposed to a "trip-based" mode choice model, which is consistent with directions taken in the development of activity- and tour-based model systems. In summary, the integrated model development and application in this chapter contributes to, and brings together, three different fields of research – integrated land-use travel demand modeling, built environment and travel behavior modeling, and tour-based travel demand modeling.

Model estimation results are found to be plausible and intuitively appealing from a behavioral and transportation systems analysis perspective. First, the model estimation results showed that the impact of land use variables such as zonal-density can be over-estimated when variables describing transportation network and modal accessibility are excluded from the models. Second, the results suggest a significant impact of the work-end built environment variables; specifically the transportation network and density variables. The neglect of work-end built environment variables can result in an overestimation of the impact of residence end built environment variables. This suggests that land-use transportation policies need to modify the land-use and transportation characteristics of both work and residence places in order to achieve desirable shifts in travel behavior. Third, and more importantly, all of the findings speak to the compelling need to model multiple choice dimensions simultaneously while explicitly accounting for the effects identified above, including self-selection and endogeneity. The model estimation results showed the significant presence of common unobserved factors affecting multiple choice dimensions. In the case of residential location choice and bicycle ownership (in the context of household sensitivity to commute travel time) significant unobserved lifestyle factors contribute to higher bicycle ownership as well as to residing in locations that reduce commute travel time, thus leading to a negative error correlation between these two choice dimensions. In simpler terms, this may perhaps indicate that bicycling oriented individuals own more bicycles and stay closer to work places to be able to bicycle to work. More generally, model estimation results showed significant self-selection effects, where individuals with certain modal preferences are found to self-select into residential zones that support their mobility preferences. Further, the results showed significant endogeneity effects, for example, where auto ownership and bicycle ownership are endogenous to mode choice decisions. Also, the results showed significant common unobserved heterogeneity between auto and bicycle ownership decisions.

The findings of this chapter have important implications, both from a modeling perspective and a policy perspective. From a modeling perspective, it is clear that activity-based microsimulation model systems should increasingly move towards the incorporation of integrated models of location and travel choices similar to that developed in this paper. The incorporation of such model systems allows one to account for the various behavioral phenomena and aspects of interest outlined earlier. From a policy perspective, the results make it clear that *not* using a model system to jointly model multiple choice dimensions in a unified framework could lead to erroneous estimates or forecasts of the impacts of changes in built environment attributes on location and travel choices of households and individuals. For example, in this chapter, it is noted that erroneous estimates of built environment effects on mode choice decisions would have been obtained had the endogeneity of auto ownership and bicycle ownership not been recognized explicitly in the model. Erroneous estimates would also be obtained if residential self-selection effects (due to observed or unobserved factors) were ignored. Unfortunately, most model systems in place today treat multi-dimensional choice processes as a series of sequential independent decisions or attempt to connect the various decisions through the use of deeply nested logit models that are quite restrictive with regard to their ability to reflect the various aspects of behavior highlighted in this chapter. Given recent advances in analytical and computational methods and the greater availability of comprehensive data sets that include both land use and travel choices, the time is ripe to begin implementing models of the type developed in this chapter in future model systems. Such efforts would go a long way towards the integrated modeling of long term, medium term, and short term choices (choice continuum).

The model system presented in this chapter can be extended to include additional choice dimensions and endogenous variables, although the inclusion of additional endogenous variables will add computational burden. Nevertheless, future research efforts could focus on further expanding the joint model system presented here or on developing multi-dimensional choice models that incorporate other variables of interest such as trip departure time choice, destination choice, vehicle type choice, and work location choice. An important aspect of interest that needs to be integrated into such multidimensional land-use travel demand models is the individual activity participation and time allocation behavior (see Chapter 5, for such an effort). Attributes of locations (built environment attributes and transportation network attributes) used in this study were zone-level attributes; future research should explore the sensitivity of model results to using alternative aggregations of spatial units for representing built environment attributes and location choice processes.

CHAPTER 4

NEIGHBORHOOD TYPE CHOICE AND BICYCLE OWNERSHIP: HETEROGENEITY IN RESIDENTIAL SELF-SELECTION

4.1 HETEROGENEITY IN RESIDENTIAL SELF-SELECTION

A substantial part of the preceding chapters focused on the phenomenon of residential self-selection. To control for such self-selection effects, the chapters developed a joint modeling methodology in which the residential location choice is jointly modeled with the travel behavior dimension of interest. These joint models allow for the accommodation of residential self-selection due to both observed socio demographics and unobserved attitudes and preferences. This is achieved by the use of common socio-demographic explanatory variables and common random terms in the residential location choice and travel behavior choice equations. The presence of common observed variables and unobserved random terms (across the residential location choice and travel behavior choice models) indicates the extent to which self-selection may be taking place.

In this chapter, the joint modeling approach is further enhanced to account for heterogeneity in residential self-selection effects. For example, although each household (or individual) may have its own life style preferences and corresponding residential self-selection preferences, low income households may face financial deterrents and other constraints (such as housing availability/affordability, market conditions, *etc.*) to self-select more into neighborhoods of their choice, when compared to higher income households. In another example, households with children may have a higher magnitude of residential self-selection preferences (effects) when compared to households without children, because of their desire to provide children with a family-oriented residential environment. The heterogeneity in the jointness of the residential location and travel behavior equations, represented by the heterogeneity in the error covariances, captures such variation among households in residential self-selection effects.

The heterogeneous residential self-selection modeling methodology is applied to the case of a joint residential neighborhood type choice and bicycle ownership level choice model using data from the San Francisco Bay Area. The model takes the form of a joint binary logit and order logit formulation, in which the binary logit component is employed to model the residential neighborhood type choice, and the ordered logit component is used to model the household bicycle ownership level choice. In the context of the residential neighborhood type choice, the analysis in this chapter is different from the analysis in the previous chapters, because a binary neighborhood type choice model replaces the traffic analysis zone (TAZ)-based location choice model employed to model residential choices in the previous chapters. More specifically, in this chapter, the TAZs of the San Francisco Bay Area are classified into bicycle-friendly neighborhoods and less bicycle-friendly neighborhoods using a combination of factor analysis and clustering techniques. This binary variable is used to represent bicycle-friendly and less bicycle-friendly neighborhoods. The residential self-selection effects and corresponding heterogeneity are captured within the context of this binary variable (*i.e.,* in the context of the impact of bicycle-friendly neighborhoods on bicycle ownership levels).

The remainder of the chapter is organized as follows. The next section describes the factor analysis and clustering techniques used to classify the TAZs of the San Francisco Bay Area into bicycle-friendly and less bicycle-friendly neighborhoods. Then the econometric model formulation is presented in Section 4.3. Model estimation results and policy analysis results are discussed in Sections 4.4 and 4.5, respectively. Key conclusions are presented Section 4.6.

4.2 RESIDENTIAL NEIGHBORHOOD TYPE DEFINITION

The San Francisco Bay Area consists of 9 counties and 1099 Traffic Analysis Zones (TAZs) in all. This study uses factor analysis and clustering techniques to define a binary variable that distinguishes the TAZs of the San Francisco Bay Area into bicycle-friendly and less bicycle-friendly neighborhoods. This binary variable is used as a dependent variable in the neighborhood type choice model and as an explanatory variable in the bicycle ownership model to represent bicycle-friendly neighborhoods.

Several studies in the past have used either the clustering technique (see, for example, Song and Knapp, 2004; and Rodriguez *et al.,* 2005) and or the factor analysis method (see, for example, Cervero 1989; Cervero and Kockelman, 1997; Ewing *et al.,* 2002; and Srinivasan, 2002) to categorize residential locations into walking/bicycling friendly neighborhoods. In this context, it is important to note that a multitude of zonal land-use characteristics define the built environment and the bicycle-friendliness of a zone. The attributes include bicycling facilities (zonal bicycle lane density, length of bicycle lanes in the zone), bicycle route network (such as the number of zones accessible by the bicycle route network), other characteristics that may encourage bicycling (for example, the number of natural and physically active recreation centers in the zone), zonal density characteristics (zonal employment, population, and household densities), and the land-use structure (fraction of area under residential, commercial and other land uses, and the land-use mix). Also, many of these attributes may be very significantly correlated to each other (see Cervero and Duncan, 2003). Thus, a combination of both the techniques should be used to come up with the definition of a bicycle-friendly neighborhood. Factor analysis helps in reducing the data (*i.e.,* the various correlated attributes or *factors*) into a manageable number of *principal components* (or variables) that define the built environment of a neighborhood, and the clustering technique helps in using these *principal components* to divide the zones into bicycle-friendly and less bicycle-friendly neighborhoods.

Table 4.1 (in the next page) shows the results of the factor analysis (in the first block of the table) and cluster analysis (in the second block of the table) carried out for the San Francisco Bay Area. The six built environment characteristics (or *factors*) listed in the first column of the table were reduced to two *principal components* using the factor analysis. The factor loadings of the first *component* (in the second column) indicate that this *component* represents the residential density and land-use and the bicycling facilities in a zone. Thus, if a zone exhibits a high value of this *component*, that zone can be labeled as a residential type of zone with good bicycling facilities. Similarly, the second *component* captures zonal characteristics such as number of physically active centers such as sports centers, gymnasiums, and playing fields, *etc.*, and number of natural recreation centers such as parks and gardens that can potentially encourage bicycling. The non-negligible loading (0.357) of the factor "bicycle lane density" on this component supports the notion that such activity centers may be associated with good bicycle facilities. Thus both the *components* represent bicycle-friendliness. The summary statistics indicate that the two *components* have eigen values above 1, and account for 67% of the variability in the six *factors* listed in the table.

Table 4.1 Results of Factor Analysis and Cluster Analysis

Factor Analysis Results: Factor Loadings and Summary*		
Components / **Factors****	Bicycle facilities, residential density and land-use	Activity centers that encourage bicycling
Bicycle lane density (mileage per square mile)	0.513	0.357
Number of zones accessible from the home zone by bicycle	0.784	
Street block density (mileage per square mile)	0.924	
Household population density (per acre)	0.839	
Fraction of residential land use in the zone	0.716	
Number of physically active and natural recreation centers in zone		0.914
Summary statistics		
Eigen value	2.95	1.06
Percentage of variance accounted by the component	49.12	17.71
Cluster Analysis Results: Zonal-level Land Use Characteristics (Averages) by Neighborhood Type		
Neighborhood / **Type**	Bicycle-friendly neighborhood	Less bicycle-friendly neighborhood
Bicycle lane density (mileage per square mile)	4.92	1.88
Number of zones accessible from the home zone by bicycle	60.97	29.31
Street block density (mileage per square mile)	21.00	13.92
Household population density (per acre)	20.70	7.73
Fraction of residential land use in the zone	0.56	0.49
Number of physically active and natural recreational centers in the zone	5.23	1.16

* Principal components estimation and varimax rotation were used in deriving the results
**Factor loadings below 0.35 below are considered insignificant and not shown in the table

After extracting the above mentioned two *components* from the factor analysis, a two-step cluster analysis is employed to divide the 1099 zones of the San Francisco Bay Area into two clusters, based on the two *components*. Subsequently, a descriptive analysis (for all the 1099 TAZs) was undertaken to analyze the zonal land use and bicycle facility characteristics (*i.e.,* the *factors* used in the factor analysis) in the two clusters. Table 4.1 (in the second block) shows the average values of the zonal (or neighborhood) characteristics for the two clusters. Based on these values, the zones belonging to the cluster for which the average values of the *factors* are higher are labeled as bicycle-friendly neighborhoods and the zones belonging to the other cluster are labeled as less bicycle friendly neighborhoods. As can be seen, the bicycle friendly neighborhoods are characterized by better bicycling facilities, better accessibility by bicycle, higher density (street block density, and population density), and a larger number of physically active and natural recreational facilities. The fraction of residential land use was not substantially different across the two clusters. Overall, the neighborhood type definition based on a combination of factor analysis and cluster analysis appears to be intuitive and reasonable. This definition was used as a binary residential neighborhood type choice variable in the estimation of the heterogeneous-joint model. Of the 1099 TAZs, 320 were characterized as bicycle-friendly neighborhoods while the remaining were characterized as less bicycle-friendly neighborhoods. Further, among the 5147 households selected from the five counties of the Southern San Francisco Bay Area for further empirical analysis, a descriptive analysis indicates that 33.6% of the households reside in bicycle-friendly neighborhoods, while the remaining 66.4% of the households reside in less bicycle-friendly/suburban neighborhoods.

4.3 ECONOMETRIC MODELING FRAMEWORK

4.3.1 Model Structure

Let q (q = 1, 2, …, Q) be an index to represent households, and k (k = 1, 2, 3, …, K) be an index to represent bicycle ownership. r_q represents the residential neighborhood type chosen by household q; r_q = 1 if household q chooses a bicycle-friendly neighborhood and r_q = 0 if household q chooses a less bicycle-friendly neighborhood. Using these notational preliminaries, the structure of the residential neighborhood type choice model component is discussed first, the bicycle ownership model component is discussed second, the joint nature of the two components is discussed third, and the heterogeneity in the jointness of the two components is discussed at the end of this subsection.

4.3.1.1 The Residential Neighborhood Type Choice Component

The residential neighborhood type choice component takes the familiar binary logit formulation, as presented below, with r_q as the dependent variable:

$$u_q^* = (\beta' + \gamma_q')\, x_q + \eta_q + \varepsilon_q \,,\ r_q = 1 \text{ if } u_q^* > 0\,;\ r_q = 0 \text{ otherwise} \qquad (5.1)$$

In the equation above, u_q^* is the indirect utility that household q obtains from locating in a bicycle-friendly residential neighborhood, x_q is an (M x 1)-column vector of socio-demographic attributes (including a constant) associated with household q (for example, household size, income, housing type, *etc.*). β represents a corresponding (M x 1)-column vector of mean effects of the elements of x_q on the utility associated with neighborhood choice, while γ_q is another (M x 1)- column vector with its m^{th} element representing unobserved factors specific to household q that moderate the influence of the corresponding m^{th} element of the vector x_q. η_q captures common unobserved factors influencing household q's utility for a neo-traditional/bicycle-friendly neighborhood type choice and the household's bicycle ownership propensity (more details on this later in this subsection). ε_q is an idiosyncratic random error term assumed to be identically and independently standard logistic distributed across individuals q.

4.3.1.2 The Bicycle Ownership Model Component

The household bicycle ownership component takes the ordered logit formulation, as presented below:

$$y_q^* = (\alpha' + \delta_q') z_q \pm \eta_q + (\theta + \mu' w_q + \lambda_q) r_q + \xi_q \,,\; y_q = k \text{ if } \psi_{k-1} < y_q^* < \psi_k \tag{5.2}$$

In the equation above, y_q^* is the latent propensity associated with the bicycle ownership of household *q*. This latent propensity y_q^* is mapped to the actual bicycle ownership level y_q (*i.e.,* the number of bicycles owned by the household) by the ψ thresholds ($\psi_0 = -\infty$ and $\psi_K = \infty$) in the usual ordered-response fashion. z_q is an (*L* x 1) column vector of attributes (not including a constant and not including the household's residential neighborhood type) that influences the propensity associated with bicycle ownership. α is a corresponding (*L* x 1)-column vector of mean effects, and δ_q is another (*L* x 1)-column vector of unobserved factors moderating the influence of attributes in z_q on the bicycle ownership propensity for household *q*. As discussed in the previous section, η_q captures common unobserved factors influencing household *q*'s utility for a neo-traditional/bicycle-friendly neighborhood type choice and the household's bicycle ownership propensity. θ is a scalar constant representing the effect of residential neighborhood type (*i.e.,* r_q) on household bicycle ownership, w_q is a set of household attributes that moderate the effect of residential neighborhood type on household bicycle ownership, and μ is a corresponding vector of coefficients. λ_q is an unobserved component influencing the impact of residential neighborhood type for household *q*, and ξ_q is an idiosyncratic random error term assumed to be identically and independently standard logistic distributed across households.

4.3.1.3 The Joint Model System

The model system allows self-selection of households (based on their bicycle ownership preferences) into neighborhoods based on observed socio-demographics, unobserved preferences and other unobserved factors. This is achieved by the use of common socio-demographic variables and common random terms in the neighborhood type choice and bicycle ownership equations. The presence of common unobserved factors (captured by the common unobserved term η_q across the two equations) leads to the joint nature of the model system.

The '±' sign in front of the η_q term in the bicycle ownership propensity equation indicates that the correlation in the unobserved factors may be positive or negative. If the sign is positive (negative), it implies that individuals who intrinsically have a higher (lower) inclination to reside in bicycle-friendly neighborhoods tend to have a higher bicycle ownership propensity. One can empirically test the models with both '+' and '−' signs to determine the best empirical result.

Finally, to complete the structure of Equations (5.1) and (5.2), it is assumed that the γ_q and δ_q elements, and λ_q and η_q, are independent realizations from normal population distributions; $\gamma_{qm} \sim N(0, \upsilon_m^2)$, $\delta_{ql} \sim N(0, \omega_l^2)$, $\lambda_q \sim N(0, \tau^2)$, and $\eta_q \sim N(0, \sigma^2)$.

4.3.1.4 The Heterogeneous-Joint Model System

The joint nature of the model system may be allowed to vary across households by allowing the magnitude of the common unobserved factors to vary based on household characteristics. That is, the common unobserved η_q term in Equations (5.1) and (5.2) is expressed as $\eta_q = \vartheta_q \exp(\iota + \varpi' v)$, where $\vartheta_q \sim N(0,1)$, ι is a constant, v is a vector of household characteristics, and ϖ is the corresponding coefficient vector. Thus, the joint model system accounting for unobserved heterogeneity in residential self-selection effects can be expressed as:

$$u_q^* = (\beta' + \gamma_q') x_q + \vartheta_q \exp(\iota + \varpi' v) + \varepsilon_q, \; r_q = 1 \text{ if } r_q^* > 0; \; r_q = 0 \text{ otherwise,}$$
$$y_q^* = (\alpha' + \delta_q') z_q \pm \vartheta_q \exp(\iota + \varpi' v) + (\theta + \mu' w_q + \lambda_q) r_q + \xi_q, \; y_q = k \text{ if } \psi_{k-1} < y_q^* < \psi_k \tag{5.3}$$

In the above heterogeneous-joint model formulation, the first equation represents the binary logit model component for the household's choice of residing in a bicycle-friendly neighborhood, while the second equation represents the ordered logit model component for household bicycle ownership. The common $\vartheta_q \exp(\iota + \varpi' v)$ term across both of the model components, which is a function of household characteristics, allows for the possibility that the residential self-selection effects due to common unobserved factors vary across households.

4.3.2 Model Estimation

Let Ω be a vector that includes all of the parameters to be estimated, c_q be a vector that vertically stacks the γ_q and δ_q vectors, and the λ_q and ϑ_q scalars, Σ be another vertically stacked vector of standard errors υ_m, ω_l, and τ. Let $\Omega_{-\Sigma}$ be a vector of all parameters except the standard error terms. Let d_{qk} be a dummy variable taking the value 1 if household *q* owns *k* number of bicycles and 0 otherwise. Finally, let *G*(.) be the cumulative distribution of the standard logistic distribution. Then, the likelihood function, for a given value of $\Omega_{-\Sigma}$ and error vector c_q, may be written for household *q* as:

$$L_q(\Omega_{-\Sigma} \mid c_q) =$$

$$\left\{ \frac{\exp\left[(\beta' + \gamma_q')x_q + \vartheta_q \exp(\iota + \varpi' v)\right]}{\exp\left[(\beta' + \gamma_q')x_q + \vartheta_q \exp(\iota + \varpi' v)\right] + 1} \right\}^{r_q} \left\{ \frac{1}{\exp\left[(\beta' + \gamma_q')x_q + \vartheta_q \exp(\iota + \varpi' v)\right] + 1} \right\}^{1-r_q} \quad (5.4)$$

$$\times \left\{ \begin{array}{l} G\left[\psi_k - \left\{(\alpha' + \delta_q')z_q + (\theta + \mu w_q + \lambda_q)r_q \pm \vartheta_q \exp(\iota + \varpi' v)\right\}\right] \\ - G\left[\psi_{k-1} - \left\{(\alpha' + \delta_q')z_q + (\theta + \mu w_q + \lambda_q)r_q \pm \vartheta_q \exp(\iota + \varpi' v)\right\}\right] \end{array} \right\}^{d_{qk}},$$

The unconditional likelihood function can be computed for household *q* as:

$$L_q(\Omega) = \int_{c_q} (L_q(\Omega_{-\Sigma}) \mid c_q)\, dF(c_q \mid \Sigma),$$

(5.5)

where *F* is the multidimensional cumulative normal distribution. The log-likelihood function for all the households can be written as: $L(\Omega) = \sum_q L_q(\Omega)$. Simulation techniques are applied to approximate the multidimensional integral in Equation (5.5), and the resulting simulated log-likelihood function is used in the maximum likelihood estimation. Gauss matrix programming language was used to code the simulated log-likelihood functions and corresponding simulated gradients.

4.4 MODEL ESTIMATION RESULTS

This section presents a summary of the model estimation results together with key findings and behavioral interpretations that may be drawn from the models. A series of models were estimated, including: (1) A heterogeneous-joint model system of neighborhood type choice and bicycle ownership, (2) A homogenous-joint model system of neighborhood type choice and bicycle ownership, (3) A disjoint (or independent) model system of neighborhood type choice and bicycle ownership, and (4) A disjoint (or independent) model system including only a constant in the neighborhood type choice model and no explanatory variables in the bicycle ownership model. For the sake of brevity, only the first model listed above, *i.e.,* the heterogeneous-joint model system, is presented in this Chapter in its entirety (see Table 4.2 in the next page). Appropriate log-likelihood ratio tests are applied to test the significance of residential self-selection effects and heterogeneity by comparing the model systems listed above.

Model estimation results for the heterogeneous-joint model system are presented in Table 4.2. The first part of the table shows the binary logit model of residential neighborhood type choice (bicycle-friendly neighborhood type choice = 1). The constant does not have a substantive interpretation and is statistically insignificant. Similarly, the age of the householder is statistically insignificant. The weak negative coefficient suggests that more mature households with older householders are less inclined to locate in bicycle-friendly neighborhoods. It is interesting to note that the number of children under 16 years of age, living in a single-family dwelling unit, and owning a house are all negatively associated with choosing to live in a bicycle-friendly neighborhood. The negative coefficients on these three attributes appear to be unintuitive at the first glance, since one may associate households with children, living in a single-family dwelling unit, or owning a house with positive preferences for bicycle-friendly neighborhoods.

TABLE 4.2 Estimation Results of the Heterogeneous-Joint Residential Neighborhood Choice and Bicycle Ownership Choice Model

Variables in the residential neighborhood component[a]	**Parameter**	**t-stat**
Constant	0.1146	0.80
Age of the householder	-0.0033	-1.00
Number of children (of age < 16 years) in the household	-0.1431	-2.91
Household lives in a single family dwelling unit	-0.6030	-6.78
Own house	-0.6224	-6.71
Variables in the bicycle ownership choice component		
Number of active adults in the household	0.3043	5.53
Number of children (of age < 5 years) in the household	0.4224	6.49
Number of children (of age between 5 and 16) in the household	1.0691	15.08
Number of students in the household	0.3220	5.70
Single person household	-0.3047	-3.31
Age of householder greater than 60 years	-0.6381	-6.37
Householder is male	0.1248	2.38
Caucasian household	0.5977	9.69
Household annual Income in 10,000s of dollars	0.4500	7.53
Household lives in a single family dwelling unit	0.3962	5.73
Own household	0.2788	4.03
Household location in a neo-traditional/bicycle-friendly neighborhood	0.1794	2.96
Variables in the standard deviation equation of the common error component between residential neighborhood and bicycle ownership		
Constant	-4.2668	-4.18
Number of children (of age < 5 years) in the household	0.7850	3.63
Number of children (of age between 5 and 16) in the household	1.3818	5.13
Household annual Income less than $35K	0.8230	1.04
Log-Likelihood Measures		
Model	**Log-likelihood**	**No of parameters**
Heterogeneous-joint model	-10259.55	27
Homogenous-joint model	-10275.23	24
Independent models	-10275.32	23
Naive independent models	-11461.52	7

[a] The variables are in the utility equation of bicycle-friendly neighborhood type choice.

This seemingly unintuitive result can be explained as follows. Recall that the neighborhood attributes may be highly correlated with each other (see Cervero and Duncan, 2003). The factor analysis results (Section 4.2) indicate the same in the current empirical setting. The fact that six neighborhood attributes (including bicycle facility density, accessibility by bicycle mode, density measures, land-use type, and the opportunities for recreational activities by bicycle) could be collapsed into just two principal components suggests the extent of correlation between these attributes (see first block of Table 4.1). In other words, as one may observe from the second block of Table 4.2, because of the high correlation among the neighborhood attributes, bicycle-friendly neighborhoods are not only rich in bicycling facilities, connected well by the bicycle transportation network and abundant in opportunities for recreational activities involving bicycles, but also characterized by high density of street blocks and residential population. And it is possible that households with children, living in a single-family dwelling unit, or owning a house stay away from such neighborhoods with higher street block density and higher residential density. In fact, these results point to a notable finding that such households self-select to live in exclusive and sprawling sub-urban neighborhoods which also happen to be less bicycle-friendly.

The ordered-response logit model of bicycle ownership is presented in the second block of Table 4.2. All of the explanatory variables included in the model are statistically significant. Bicycle ownership is positively associated with the number of active adults in the household, the number of children in the household, and the number of students in the household. However, single individuals and older households show a negative tendency towards owning bicycles. Where the householder is male, the household is Caucasian, and the household annual income is high, there is a tendency to own more bicycles. Similarly, residing in a single-family dwelling unit and owning a household are positively associated with bicycle ownership.

In view of the findings with respect to number of children, dwelling unit type, and house ownership in the context of neighborhood type choice (in the previous paragraph), it is interesting to note that such households prefer to live in less bicycle-friendly neighborhoods, but tend to own more number of bicycles. This indicates that even if such households prefer to (and thus, self-select to) live in low density neighborhoods that also happen to be less bicycle-friendly, they do prefer to own (and use) bicycles. It is possible that although the neighborhoods preferred by these households are less bicycle-friendly at a macro scale of geography (a TAZ is a neighborhood here; hence zonal level attributes define the neighborhood type), households living in such neighborhoods may create their own opportunities for bicycling within in their backyards and around their houses (*i.e.,* in the micro scale of geography) to satisfy their bicycling preferences. Thus it is the preferences of these households formed based on such socio-demographics as presence of children, dwelling type and house ownership, that make them own (and use) more bicycles irrespective of the neighborhoods they live in. Another explanation is that in the San Francisco Bay Area, it is likely that the temperate climate and active lifestyle preferences contribute to higher levels of bicycle ownership even in traditional suburban neighborhoods.

The above discussion is not to say that bicycle-friendly neighborhoods have no impact on bicycle ownership. Even after controlling for all other socio-economic and demographic variables, it is found that a bicycle-friendly neighborhood type significantly impacts bicycle ownership in a positive way (see the last variable in the second block of Table 4.2). In fact, the coefficient on the neighborhood type variable ceased to be statistically insignificant when the socio-demographic variables pertaining to the number of children, dwelling type, and house ownership were dropped from the model. This is because ignoring the residential self-selection effects due to these observed attributes (*i.e.,* ignoring their preferences to live in less bicycle-friendly neighborhoods) might have been confounded with the "true" impact of the neighborhood type. This is an important and a notable result in the context of self-selection effects. Almost all of the self-selection studies till date have reported the overestimation of neighborhood effects when the residential preferences due to socio-demographic attributes were not accounted for. In our knowledge, this is the first empirical study that indicates the possibility of an underestimation of neighborhood effects when the residential self-selection is not accounted for. Thus a general finding from this study is that ignoring residential self selection effects can potentially result in biased estimation of the neighborhood effects on travel behavior (bicycle ownership in this case). The bias could be either upward or downward depending on the specific attributes (neighborhood attributes as well as decision maker attributes) under consideration and the travel behavior context at hand.

The third block of the table shows coefficient estimates for variables in the standard deviation equation of the common error component between the residential location choice and bicycle ownership equations (the η_q term). Recall that these variable effects are representative of the heterogeneity in residential self-selection effects due to unobserved factors (*i.e.,* unobserved self-selection effects). It is found that heterogeneity in unobserved residential self-selection effects is primarily due to the presence of children, both young children less than 5 years of age and older children between 5 and 16 years of age. In addition, a modest impact of income on heterogeneity in residential self-selection effects is seen.

The final block of the table presents a comparison of log-likelihood measures for the four different model systems listed earlier in this section. A rather interesting finding from this table is that the homogenous-joint model (not reported in this Chapter) did not show the presence of unobserved residential self-selection effects. A chi-square test between the homogenous-joint model and the independent residential neighborhood type choice and bicycle ownership models with a log-likelihood ratio statistic = 0.18 with 1 degree of freedom rejects the presence of any unobserved residential self-selection in the homogenous-joint model. The heterogeneous-joint model, on the other hand, showed a statistically significant improvement in the log-likelihood. From the table, a chi-square test between the heterogeneous-joint and homogenous-joint models (a log-likelihood ratio statistic = 31.36 with 3 degrees of freedom) suggests the presence of significant variation in the unobserved residential self-selection effects in the population. However, we would like to caution the readers that improvement in the log-likelihood should not be used as sole criterion to determine the presence of self-selection effects.

In order to further assess the presence of unobserved residential self-selection effects and corresponding heterogeneity, the t-statistic of the standard deviation of the common error component (*i.e.,* $\sigma = \exp(\iota + \varpi' v)$) in the heterogeneous-joint model was calculated for different population segments such as households with children, households without children, households with low annual income, and households with high annual income. Since the estimates and the t-statistics of ι and ϖ are known, the t-statistics of σ for each demographic segment (*i.e.,* for corresponding values of v) could be computed in a straightforward manner applying the delta method. Details of the delta method are available in any standard econometrics text book such as Wooldridge (*30*).

The t-statistics of σ for each segment were around 1.0, indicating only a marginally significant presence of residential self-selection effects in each segment. These t-statistics are, however, much higher than that of the t-statistic of σ in the homogenous-joint model (which was close to zero). This is perhaps why there was a statistically significant improvement in the log-likelihood; the net effect of capturing differential residential self-selection effects in different demographic segments might have contributed to the improvement in the log-likelihood. However, the t-statistics are not large enough to indicate a presence of significant heterogeneity in unobserved self-selection effects. The improvement in the log-likelihood and the increase in the t-statistics of σ for each segment supports the notion that one needs to account for heterogeneity in unobserved self-selection effects However, in the absence of significant unobserved residential self-selection effects to begin with (in the homogenous-joint model), it is unclear whether accounting for any further heterogeneity in such effects offers significant advantages in a policy analysis context. Hence, a policy simulation analysis was undertaken to assess if accounting for heterogeneity in residential self-selection would offer any discernible advantages.

The policy simulation analysis (not presented in tables) indicated no significant differences between the results (*i.e.,* the impact of the neighborhood type on bicycle ownership) of an independent bicycle ownership model (*i.e.,* a model with no residential self-selection effects) and those of the heterogeneous-joint model (*i.e.,* a model with heterogeneous residential self-selection effects). The elasticity effects were not different even for the different socio-demographic segments (such as households with children and households without children) for which the model estimation results indicated a heterogeneity in self-selection. This indicates that in the current empirical context, there are no significant unobserved residential self-selection effects. Although there was a significant improvement in the log-likelihood due to capturing heterogeneity in unobserved residential self-selection effects, the effect of self-selection is itself not significant enough to gain much advantage from capturing any further heterogeneity. This may be because the residential self-selection preferences are already captured in the bicycle ownership model by including such socio-demographic variables as number of children, dwelling type, and house ownership. Thus a simple ordered logit model would have sufficed to estimate the impact of neighborhood type on bicycle ownership. However, it is important, at the least, to test for the presence of residential self-selection and the heterogeneity in self-selection before using a simple ordered logit model for bicycle ownership.

4.5 SUMMARY AND CONCLUSIONS

This paper presents a joint model of residential neighborhood type choice and bicycle ownership. The model isolates the true causal effects of the neighborhood attributes on household bicycle ownership from spurious association due to residential self-selection effects. The joint model accounts for residential self-selection due to both observed socio-demographic characteristics and unobserved preferences. In addition, the model allows for differential residential self-selection effects (*i.e.,* heterogeneity in residential self-selection effects) across different socio-demographic segments. The model is estimated using a sample of more than 5000 households from the San Francisco Bay Area.

The model results show a substantial presence of residential self-selection effects due to observed socio-demographics such as number of children, dwelling type, and house ownership. An important finding in this context is that ignoring such self-selection effects may lead to severe underestimation of the impact of bicycle-friendly neighborhood type on bicycle ownership. This is the first empirical study that reports the danger of under estimation of the neighborhood type impact on travel related behavior (bicycle ownership choice, in this Chapter). All other self-selection studies till date appear to have reported an overestimation of the neighborhood impacts due to ignoring residential self-selection effects. Thus, it is shown for the first time in the self-selection literature that ignoring such observed self-selection effects may not always lead to overestimation of the impact of neighborhood attributes on travel related choices such as bicycle ownership. In the current context, ignoring self-selection due to socio-demographic attributes resulted in an underestimation of the impact of neighborhood attributes on bicycle ownership. In the context of unobserved factors, no significant self-selection effects were found. In the current empirical context, it appears that the observed variables included in the model may have captured many of the self-selection effects and it is questionable whether additional benefits would be obtained by accounting for unobserved residential self-selection effects or any further heterogeneity in such effects. However, it is recommended to test for such effects as well as heterogeneity in such effects before concluding that there are no unobserved factors contributing to residential self-selection.

Future research in this arena should focus on examining the consistency of this finding across multiple geographic contexts and associated data sets. In addition, the spatial unit of analysis used in the context of residential neighborhood type choice modeling (the traffic analysis zone is used in this Chapter) merits further consideration to examine whether the findings are robust and consistent even if different levels of spatial resolution are used to model residential neighborhood type choice. The binary neighborhood type choice modeling methodology presented in this chapter suits very well for such an analysis at different scales of geographic resolution (as opposed to a multinomial residential location choice modeling approach, which would become computationally unfeasibly intensive with finer spatial resolutions).

CHAPTER 5

INCORPORATING RESIDENTIAL SELF-SELECTION EFFECTS IN ACTIVITY TIME-USE BEHAVIOR: FORMULATION AND APPLICATION OF A JOINT MIXED MULTINOMIAL LOGIT – MULTIPLE DISCRETE CONTINOUS CHOICE MODEL

5.1 ACTIVITY-BASED MODELING AND ACTIVITY TIME-USE BEHAVIOR

5.1.1 Activity-based Travel Demand Modeling and Time-use Behavior

The activity-based approach to travel demand modeling views travel as a demand derived from the need to pursue activities distributed in time and space (see Jones *et al.,* 1990; and Bhat and Koppelman, 1999). The conceptual and behavioral appeal of this approach originates from the recognition that the need and desire to participate in activities is more basic than the travel that some of these participations may entail (Bhat *et al.,* 2004). Other salient features of the activity-based approach include: (1) The focus on entire sequences and patterns of activities and travel (using the whole day or longer periods of time as the unit of analysis) rather than individual trips, (2) The recognition of the linkages among activity-travel decisions of an individual across different time periods of the day, (3) The explicit modeling of the temporal dimension of activity participations and travel, (4) the accommodation of space-time interactions in activities and travel. In essence, the activity-based approach to travel analysis attempts to better understand the behavioral basis for individual decisions regarding participation in activities in certain places at given times (and hence the resulting travel needs). This behavioral basis includes all the factors that influence the why, how, when and where of performed activities and resulting travel. Among these factors are the needs, preferences, prejudices, and habits of individuals (and households), the cultural/social norms of the community, and the travel service characteristics of the surrounding environment.

All of the above-mentioned aspects of the activity-based approach may be traced back to a single fundamental difference between the trip-based approach and the activity-based approach, which is in the way time is conceptualized and represented in the two approaches (see Pas, 1996, Pas and Harvey, 1997, and Bhat and Koppelman, 1999). The appropriate incorporation of time is perhaps the most important prerequisite to accurately forecasting activity-travel patterns. In the trip-based approach, "*time is reduced to being simply a "cost" of making a trip*" (Bhat and Koppelman, 1999). The activity-based approach, on the other hand, treats time as an *"all-encompassing continuous entity"* within which individuals make activity/travel participation decisions (see Kurani and Lee-Gosselin, 1996). Thus, the central basis of the activity-based approach is that individuals' activity-travel patterns are a result of their time-use decisions. Individuals have 24 hours in a day (or multiples of 24 hours for longer periods of time) and decide how to use that time among activities and travel (and with whom) subject to their sociodemographic, spatial, temporal, transportation system, and other contextual constraints. In the activity-based approach, the impact of land-use and demand management policies on time-use behavior is an important precursor step to assessing the impact of such polices on individual travel behavior. For example, one may analyze whether improving a neighborhood with walkways, bikeways, and recreational parks encourages individuals to invest more time in physically active recreation pursuits in the place of in-home passive recreation (such as watching television or playing computer games). The travel dimensions then can be "derived" from the changes in time-use and activity-scheduling patterns.

To be sure, there have been several studies of individual time-use in the past decade. However, most of these studies have focused on understanding time-use patterns as a function of individual and household sociodemographics (see, for example, Bhat and Misra, 1999, Chen and Mokhtarian, 2006, Gliebe and Koppelman, 2002, Golob and McNally, 1995, Goulias and Henson, 2006, Harvey and Taylor, 2000, Kapur and Bhat, 2007, Kraan, 1996, Levinson, 1999, Lu and Pas, 1997 and 1999, Meloni *et al.,* 2004, Meloni *et al.,* 2007, Yamamoto and Kitamura, 1999, and Ye and Pendyala, 2005). Such time-use studies, while contributing to the literature in important ways, are unable to examine the impact of land-use policies. Some more recent studies by Bhat and colleagues do include the impact of a comprehensive set of land use attributes in their time-use analyses (see, for example, Bhat, 2005, Bhat *et al.,* 2006, Copperman and Bhat, 2007). However, these studies are focused on weekend day time-use behavior and not weekday time-use behavior. Besides, an important limitation of all these earlier activity-based time-use studies is that they do not consider residential sorting effects when assessing the impact of land use attributes. Rather, they assume land use as being pre-determined and exogenous. The discussion below expands on this issue.

5.1.2 Residential Self-Selection in Activity Time-Use Behavior

In the activity-based travel demand models developed to date, land-use attributes are considered pre-determined and exogenous, and are used as independent variables to explain travel behavior. Such an approach implicitly assumes a one-directional relationship between land-use and travel demand. In the past decade, there has been an increasing amount of research focused on revisiting this simplistic assumption, and considering more complex inter-relationships that may exist between land-use and travel demand. At the center of this debate and investigation is whether any effect of land-use on travel demand is causal or merely associative (or some combination of the two). To understand this issue better in the context of activity time-use behavior, consider a land-use policy to improve bicycling facilities, with the objective of reducing automobile dependence and increasing physically active recreational pursuits. To assess the impacts of such a policy, assume that a data collection effort has been undertaken to examine the bicycling levels of individuals in neighborhoods with different levels of existing bicycling facilities. An analysis of this data may find that individuals residing in neighborhoods with good bicycling facilities pursue more bicycling-related activities. The question is whether this relationship implies that building neighborhoods with good bicycling facilities would result in higher bicycling levels in the overall population (*i.e.,* a causal relationship), or whether this relationship is an artifact of individuals who are bicycling-inclined (because of, say, physical fitness consciousness) self-selecting themselves to reside in neighborhoods with good bicycling facilities (*i.e.,* an associative relationship). If the latter "residential sorting process" is at work, building neighborhoods with good bicycling facilities would not result in higher bicycling levels in the overall population, but simply lead to an alteration of spatial residence patterns of the population based on physical fitness consciousness.

In reality, the nature of the relationship between land-use and travel behavior may be part causal and part associative. Thus, any attempt to examine the land use-travel behavior connection should disentangle the causal and associative elements of the relationship to inform and contribute to the credible assessment of the impact of land-use policies on travel behavior. Several recent studies have alluded to and/or accommodated this residential sorting process in one of several ways. However, all these earlier studies study the land use and travel behavior relationship by directly focusing on specific travel behavior dimensions, such as trip frequency or trip mileage for one or more trip purposes. Essentially, these earlier studies ignore the importance of understanding the land-use activity time-use relationship (and corresponding residential self-selection effects) as a precursor step to accurately assessing the impact of land-use polices on individual travel behavior. To our knowledge, no study to date has addressed the issue of residential self-selection in conjunction with the activity-based approach to modeling travel demand; especially so in the context of activity time-use behavior.

5.1.3 Contribution and Organization of the Chapter

In view of the preceding discussion, from a substantive viewpoint, this chapter is an effort toward bringing together activity-based travel demand modeling research and built environment and travel behavior modeling research. The chapter develops an integrated model of residential location choice and activity participation and time-use (or activity time-use) that can be used to account for residential self-selection effects while estimating the impacts of the built environment on daily activity participation and time-use behavior. Specifically, the model accommodates residential sorting effects due to observed and unobserved individual characteristics in examining the impact of activity-travel environment (ATE) variables on individual time-use in maintenance activity (grocery shopping, household chores, personal care, *etc.*) and several types of discretionary activity purposes. A salient feature of the model is that the self-selection effects are accommodated simultaneously in the activity participation and time-use behavior (or the activity time-use behavior) in multiple types of activities. That is, the model recognizes the possibility that an individual can participate in multiple types of activities (as opposed to a single activity engagement) in a given period of time. The multiple activity participation and time-use behavior (*i.e.,* the activity time-use behavior) comes under what has recently come to be labeled as multiple discrete-continuous choice making behavior (see Section 5.2 for further details). Hence, to accommodate and model residential self-selection effects in activity time-use behavior, the residential choice-activity time use model system in the chapter takes the form of a joint mixed multinomial logit – multiple discrete-continuous extreme value (MNL–MDCEV) model. To our knowledge, this is the first instance in the econometric or other literature of the development of such a model to jointly analyze an unordered discrete variable (residential location choice in the current context) and multiple discrete-continuous variables (activity participations and time-use decisions in multiple activities in the current context).

The remainder of this chapter is organized as follows. The next section (*i.e.,* Section 5.2) provides an overview of multiple discrete continuous choice situations, features of such choice situations, and a review of the literature on the methods available to model such choice situations. At the end of this section, it will become clear why the MDCEV model structure is used to model activity time-use behavior component of the joint residential choice-activity time use model system in this chapter. Section 5.3 presents the mathematical structure of the joint model and the estimation procedure. Section 5.4 discusses the data sources and describes the sample used in the analysis. Section 5.5 focuses on the empirical results. Finally, Section 5.6 concludes the chapter by summarizing important findings and identifying directions for future research.

5.2 ACTIVITY TIME-USE AS A MULTIPLE DISCRETE-CONTINUOUS CHOICE

5.2.1 Background

A number of consumer demand choice situations are characterized by the simultaneous choice of multiple alternatives (or goods, or services[23]), as opposed to a single alternative. That is, the decision-maker can potentially choose one or more alternatives from a set of available alternatives. Such choice situations have now come to be labeled as "multiple discrete" choice situations. In addition to the discrete choices, there can also be a continuous choice corresponding to the amount of consumption (or allocation of resources such as time and money for consumption) of each chosen alternative. This leads to the case of "multiple discrete-continuous" choice situations. Individual activity type choice and the duration of activity participation (*i.e.,* time-use) decisions come under the category of multiple discrete-continuous choice making. An individual may participate in different types of discretionary (or leisure) activities such as social activities, recreational activities, *etc,* within a given time period such as a day or a week. In addition to the activity participation decisions, the individual makes decisions on the amount of time allocated to each type of activity.[24]

Most of the literature in the choice modeling field is confined to analyzing the choice of a single discrete alternative from a set of mutually exclusive and collectively exhaustive alternatives. Such single discrete choice modeling approaches are not well suited to modeling multiple discrete-continuous choice situations, because they do not

---

[23] We use the terms alternatives, goods, and services interchangeably.

[24] Another example of such a choice situation is a household's choice of vehicle ownership and usage (Sen and Bhat, 2006). A household may own a mix of different types of vehicles (for example, a sedan, a pickup, a van, *etc.*) and use them to different extents based on the different needs, preferences and ownership and operating costs. Other examples include: (1) Household decisions pertaining to the expenditure allocation for different recreational purposes, and (2) Individual decisions on the different types of long distance recreational trips and the amount of money spent for those trips over a period of a month or a year. Of course, several other situations from disciplines such as economics, revenue management, and marketing can be characterized as multiple discrete-continuous choice situations. Examples include stock and bond portfolio choice of an investor, brand choice and consumption quantity of grocery items purchased by a grocery shopper, and the vehicle fleet mix and usage decisions of a courier service firm such as FedEx.

capture several distinct features underlying consumers' multiple discrete-continuous choice decisions. We discuss these distinct features in the context of the participation of individuals in different types of activities, and the time allocation to each type of activity, in Section 5.2.2. Subsequently, in Section 5.2.3, we provide a review of the literature on the methods available to model such choice situations.

5.2.2 Features of Multiple Discrete-Continuous Choice Making Situations

The important features of consumer behavior that need to be considered while analyzing the multiple discrete-continuous choice situation of activity time allocation are: (1) Satiation, (2) Resource allocation, (3) Substitutability of alternatives, and (4) Simultaneity of Decisions. All of these features together give rise to the choice of potentially multiple alternatives in a single choice occasion. Each of the features is discussed below within a general context of consumer choice behavior, while examples are provided for the context of activity participation and time-use decisions.

5.2.2.1 Satiation

According to the utility theory of consumption, consumers derive *utility* from consuming goods/services (or alternatives). Utility is a measure of *happiness* or *satisfaction* gained by consumption. In general, as the consumption of a particular good/service increases, the utility accrued increases at a decreasing rate. That is, as the consumption increases, the utility accrued from an additional unit of consumption decreases. This is also called as the law of diminishing marginal returns. Satiation effects can be defined as the diminishing marginal returns from the consumption of an alternative (good/service) as the consumption of that alternative increases (Kim *et al.,* 2002; and Bhat 2005).

Consider the example of leisure (or discretionary) activity participation and time-use decisions. Individuals participate in various types of leisure activities such as relaxation, socialization, and recreation, *etc.* All the time available for leisure is not usually spent for just a single type of activity, because individuals experience satiation effects as the time consumption for a particular type of activity increases. Because of such satiation effects, individuals participate in multiple types of activities.

5.2.2.2 Resource Allocation

In general, multiple discrete-continuous choice making situations involve a finite amount of resources (*i.e.,* a budget) under which the decision-makers operate. The decision making mechanism is driven by an allocation of the limited amount of resources such as time and/or money to consume goods/services/alternatives. As can be seen, the activity participation and time-use decisions clearly involve decisions regarding the allocation of time to various activities.

5.2.2.3 Substitutability of Alterantives

Sen (2006) describes substitutability of alternatives as an important aspect to characterize a choice situation. The term *substitutability of alternatives* refers to the extent to which the consumption of an alternative can be substituted by the consumption of another alternative. To understand this, consider the two extreme cases of substitutability: (1) Perfect substitutability and (2) No substitutability. If the alternatives in a subset are perfect substitutes to each other, then the consumption of one of the alternatives precludes the consumption of all other alternatives in that subset. Such perfect substitutable alternatives are also called mutually exclusive alternatives. In such cases, only one alternative that offers the maximum utility is chosen from the subset of the perfectly substitutable alternatives[25]. On the other hand, if the alternatives in a subset cannot be substituted to each other, then all of the alternatives in the subset must be consumed. In general, the substitutability of alternatives can lie any where in the continuum between perfect substitutability to no substitutability. The alternatives that do not belong to either of the extremes on the continuum of substitutability are called imperfect substitutes. It is because of the imperfect substitutability that multiple alternatives are consumed as opposed to a single alternative.

Perfect substitutability arises because of the same need served by different alternatives. In such situations, the alternative that offers the maximum utility is consumed. On the other hand, imperfect substitutability arises due to the different needs

[25] It is this premise that most of the single discrete choice models are built upon.

served by the alternatives. Consumer choice situations may involve alternatives with different levels of substitutability. Some of the alternatives may be perfect substitutes to each other, while others may be imperfect substitutes. Along with the perfect and imperfect substitutes, several consumer choices may include essential alternatives also. Essential alternatives are not substitutable by other alternatives, because of the compulsory nature of the needs they serve. For the same reason, a non-zero amount of consumption may be necessary in the case of essential (non-substitutable) alternatives.

In the case of daily activity time-use behavior, leisure (or discretionary) activity participation and time use decisions involve imperfect substitutes. That is, all the leisure (or discretionary) activities (relaxation, socialization, physically active recreation, physically passive recreation, *etc.*) are imperfect substitutes to each other. It is possible that these activities, when categorized in a fine manner, constitute of perfect substitutes. Consider, for example, the physically active recreation activity such as exercising, jogging, swimming, tennis, *etc.* Individuals may participate in only one of the physically active recreation activity on a day. Thus, the different types of physically active recreation activities are perfect substitutes to each other. Further, barring unusual cases, individuals' daily activity decisions may also involve certain essential alternatives, such as sleep and maintenance activities. Individuals must participate in these activities in order to satisfy basic physiological needs. In this chapter, we will deal with one essential alternative (maintenance activity), and a number of discretionary alternatives that are imperfect substitutes to each other.

5.2.2.4 Simultaneity of the Decisions

Multiple discrete-continuous choice situations are characterized by simultaneity in decision-making. Especially, there may be a high degree of jointness between the discrete and continuous choices. Consider for example, the case of an individual's daily activity participation and time-use decisions. An individual chooses to participate in different types of activities in a day. The amount of time (s)he chooses to allocate to each of the activities is dependent on the types of activities (s)he participates in. Hence, the discrete and continuous choice components in a multiple discrete-continuous activity participation and time-use choice situation are closely tied to each other.

5.2.3 Review of the Literature on Modeling Multiple Discrete/Discrete-Continuous Choices

5.2.3.1 Classification of the Literature

The multiple discrete/discrete-continuous choice models developed to date can be broadly classified into two categories: (1) The multivariate single discrete/discrete-continuous choice models, and (2) The utility-maximization-based multiple discrete-continuous choice models. The multivariate single discrete/discrete-continuous choice models, as the name suggests, are used to analyze only the discrete choices. However, these methods can be (and have been) extended to include continuous choice components. The utility maximizing models, on the other hand, have been used to model both the discrete and the continuous components jointly. The following section provides a review of the multivariate single discrete/discrete-continuous choice models and the subsequent section provides a review of the utility-maximization models.

5.2.3.2 Multivariate Single Discrete/Discrete-Continuous Choice Models

Multivariate single discrete choice methods have been used mostly in the field of marketing research to model multiple discrete choice behavior. Based on the nature of the outcome variables, the models can be categorized into multivariate binary choice models, and multivariate ordered response models.

Multivariate binary choice models, as the name suggests, comprise of two or more binary choice components that are jointly modeled. Each of the binary choice components are used to determine if an alternative is chosen by the decision-maker. Thus the number of binary choice components is equal to the number of choice alternatives under consideration. Different distributional assumptions of the stochastic components give rise to different model structures. The two widely used model structures are the multivariate binary probit (based on a joint normal distribution), and the multivariate binary logit (based on a joint logistic distribution, originally formulated by Cox, 1972) models. Examples of multivariate binary probit applications include Greene (1997), Chib and Greenberg (1998), Manchanda *et al.* (1999), Chib *et al.* (2002), Deepak *et al.* (2002), Edwards and Allenby 2003; and Chib *et al.* (2005). Examples of multivariate binary logit models include Ma and Seetharaman (2004), Russel and Peterson (2000), and Ma *et al.* (2005).

The multivariate ordered response models, as the name suggests, comprise of multiple ordered response models that are tied together using multivariate distributional assumptions on the stochastic terms. This framework is typically used to model the integer-quantity outcomes of multiple choice alternatives (for example, the number of vehicles of each type owned by a household). The ordered response structure can be either an ordered probit/logit structure or a count data model (Poisson, negative binomial, *etc.*) structure. Applications of the multivariate ordered probit structure include Bhat and Srinivasan (2005), and applications of the multivariate count data model structure include Terza and Wilson, 1990; and Russel and Kakamura, 1997.

A major drawback with the above-mentioned models (*i.e.,* the multivariate single discrete/discrete-continuous models) is that they are not associated with closed-form likelihood expressions, because the individual probability expressions involve multi-dimensional integrals that are not analytically tractable. Thus, the analyst has to resort to simulation techniques to evaluate the probability expressions and carryout the estimation. Moreover, it is cumbersome to extend these models to incorporate continuous components of choice (see Srinivasan and Bhat, 2006 for a multiple binary logit/linear regression model). Further, the over all choice-making framework of these models does not fall under the realm of the well established theory of random utility maximization. As stated by Bhat (2005), the multivariate single discrete/discrete-continuous choice modeling methods are a result of a mere "statistical stitching" of several single discrete choice models.

5.2.3.3 Utility Maximization Frameworks

In the recent past, there have been an increasing number of multiple discrete-continuous choice analyses using the utility maximization framework. The central idea of the utility maximization framework is that the consumers operate with in a limited amount of resources (*i.e.,* a budget) and make their consumption choices in order to maximize the utility of consumption. Thus, a multiple discrete-continuous choice situation can be viewed as a consumption-utility maximization problem with constraints on the overall budget available for consumption. In other words, a discrete-continuous choice situation can be viewed as a resource allocation (to various goods or alternatives) problem. More formally, this approach assumes a direct utility function with respect to the consumption quantity vector. The consumers are assumed to maximize the utility function with respect to a budget constraint, and non-negativity constraints on the amount of consumption of each good/alternative.

Within the utility maximization framework, Wales and Woodland (1983) proposed two alternative ways of dealing with multiple discrete-continuous choice situations. They label the first approach as the Amemiya-Tobin approach and the second approach as the Kuhn Tucker (or KT) approach. The two approaches differ in the manner in which stochasticity is introduced. The Kuhn Tucker approach (which was also used by Hanemann, 1978 under the label of "generalized corner solution problem") has gained wider attention than the former approach, because of its behavioral consistency. We first provide a brief overview of the Amemiya-Tobin approach and then turn to a detailed review of the literature that employed the Kuhn Tucker approach.

The Amemiya-Tobin approach proposed by Wales and Woodland is an extension of the work by Amemiya (1958) and Tobin (1974). According to this approach, the direct utility is assumed to be deterministic. Subsequently, the utility maximization problem (taking into consideration the budget constraint) is solved to obtain the deterministic consumption shares. Stochasticity is then introduced into the deterministic consumption shares obtained from the deterministic utility maximization process. This stochasticity is assumed to allow for the possibility that consumers may make errors in the utility maximization process, and that there may be unobserved (to the analyst) factors affecting the consumption shares. Wales and Woodland extended this approach by ensuring that (a) the stochastic consumption shares are non-negative and less than one, (b) they add up to one, and (c) corner solutions (*i.e.,* the possibility that some alternatives are not chosen or consumed) are allowed. The primary drawback of this approach is that the corner solutions are not incorporated in a behaviorally consistent manner.

According to the KT approach, the utility function is assumed to be random over the population. Subsequently, the first order Kuhn Tucker (1951) conditions are used to derive the consumption quantities by solving the random utility maximization problem subject to the linear budget constraint and the non-negativity constraints. The KT conditions include the stochastic conditions for zero consumption (corner solutions) of some goods and positive consumption of other goods. These stochastic KT conditions can be used to derive the probabilities of the consumption (including zero consumption of some goods). Thus, the KT approach is a unified structural framework to simultaneously determine the consumption quantities while allowing for corner solutions. However, the stochastic distributional assumptions used in the past (till about six years before) lead to multi-dimensional integrals with no closed-form expressions. One has to resort to simulation techniques to evaluate these integrals. Thus, there have been only a few applications using the KT framework until recently (see Bockstael *et al.,* 1987; Lee and Pit, 1986, and Srinivasan and Winer, 1994 for earlier works using the KT approach[26] . Also, see Kim *et al.,* 2002 for a recent application using the Geweke-Hajivassiliou-Keane technique to evaluate the truncated multi-dimensional normal integral resulting from a KT-based model multivariate normal stochastic assumption). It is only in the recent past that the KT approach has gained attention; especially in the Environmental Economics and the Transportation fields.

In the Environmental Economics field, Phaneuf *et al.* (2000) used a type-I extreme value stochastic specification of the consumption utility structure within the KT-based utility maximization framework. The type-I extreme value stochastic structure, when used in the KT-based constrained utility maximization framework, gives rise to analytically tractable KT first-order conditions and probability expressions. Following

[26] Bockstael *et al.,* 1987; Lee and Pit, 1986a extended the KT approach (a primal problem with direct utility specification) by reformulating the utility maximization problem as a dual problem with indirect utility specification. See Phaneuf, 1999; and Wang 2003 for recent efforts on the dual KT approach.

this specification, a number of empirical studies were undertaken in the area of Environmental Economics[27]. These studies include Phaneuf and Herriges (2000), von Haefen (2003), Herriges *et al.* (2004), and von Haefen and Phaneuf (2005). However, a drawback of these models is that stochasticity is introduced in the sub-utility functions of only the non-essential goods. That is, there is no stochasticity in the sub-utility function of the essential good[28]. There is no intuitive (or theoretical) basis for assuming such a deterministic utility contribution from the consumption of the essential good. Further, as illustrated by Bhat (2008), this assumption can lead to incorrect probability calculations. Moreover, such a specification requires the presence of at least one essential alternative (or good) in the choice set.

A recent and significant development in the area of multiple discrete-continuous choice modeling research is the formulation of the Multiple Discrete-Continuous Extreme Value (MDCEV) Model by Bhat (see Bhat 2005 and Bhat 2007). The MDCEV formulation is based on the generalized variant of a translated constant elasticity of substitution (CES) utility function (also used by Kim *et al.,* 2002) with multiplicative log-extreme value error terms. It is to be noted here that, unlike in the Phaneuf *et al.* (2000) framework, stochasticity is introduced into the sub-utility of each and every alternative (including the essential alternatives). The probability expressions resulting from the MDCEV model structure are analytically tractable. In fact, the MDCEV model has a remarkable property that the probability expressions simplify to the well known multinomial logit (MNL) probabilities when each and every consumer chooses only one alternative. Further, it is straight forward to extend the MDCEV model to incorporate random coefficients and/or error term correlations.

Following the MDCEV model formulation, Bhat and colleagues (Bhat and Sen, 2005; Bhat *et al.,* 2006; Sen and Bhat, 2006; Kapur and Bhat, 2007; and Sener and bhat, 2007) undertook several multiple discrete-continuous choice analyses for different empirical contexts on activity participation and time-use. Among these studies, Bhat *et*

[27] Habib and Miller (2006) used the Phaneuf *et al.* (2000) specification in Transportation research.
[28] Bhat (2006) illustrated that this assumption is not innocuous.

al. (2006) formulated a joint MDCEV-MNL model to analyze the perfect and imperfect substitutable alternatives in a unified framework. Sen and Bhat (2006) and Sen (2006) provided a compelling application of the joint MDCEV-MNL model for analyzing household vehicle ownership and usage. Kapur and Bhat (2007) and Sener and bhat (2007) used the MDCEV modeling framework to analyze multiple discrete-continuous outcomes in multiple dimensions (*i.e.,* they analyzed multiple discrete-continuous choices of time allocation by activity purpose and accompaniment).

Another important study by Bhat (2008) provides a reformulated functional form for the utility specification of the MDCEV model, in order to identify the distinguished role played by each parameter of the utility function. In this chapter, we use the MDCEV model formulation of Bhat (2008) to model the activity participation and time-use component of the joint residential location and activity time-use model system. Such a joint model system is used to accommodate residential self-selection effects in activity time-use behavior.

5.2.3.3 Summary of the Literature

The multiple discrete/discrete-continuous choice models available to date can be broadly classified into two categories: (1) The multivariate single discrete/discrete-continuous choice models, and (2) The utility maximization models. The multivariate single discrete/discrete-continuous choice models are not based on a unifying theory of utility maximization. On the other hand, the Kuhn Tucker (KT) based models, which are based on the theory of random utility maximization, have gained increasingly high attention in the recent past. In the past five years, there have been important contributions in this area from the Transportation, Marketing, and Environmental Economics fields. A recent and very remarkable development is the formulation of the MDCEV model by Bhat (2005, and 2008). The MDCEV model is elegant in that it is based on a unified KT based random utility maximization approach, and yields closed-form probability expressions. Further, it simplifies to the well-known MNL model when all the decision makers choose only one alternative.

The MDCEV model structure is well suited to model several multiple discrete-continuous choice situations, including individual activity participation and time-use decisions, because the model accommodates the various features of consumer choice making such as satiation, resource allocation, imperfect substitutable alternatives (various discretionary activity types, in the case of activity time-use decisions), essential alternatives, and the simultaneity of decisions. Hence, in the activity time-use model component of the joint residential location and activity time-use model system developed in this chapter, we used the MDCEV model structure.

5.3 ECONOMETRIC MODELING FRAMEWORK

5.3.1 Model Structure

Let q (q = 1, 2, …, Q) be an index for the decision-maker, k (k = 1, 2, …, K) be the index for activity purpose, and i (i = 1, 2, …, I) be the index for the spatial unit of residence. Let T_q be the total amount of time available to individual q for participation in maintenance and discretionary activity purposes, and let $\boldsymbol{t}_q = \{t_{q1}, t_{q2}, ..., t_{qk}\}$ be the vector of time investments in maintenance activity (t_{q1}) and discretionary activities $(t_{q2}, t_{q3}, ..., t_{qk})$. All individuals in the sample participate for some non-zero amount of time in maintenance activity, and hence this alternative constitutes the “outside good” (or the “essential good”) in the MDCEV component of the model system (see Bhat, 2008).[29]

Using the above notational preliminaries, we next discuss the structure of the residential location choice model component (Section 5.3.1.1), the time use model component (Section 5.3.1.2), and then highlight the joint nature of the two components (Section 5.3.1.3).

[29] The term “outside good” (or the “essential good”), as described in Section 5.2.1.3, refers to a good that is “outside” the purview of the choice of whether to be consumed or not. That is, the “outside good” is a good that is always consumed by all consumers.

5.3.1.1 The Residential Location Component

The residential location component takes the familiar discrete choice formulation, as presented below:

$$u^*_{qi} = \varphi'_q z_{qi} + \xi_{qi}\text{, spatial unit } i \text{ chosen if } u^*_{qi} > \max_{\substack{d=1,2,\ldots,I \\ d\neq i}} u^*_{ql}. \tag{5.1}$$

In the equation above, u^*_{qi} is the indirect (latent) utility that the q^{th} individual (as part of her/his household) obtains from locating in spatial unit *i*, z_{qi} is a vector of ATE attributes corresponding to individual *q* and spatial unit *i* (such as land-use mix and measures of activity accessibility), and φ_q is a coefficient vector capturing individual *q*'s sensitivity to attributes in z_{qi}. We parameterize each element *l* of φ_q as $\varphi_{ql} = (\varphi_l + \lambda'_l w_{ql} + v_{ql} + \sum_k \omega_{qkl})$, where w_{ql} is a vector of observed individual-specific factors (such as income and/or household size of individual *q*'s household) affecting sensitivity to the l^{th} attribute in vector z_{qi}, and v_{ql} and $(\sum_k \omega_{qkl})$ are individual-specific unobserved factors impacting individual *q*'s and her/his household's sensitivity to the l^{th} attribute in vector z_{qi}. v_{ql} includes only those individual-specific unobserved factors that influence sensitivity to residential choice, while each ω_{qkl} (k = 2, 3, …, K) includes only those individual-specific unobserved factors that influence both residential choice and time use in discretionary activity purpose *k*. For instance, consider an individual's sensitivity to bicycling facilities around her/his household. The individual may have a higher sensitivity (than her or his observationally equivalent peer group) to bicycling facility density because of general auto disinclination. This auto disinclination may not, however, impact time use in activities. This would be captured in v_{ql}. Now, another unobserved individual factor may be fitness consciousness. This is likely to impact the sensitivity to bicycling facility density in residential choice (because better bicycling facilities are more conducive to bicycling activity) and also influence time invested in physically active leisure. This would be included in ω_{qkl} (more on this later). Finally, in

Equation (5.1), ξ_{qi} is an idiosyncratic error term assumed to be identically and independently extreme-value distributed across spatial alternatives and individuals.

5.3.1.2 The Activity Time-Use Model Component

Designate the first alternative as maintenance activity, which is also the outside good that is always consumed. The rest of the (K-1) alternatives correspond to discretionary activities. Consider the following additive utility function form[30]:

$$U_{qi}(\boldsymbol{t}_q) = \frac{1}{\alpha_1}\psi_{q1i}\, t_{q1}^{\alpha_1} + \sum_{k=2}^{K}\gamma_k\, \psi_{qki}\, \ln\left(\frac{t_{qk}}{\gamma_k}+1\right). \tag{5.2}$$

In the above utility function, $\boldsymbol{t}_q$ is the vector of time investments (t_{q1}, t_{q2}, …, t_{qk}) of individual q, or equivalently, the time spent in each activity purpose.[31] $U_{qi}(\boldsymbol{t}_q)$ refers to the utility accrued to the individual due to time investment $\boldsymbol{t}_q$ if s/he resides in spatial unit i. The term ψ_{qki} corresponds to the marginal random utility of one unit of time investment in alternative k at the point of zero time investment for the alternative for the individual residing in spatial unit i (as can be observed by computing $\partial U_{qi}(\boldsymbol{t}_q)/\partial t_{qk}\,|_{t_{qk}=0}$). Thus ψ_{qki} controls the discrete choice participation decision in alternative k for individual q residing in spatial unit i. We will refer to this term as the baseline preference for alternative k. The term γ_k is a translation parameter that serves to allow corner solutions (zero consumption) for the "inside" alternatives k = 2, 3, …, K

[30] The reader will note that the utility form provided below is a special case of a general form provided by Bhat (2008). Some other utility function forms were also considered, but the one provided below the best data fit. For conciseness, we do not discuss these alternative forms. The reader is referred to Bhat (2008) for a detailed discussion of alternative utility forms. The reader will also note the implicit assumption in the formulation below that there is utility gained from investing time in discretionary activities. This is a reasonable assumption since individuals have the choice not to participate in such activities. Also the reader will note that the inclusion of the maintenance activity type (the first alternative) allows the analyst to endogenously estimate the total amount of time invested in discretionary pursuits.

[31] The individual has the vector $\boldsymbol{t}_q$ as the decision vector. The second through K^{th} elements of $\boldsymbol{t}_q$ can either be zero or some positive value (the first element of $\boldsymbol{t}_q$ should be positive). Whether or not a specific t_{qk} value (k = 2, 3, …, K) is zero constitutes the discrete choice component, while the magnitude of each non-zero t_{qk} value constitutes the continuous choice component. In the rest of this chapter, we will use the terms "time investments" and "time use" interchangeably to refer to these discrete-continuous t_{qk} values.

(γ_k >0). However, it also serves as a satiation parameter for these inside alternatives, with values of γ_k closer to zero implying higher satiation (or lower time investment) for a given level of baseline preference (see Bhat, 2008). There is no γ_1 term for the first alternative in Equation 2 because it is always consumed. However, satiation effects in the consumption of this first alternative are captured through the exponential α_1 parameter, which is bounded from above by a value of 1 (so that marginal utility decreases with increasing time investment in maintenance activity). The constraints that $\alpha_1 \leq 1$ and γ_k >0 (k = 2, 3, …, K) are maintained through appropriate parameterizations (see Bhat, 2008 for details)[32].

To complete the model specification, we express the baseline parameters as functions of observed and unobserved attributes corresponding to the individual q and spatial unit i as follows: $\psi_{qki} = \exp(\beta'_{qk} z_{qi} + \theta'_k x_q + \varepsilon_{qki})$, where z_{qi} is an ATE vector of observed attributes with individual-specific parameter vector β_{qk}, and x_q is a vector of pure individual sociodemographics with coefficient vector θ_k [33]. We further parameterize β_{qk} as follows: $\beta_{qkl} = \beta_{kl} + \Delta_{kl} s_{ql} + \eta_{qkl}$, where s_{ql} is a vector of observed individual-specific factors influencing the sensitivity of individual q to the l^{th} ATE attribute in vector z_{qi} through the Δ_{kl} coefficient, and η_{qkl} is a term capturing the impact of individual-specific and purpose-specific unobserved terms on the sensitivity to the l^{th} ATE attribute in z_{qi}. η_{qkl} is next partitioned into two components: $\sum_l (\pm\omega_{qkl}) z_{qil} + \zeta_{qki}$. The $(\pm\omega_{qkl}) z_{qil}$ terms are the common error components in the residential location and time

[32] The α and γ parameters are subscripted only by activity purpose k (unlike the ψ parameters) because specification tests in our empirical analysis did not show statistically significant variation in these parameters based on individual and spatial unit-related characteristics.
[33] Note that we are introducing the full vector z_{qi} of ATE attributes in both the residential choice and time investment model components. In general, some ATE attributes will not impact residential choice (the corresponding element of φ_q is zero for all q) and some will not influence time investments in activity type k (the corresponding element of β_{qk} is zero for all q). Additionally, it is possible that ATE attributes have a mean effect of zero across individuals for residential choice and/or one or more activity purpose time investments, but have a significant distribution around the zero mean.

investment model components, while ζ_{qki} is an idiosyncratic term assumed to be identically and independently standard type I extreme-value distributed across individuals, activity purposes, and spatial units.[34]

From the analyst's perspective, the individual maximizes the random utility given in the Equation (5.1) subject to the binding linear budget constraint that $\sum_{k=1}^{K} t_k = T$ (where T is the total budget), and the non-negativity constraints on t_k ($k = 1,2,...,K$). Thus, the activity participation and time-use model component of the joint model system is a constrained optimization problem, where a given amount of time is optimally allocated among different types of activities to maximize the utility accrued from the time allocation. The analyst can solve for the optimal consumptions by forming the Lagrangian and applying the Kuhn-Tucker (KT) conditions. The resulting stochastic first order conditions, and a type-I extreme value distribution on the ζ_{qki} terms results in a closed form conditional probability expression (conditional upon the other error terms in the utility specification) for the time-use component of the joint model system. The reader is referred to Bhat (2008) for the derivation of the probability expressions.

5.3.1.3 The Joint Model System

The specifications in the previous two sections may be collected and brought together in the following equation system:

$$u_{qi}^{*} = \sum_{l} \left(\varphi_l + \lambda_l' w_{ql} + v_{ql}\right) z_{qil} + \sum_{k} \left(\sum_{l} \omega_{qkl}\right) z_{qil} + \xi_{qi}$$

[34] In our empirical analysis, we further partitioned ζ_{qki} into error components that generate covariance across discretionary activity purposes (for example, individuals who are generically inclined toward social activities may have a higher baseline preference than their observationally equivalent peers for both in-home and out-of-home social activities). However, in our final specification, we did not find any such statistically significant covariance terms. Thus, for simplicity in presentation, we are imposing the restriction of zero covariances across discretionary activity purposes in the model structure presentation.

$$U_{qi}(\boldsymbol{t}_q) = \frac{1}{\alpha_1}\exp(\zeta_{q1i})t_{q1}^{\alpha_1} +$$
$$\sum_{k=2}^{k}\left\{\gamma_k \exp\left[\left\{\sum_{l}(\beta_{kl}+\Delta_{kl}s_{ql}+\eta_{qkl})z_{qil}\right\}+\theta_k' x_q+\left\{\sum_{l}(\pm\omega_{qkl})z_{qil}\right\}+\zeta_{qki}\right]\ln\left[\frac{t_{qk}}{\gamma_k}+1\right]\right\}$$
(5.3)

In the above equation system, the vectors z_{qi} and x_q do not appear for the first alternative in $U_{qi}(\boldsymbol{t}_q)$ because only utility differences matter. From the analyst's perspective, individuals are choosing a residential location *i* and their time use profile by jointly maximizing u_{qi}^* and $U_{qi}(\boldsymbol{t}_q)$ subject to the budget constraint that $\sum_k t_{qk} = T_q$, where T_q, as mentioned earlier, is the total time available to individual *q* for participation in maintenance and discretionary activities.

The joint nature of the model system arises because of the presence of the common unobserved $\omega_{qkl}z_{qil}$ terms in the residential choice and time investment model components. More generally, the model system allows self-selection of individuals (based on their time-use preferences) into neighborhoods due to both observed and unobserved factors. In the context of observed factors, consider the situation where individuals who own bicycles are attracted toward residential neighborhoods with very good bicycling facilities. This can be reflected by including bicycle ownership as a variable in the w_{ql} vector that corresponds to the "bicycling facility density" variable in the z_{qi} vector of the residential choice equation. Individuals who own bicycles may also be more likely to spend time in physically active recreational pursuits, which can be accommodated by including bicycle ownership as a variable in the x_q vector in the time investment model component. If bicycle ownership were not included in the x_q vector for the physically active recreation purpose in the time investment equation, and if "bicycling facility density" were included as a variable in the z_{qi} vector of the same equation, the result could be an inflated estimate of the positive influence of bicycling facility density on time invested in physically active recreation. Similarly, consider an unobserved individual factor, such as fitness consciousness, that makes some individuals locate in areas with good bicycling facility and also participate more than their observationally equivalent peers in physically active recreation. Such factors are captured in the common $\omega_{qkl} z_{qil}$ terms in the two equations. The '±' sign in front of the $\omega_{qkl} z_{qil}$ terms in the time-investment model component indicates that the correlation in the unobserved factor *l* may be positive or negative. If the sign is positive (negative), it implies that individuals who intrinsically prefer the ATE characteristics represented by z_{qil} also have a higher (lower) generic preference for activity purpose *k*.

5.3.2 Model Estimation

The parameters to be estimated in the equation system (5.3) include the φ_l, λ_l, Δ_{kl}, and θ_k vectors, the γ_k ($k = 2, 3, \ldots, K$), α_1, and β_{kl} scalars ($k = 2, 3, \ldots, K$; $l = 1, 2, \ldots, L$), and the variances of the stochastic components v_{ql}, η_{qkl}, and ω_{qkl} (all assumed to be normally distributed with variances σ_{vl}^2, $\sigma_{\eta kl}^2$, $\sigma_{\omega kl}^2$, respectively) that represent random heterogeneity.

Let Ω represent a vector that includes all the parameters to be estimated, and let $\Omega_{-\sigma}$ represent a vector of all parameters except the variance terms. Also, let g_q be a vector that stacks the v_{ql}, η_{qkl}, and ω_{qkl} terms, and let Σ be a corresponding vector of standard errors. Define $a_{qi} = 1$ if individual *q* resides in spatial unit *i* and 0 otherwise. Then the likelihood function for a given value of $\Omega_{-\sigma}$ and g_q may be written for an individual *q* as:

$$L_q(\Omega_{-\sigma}) \mid g_q = \prod_{i=1}^{I} \left\{ \left[\frac{\exp\left[\sum_l \left(\varphi_l + \lambda_l' w_{ql} + v_{ql} + \sum_k \omega_{qkl} \right) z_{qil} \right]}{\sum_j \exp\left[\sum_l \left(\varphi_l + \lambda_l' w_{ql} + v_{ql} + \sum_k \omega_{qkl} \right) z_{qil} \right]} \right] \times \left[|J_q| \frac{\prod_{h=1}^{M_q} e^{V_{qhi}}}{\left(\sum_{k=1}^{K} e^{V_{qki}} \right)^{M_q}} (M_q - 1)! \right] \right\}^{a_{qi}}, \tag{5.4}$$

where M_q is the number of activity purposes in which the individual q participates, and $|J_q|$ and V_{qki} are defined as follows:

$$|J_q| = \left(\prod_{h=1}^{M_q} c_{qh} \right) \left(\sum_{h=1}^{M_q} \frac{1}{c_{qh}} \right), \; c_{q1} = \frac{1-\alpha_1}{t_{q1}^*}, \; c_{qh} = \frac{1}{t_{q1}^* + \gamma_h}, \; (h = 2, \ldots, M_q)$$

$$V_{q1i} = (\alpha_1 - 1) \ln t_{q1}^* \;\; \forall \; i = 1, 2, \ldots, I,$$

$$V_{qki} = \left(\sum_l (\beta_{kl} + \Delta_{kl} s_{ql} + \eta_{qkl}) z_{qil} + \theta_k' x_q + \sum_l (\pm \omega_{qkl}) z_{qil} \right) - \ln\left(\frac{t_{qk}^*}{\gamma_k} + 1 \right) \; (k = 2, 3, \ldots, K). \tag{5.5}$$

In the above expression, the vector $\boldsymbol{t_q^*} = \{t_{q1}^*, t_{q2}^*, \ldots, t_{qk}^*\}$ represents the optimal (*i.e.,* observed) time investments of individual q in the various activity purpose.

Finally, the unconditional likelihood function can be computed for individual q as:

$$L_q(\Omega) = \int_{g_q} \left(L_q(\Omega_{-\sigma}) \mid g_q \right) \, d\boldsymbol{F}(g_q \mid \Sigma), \tag{5.6}$$

where $\boldsymbol{F}$ is the multidimensional cumulative normal distribution. The log-likelihood function can be written as: $L(\Omega) = \sum_q \ln L_q(\Omega)$. Simulation techniques are applied to approximate the multidimensional integral in Equation (5.4), and the resulting simulated log-likelihood function is maximized. Specifically, the scrambled Halton sequence (see Bhat, 2003) is used to draw realizations from the population normal distribution. In the current chapter, we tested the sensitivity of parameter estimates with different numbers of scrambled Halton draws per observation, and found the results to be stable with as few as 100 draws. In the analysis, we used 125 draws per observation in the estimation.

5.4 THE DATA

5.4.1 Data Sources

The primary source of data used for this analysis is the 2000 San Francisco Bay Area Travel Survey (BATS) designed and administered by MORPACE International Inc. for the Bay Area Metropolitan Transportation Commission (MTC). The survey collected information on all activity episodes (in-home and out-of-home) undertaken by individuals from over 15,000 households in the Bay Area for a two-day period (see MORPACE International Inc., 2002, for details on survey, sampling and administration procedures). Information characterizing the context (activity type, start and end times of the activity, and location of participation) of each activity episode was collected. Furthermore, data on individual and household socio-demographics was also obtained.

In addition to the 2000 BATS data, several other secondary data sources were used to derive spatial variables characterizing the activity-travel environment in the region. These include: (1) Land-use/demographic coverage data, obtained from the MTC, (2) The zone-to-zone motorized travel level of service files, obtained from MTC, (3) The Census 2000 population and Housing data summary files (SF1), (4) GIS layers of businesses (automotive businesses, shopping and grocery stores, medical facilities and personal services, food stores, sports and fitness centers, parks and gardens, restaurants, recreational businesses, and schools), obtained from the InfoUSA business directory, (5) GIS layers of bicycling facilities, also obtained from MTC, and (6) GIS layers of highway (interstate, national, state and county highways) network and local roadways (local, neighborhood, and rural roads) network, extracted from the Census 2000 Tiger files.

5.4.2 Sample Formation

5.4.2.1 Sample Extraction

Several steps were undertaken to generate the data for the current analysis. First, only adults (individuals 16 years or older) were selected to focus the analysis on the demographic segment that clearly exercises a choice in activity participation and time allocation. Second, only weekday data was selected from the sample, since the focus of our analysis is on the time-use patterns on weekdays. Third, of the two-day activity diary data available for each individual, only one randomly chosen day was picked. In addition, we restricted the analysis to the time-use patterns of a single randomly chosen individual from each household. These decisions were made to keep the sample size manageable in the estimations, and also to avoid the problem of repeated data measurement from the same individual and/or household. Fourth, we selected only the non-work episodes for each individual. Finally, we confined the analysis to Alameda County in the San Francisco Bay area.

5.4.2.2 Activity Type Classification

The survey instrument used a 17-category activity purpose classification scheme, and collected information on the location of each activity episode. The activity purpose and location information was used to create a 13-category activity purpose typology in the current study to classify non-work episodes. The activity purposes are: (1) Maintenance (household chores, personal care, meal preparation, grocery shopping, and medical appointments), (2) In-home (IH) internet browsing, (3) Out-of-home (OH) volunteering (including religious and civic activity participation), (4) OH non-maintenance shopping (*i.e.,* shopping for non-grocery purposes), (5) IH socializing, (6) OH socializing, (7) IH relaxing (resting, reading, listening to music, *etc.*), (8) IH recreation (hobbies, television viewing, *etc.*), (9) OH meals, (10) OH physically active pure recreation (episodes undertaken using non-motorized modes, and without any specific destination, such as walking, jogging, or bicycling around the neighborhood, are classified in this category), (11) OH physically inactive pure recreation (episodes undertaken using motorized modes, and without any specific destination, such as joy-riding around the block, belong to this category), (12) OH physically active recreation (exercising at the gym, playing tennis, *etc.*), and (13) OH physically passive recreation (going to the movies, opera show, *etc.*). The dependent variables in the time investment component of the model system correspond to the time allocated during the 24-hour survey weekday to each of the 13 activity purposes just identified. These variables are obtained by appropriate time aggregation across all episodes within each purpose category for each individual. The total time across all these 13 categories is considered exogenous, and the model focuses on residential choice and daily time-use in each of the 13 activity purpose categories.

5.4.2.3 Activity Travel Environment (ATE) Measures

The data from the secondary sources identified in Section 5.4.1 were used to compute a host of built environment measures for each traffic analysis zone (TAZ), including:

7. Zonal size and density measures, such as total population, population density, household density, density of employment by each of several employment categories, and dummy variables for central business district (CBD), urban, suburban, and rural areas (computed based on employment density). These attributes were obtained from the zonal land-use data file.
8. Zonal land-use structure variables, such as housing type measures (fractions of single family, and multiple family dwelling units), fractions of zonal area in residential and commercial land-uses, and land-use mix (see Bhat and Gossen, 2004 for a description of the land-use mix variable; the variable takes a value of zero for zones with only one type of land-use and a value of 1 for zones with equal distributions in area among residential, commercial, and other land-uses). The zonal land-use structure variables were constructed from the zonal land-use data file.
9. Regional accessibility measures, such as shopping accessibility, recreational accessibility, and employment accessibility. These are Hansen-type (see Fotheringham, 1983) accessibility measures computed from the zonal land-use and level-of-service data.

10. Zonal demographics, such as average household size and median household income. These demographic measures were derived from the Census 2000 population and housing data summary file (SF1).
11. Zonal ethnic composition measures, constructed as fractions of Caucasian, African-American, Hispanic, Asian, and other ethnicity populations, also derived from the Census 2000 population and housing data summary file.
12. Zonal activity opportunity variables, such as activity center intensity (*i.e.,* the number of business establishments per square mile) and density (*i.e.,* the number of business establishments per square mile) for each of the following activity types (extracted from the InfoUSA business establishments data): (a) maintenance (grocery stores, gas stations, food stores, car wash, automotive businesses, banks, medical facilities, *etc.*), (b) physically active recreation (fitness centers, sports centers, dance and yoga studios, parks, gardens, *etc.*), (c) Physically inactive recreation (theatres, amusement centers, arcades, *etc.*), and (d) eat-out (restaurants and eateries).
13. Zonal transportation network measures, such as highway density (miles of highway facilities per square mile), bikeway density (miles of bikeway facilities per square mile), and local roadway density (miles of roadway density per square mile). These variables were extracted from the GIS layers of bikeways and roadways.

All of these ATE attributes were merged with the activity time-use and individual/household demographic data to form a comprehensive database suitable for modeling residential choice and time-investment decisions.

5.4.3 Sample Description

The final estimation sample consists of 2793 individuals residing in Alameda County in the San Francisco Bay area. Each individual has the choice of residing in any one of 236 zones in Alameda County, which is modeled using a standard discrete choice formulation. In this residential choice model, we do not undertake a sampling scheme to reduce the number of alternatives because our overall residential choice model does not correspond to a simple multinomial logit model (rather there is mixing in the model).

The time-use of individuals in each of 13 activity purposes constitutes the dependent variables of the MDCEV component of the model system. Each (and all) of the 2793 individuals participated in maintenance activity. The frequency distribution of individuals based on the number of discretionary activity types they participated in is provided in Table 5.1.

Table 5.1 Distribution of Individuals by Number of Discretionary Activity Purposes Participated in

Number of discretionary activity types participated in	Number of individuals participated	(%) of individuals participated
0	910	(32.6%)
1	919	(32.9%)
2	610	(21.8%)
3	241	(8.6%)
4	88	(3.2%)
5	20	(0.7%)
6	5	(0.2%)
Total	2,793	100%

As can be observed, 32.6% of individuals did not participate in any discretionary activity, while the remaining 67.4% of individuals participated in at least one type of discretionary activity. Of the individuals who participated in at least one type of discretionary activity, 51.2% of individuals participated in multiple types of discretionary activities. Overall, the results clearly illustrate the high prevalence of participating in multiple non-work activity purposes on a single weekday, providing strong support for the use of the MDCEV model for time-use analysis.

5.5 EMPIRICAL ANALYSIS

Several types of variables were considered in the joint modeling system. These included: (1) household socio-demographics (household size, household composition and family structure, vehicle and bicycle ownership, ethnicity, income, whether or not household owned or rented its residence, dwelling type, availability of internet, *etc.*), (2) individual demographics and employment characteristics (age, license holding to drive, employment status, number of hours of work on the day, physical disability status *etc.*), (3) contextual variables such as season of the year, day of the week, rain-fall (and the amount of rain fell) in the day, temperature (minimum and maximum values in the day, and the variation in the day), (4) a host of activity-travel environment variables, and (5) the interactions of the activity-travel environment variables with household and individual socio-demographics.

The final variable specification was obtained based on a systematic process of eliminating variables found to be statistically insignificant, and parsimony in representation. The specification was additionally guided by intuitive considerations, and results from earlier studies. In the next section (Section 5.5.1), we discuss the results of the residential location choice model component. Section 5.5.2 presents the results of the activity time-use model component. Section 5.5.3 discusses self-selection effects, and Section 5.5.4 focuses on likelihood-based measures of data fit.

5.5.1 Residential Location Choice Model Component Results

Table 5.2 (in the next page) presents the estimation results for the residential location choice component of the joint model. The first set of variables in Table 5.2 corresponds to zonal size and density measures. The coefficient on the logarithm of number of households in the zone has the expected positive sign, indicating that individuals are more likely to locate in zones with a large number of housing opportunities. The parameter on this variable is between 0 and 1, as should theoretically be the case on the size measure (see Bhat *et al.,* 1998 and Daly, 1982). The household density effects indicate that individuals whose households do not have seniors (of age > 65 years) and do not have young children ($\leq$ 15 years of age) are likely to locate in zones with high household density, while those with seniors and young children prefer low density neighborhoods.

Table 5.2. Estimation Results of the Residential Location Choice Component of the Joint Residential Location and Time-use Model

Variables	Parameter	t-stat
Zonal size and density measures (including demographic interactions)		
Logarithm of number of households in zone	0.836	16.97
Household density (#households per acre x 10^{-1})	0.238	3.61
Interacted with presence of seniors in household	-0.736	-6.41
Interacted with presence of children (of age < 5 years) in household	-0.392	-2.30
Interacted with presence of children (of age 5 to 15 years) in household	-0.216	-2.23
Employment density (#jobs per acre x 10^{-1})	-0.006	-1.02
Interacted with household income less than $ 35,000 per annum	0.196	2.38
Interacted with household belonging to the African American race	0.114	1.81
Interacted with presence of children (of age 15 years or younger) in household	-0.069	-1.81
Interacted with household belonging to the Caucasian race	-0.067	-1.48
Zonal land-use structure variables (including demographic interactions)		
Fraction of commercial land area	-0.177	-1.04
Land-use mix	-0.216	-1.98
Zonal demographics (including demographic interactions)		
Absolute difference between zonal median income and HH income ($ x 10^{-3})	-0.022	-15.29
Absolute difference between zonal average household size and household size	-0.457	-7.87
Fraction of senior (age > 65) population interacted with presence of senior adult in Household	2.760	4.67
Zonal race composition measures (including demographic interactions)		
Fraction of African-American population * African-American dummy variable	3.931	9.35
Fraction of Asian population * Asian dummy variable	3.241	8.06
Fraction of Caucasian population * with Caucasian dummy variable	1.912	13.48
Fraction of Hispanic population * Hispanic dummy variable	2.130	3.09
Zonal activity opportunity variables (including demographic interactions)		
Number of schools in the zone	0.020	1.90
Number of physically active recreation centers	0.080	1.56
Number of natural recreational centers such as parks, gardens, *etc.* interacted with Number of bicycles in the household	0.012	1.03
Zonal transportation network measures (including demographic interactions)		
Highway density (miles of highway per square mile) interacted with presence of seniors (age > 65) in household	-0.040	-1.01
Street block density (number of blocks per square mile x 10^{-1}) interacted with household income greater than $ 90,000 per annum	-0.044	-4.12
Bicycling facility density (miles of bike lanes per square mile)	0.028	3.03
Interacted with number of bicycles in the household	0.016	3.93

The latter result may be a reflection of the "urban-to-suburban flight" trend of households that have young children, as has been documented in several sociological studies (Birch, 2005, Lee and Guest, 1983, Marans, 1979, and Waddell and Nourzad, 2002). This flight trend has been attributed to the safety and neighborhood quality related preferences of households having children. The employment density coefficients indicate that low income individuals (annual household income < $35,000) and African-American households tend to locate themselves in areas with high employment density, while those who have high incomes, who have children (15 years of age or younger) in their household, and who are Caucasians prefer areas of low employment density.

The zonal land-use structure variable effects indicate that individuals shy away from zones with a high fraction of commercial land area and with good land use-mix diversity. The effect of the land-use mix variable is interesting, and suggests that individuals do, in general, prefer homogenous, exclusive enclaves, for their home location.

Our results did not indicate any statistically significant effects of the regional accessibility measures. The zonal demographic variables show strong effects on residential choice. In particular, there is a clear indication of residential clustering of individuals based on household income levels, household size, and age. Such clustering trends have long been documented in the residential analysis literature (see Waddell, 1993b, and Waddell, 2006).

The zonal race variables, when interacted with individual race, show clear evidence of racial clustering, with the effect being strongest for African-American individuals and weakest for Caucasian individuals. The clustering effect for Asian and Hispanic individuals is somewhere between those for African-Americans and Caucasians, with Asians having a higher tendency to cluster than Hispanics. The racial clustering may be a result of social networking, race-specific preferences for unobserved zonal characteristics, discrimination in the housing market place, or some combination of these. Disentangling the possible reasons for racial clustering is an important avenue for further research.

The next set of variables in Table 5.2 represents the opportunities offered by a zone for different types of activities. The positive coefficient on the number of schools in a zone indicates that households are more likely to locate in zones with better schooling opportunities. Interestingly, we did not find any statistically significant variation in the effect of number of schools based on the presence (or number) of children in the household. This is, in part, because a high fraction of individuals in our sample have children in their households, leading to inadequate variation in the children-related variables. The next variable, "number of physically active recreation centers such as fitness centers/gymnasiums, sports centers, dance and yoga centers", suggests a positive impact of the zonal level opportunities for physically active recreational activity pursuits on a household's preference to reside in that zone. The final variable under zonal activity opportunities indicates the preference of individuals with bicycles in their households to locate in zones with a higher number of natural recreation opportunities, perhaps due to the generally outdoor activity inclination of individuals with bicycles. Admittedly, none of these zonal activity opportunity variables have a highly statistically significant impact on residential choice, but they are included so that their effects can be explored more carefully in future residential choice studies that use a neighborhood-level scale for measuring these opportunity attributes (rather than the aggregate zone-level scale in the current study).

The effects of the zonal transportation network measures show that individuals with seniors (age > 65 years) in their household are likely to stay away from zones with high highway density. This is possibly a reflection of the preference for relatively quiet, retirement-like, communities amongst households with seniors (though this effect is not very statistically significant). The influence of street block density lends reinforcement to the stereotype of high income households choosing to locate in suburban-like, low density, sprawling, communities away from the "inner city" neighborhoods. Finally, in the set of zonal network measures, the results show that, without exception, households prefer zones with high bicycling facility density, though this effect is particularly strong among individuals who own several bicycles.

5.5.2 Time-use Analysis Component Results

The final specification results of the activity time-use component of the joint model (*i.e.,* the MDCEV modeling component) are presented in Table 5.3 (see next page). Each row corresponds to a parameter or a variable, and each column corresponds to a specific activity purpose category. The maintenance activity purpose serves as the base category for all variables. In addition, a blank (*i.e.,* no parameter estimate) for a variable for an alternative implies that the alternative also constitutes the base category for that variable.

5.5.2.1 Baseline Preference Constants

The baseline preference constants (see first row of Table 5.3) do not have any substantive interpretations. They capture generic tendencies to participate in each discretionary activity type category as well as accommodate the range of continuous independent variables in the model. However, all the baseline preference constants are negative, which is a reflection of the fact that all individuals participate in maintenance activity (the "outside" good), while this is not the case for the remaining discretionary activity purposes.

5.5.2.2 Satiation/Translation Parameters

The satiation parameter estimate of α_1 for the maintenance activity (*i.e.,* the outside good) alternative is 0.253, with a t-statistic value of 24.82 for the null hypothesis that α_1 = 1. This indicates the presence of strong satiation effects in the time investment on maintenance activities. That is, the marginal utility of time investment in maintenance activity drops rapidly with increasing maintenance time.

The translation parameters (γ_k) for the discretionary activity purposes are presented in the second row of Table 5.3. As discussed in Section 2.1.2, the γ_k parameters allow corner solutions (*i.e.,* zero time investment in the discretionary activity purposes) as well as serve as satiation parameters. The magnitude of the

Table 5.3. Estimation Results of the Time-use Component of the Joint Residential Location and Time-Use Model

	OH Volunteering	IH Internet use	OH Shopping	IH Socializing	OH Socializing	IH Relaxing	IH Recreation	OH Meals	OH Physically Active Pure Recreation	OH Physically Inactive Pure Recreation	OH Physically Active Recreation	OH Physically Inactive Recreation
Baseline Constants	-6.751 (-33.93)	-9.321 (-16.73)	-5.985 (-29.52)	-7.863 (-29.26)	-6.668 (-39.49)	-5.669 (-28.80)	-5.667 (-29.09)	-7.106 (-28.23)	-7.093 (-24.61)	-7.275 (-38.23)	-7.243 (-26.07)	-7.540 (-28.63)
Translation Parameters	147.85 (4.10)	138.13 (3.96)	27.87 (9.43)	145.02 (3.89)	100.20 (6.22)	251.99 (8.85)	262.38 (7.07)	42.40 (10.80)	42.02 (3.44)	36.90 (4.51)	111.06 (4.08)	127.26 (4.84)
Household Sociodemographics												
Single person family				0.475 (2.04)		-0.282 (-2.42)						
Caucasian								0.254 (2.44)				
Internet at home		2.391 (4.51)										
35k<HH income<90k						-0.362 (-2.67)	-0.387 (-2.76)	0.338 (2.42)	0.213 (1.30)	0.213 (1.30)	0.213 (1.30)	0.213 (1.30)
HH income >90K						-0.319 (-1.98)	-0.498 (-2.69)	0.371 (2.31)	0.217 (1.16)	0.217 (1.16)	0.217 (1.16)	0.217 (1.16)
Number of vehicles							-0.122 (-1.85)		-0.138 (-1.79)		-0.138 (-1.79)	
Number of bicycles						-0.042 (-1.38)			0.080 (1.32)			
Individual Sociodemographics and Employment Characteristics												
Male	-0.310 (-1.72)	0.387 (1.99)				0.333 (3.50)	0.362 (3.42)	0.307 (3.44)			0.161 (1.10)	
Age < 30 yrs		0.684 (2.72)		0.680 (2.25)	0.921 (5.12)	0.276 (2.11)	0.544 (3.93)	0.520 (4.15)		0.381 (1.32)		0.591 (2.68)
Age > 65 yrs	0.656 (3.31)											
Driver's license			0.522 (3.20)					0.522 (3.20)			0.522 (3.20)	0.522 (3.20)
Physically challenged	-0.610 (-1.19)		-0.610 (-1.19)		-0.610 (-1.19)			-0.610 (-1.19)	-0.610 (-1.19)	-0.610 (-1.19)	-0.610 (-1.19)	-0.610 (-1.19)
Employed		-0.390 (-1.60)		-0.414 (-1.79)			-0.177 (-1.58)	0.425 (4.07)				
Work duration (min)	-0.159 (-3.83)	-0.091 (-1.84)	-0.135 (-6.51)		-0.113 (-3.85)	0.060 (3.02)			-0.118 (-2.46)			-0.093 (-2.88)

Table 5.3 (continued). Estimation Results of the Time-use Component of the Joint Residential Location and Time-use Model

	OH Volunteering	IH Internet use	OH Shopping	IH Socializing	OH Socializing	IH Relaxing	IH Recreation	OH Meals	OH Physically Active Pure Recreation	OH Physically Inactive Pure Recreation	OH Physically Active Recreation	OH Physically Inactive Recreation
Day of the Week and Seasonal Variables												
Friday			0.382 (3.00)	0.769 (3.01)	0.629 (3.78)	0.322 (2.73)	0.313 (2.33)	0.436 (3.89)				0.505 (2.77)
Summer									-0.468 (-2.70)	-0.468 (-2.70)		
Fall						-0.360 (-3.39)			-0.401 (-2.23)	-0.401 (-2.23)	-0.274 (-2.21)	-0.274 (-2.21)
Activity-Travel Environment Attributes												
Employment density (#jobs per Acre x 10^{-1})		0.013 (1.61)		0.013 (1.61)		0.013 (1.61)	0.013 (1.61)					
#eat-out centers per square mile (x 10^{-1})								0.026 (2.65)				
#physically inactive recreational centers per square mile (x 10^{-1})												0.091 (3.21)
Presence of more than 4 sports/fitness centers in the residential zone interacted with presence of bicycles in Household											0.187 (2.84)	
Bicycling facility density (miles of bike lanes per square mile)									0.116 (2.47)			
Bicycling facility density interacted with #vehicles/#adults									-0.050 (-1.00)			

γ_k parameter for any activity purpose k is inversely associated with satiation for that activity purpose. Thus, a value closer to zero for activity purpose k, in general, implies higher satiation for activity purpose k (and, therefore, lower time investment in activity purpose k), while a high value for an activity purpose implies lower satiation for that activity purpose (or, equivalently, higher time investment in that activity purpose). The results show high satiation effects (low durations) for time investments in out-of-home (OH) shopping, OH meals, OH physically activity pure recreation, and OH physically inactive pure recreation. On the other hand, the results indicate the low satiation effects (high durations) for time investments in in-home (IH) relaxation and IH recreation. The satiation effects for the remaining activities are between these two extremes. These results are consistent with intuitive expectations.

5.5.2.3 Household Sociodemographics

Among the household sociodemographic variables, the effect of household structure on the baseline utilities indicates that individuals living alone are more likely to partake in IH socializing, and less likely to participate in IH relaxing, compared to individuals not living alone. This is perhaps a reflection of the higher need to socialize and interact with other individuals when living alone (see Yamamoto and Kitamura, 1999 for similar results). The specific preference for in-home social activities over out-of-home social activities among individuals living alone is interesting, and needs further exploration in future studies.

The race variable effects show that Caucasians have a higher baseline preference for OH meal activity relative to other races. Other than this, there are no statistically significant differences among races in time use patterns. This is in contrast to studies of weekend activity time-use that show that Caucasians are not only more likely than non-Caucasians to participate in OH meal activity, but also in all other kinds of OH recreational activities (see Kapur and Bhat, 2007; and Bhat *et al.,* 2006). This points to greater heterogeneity in time-use among races during the weekends relative to weekdays. The next household sociodemographic attribute is the availability of internet at home, which, as expected, is associated with a higher tendency of individuals to invest time on internet use.

Household income also impacts time investment patterns in activity purposes. Specifically, individuals in low income households have a higher baseline preference for in-home relaxing and in-home recreation relative to individuals in middle and high income households, while those in middle and high income households are more likely to participate in out-of-home meals and out-of-home recreation pursuits. These results are consistent with the higher consumption potential of services and out-of-home recreation facilities of individuals in higher income earning households.

Finally, within the set of household sociodemographic variables, the coefficient on the number of motorized vehicles suggests a positive association between high motorized vehicle ownership and an inactive life style (*i.e.,* lower preference for physically active pure recreational travel as well as physically active out-of-home recreational activities). On the other hand, high bicycle ownership is associated with a lower preference for in-home relaxing and a higher preference for physically active pure recreational travel. The latter result is quite reasonable. Households who own more bicycles are likely to be more outdoor-oriented by nature, as discussed in the residential location choice model results in Section 4.2. In that section, we also pointed to the increased likelihood of individuals with high bicycle ownership to locate themselves in zones with good bicycling facilities. Thus, by including the bicycle ownership variable in the time-use model component, we are capturing residential self-selection effects due to bicycle ownership. Specifically, if bicycle ownership were not included in the time-use model component, it would lead to an over-inflated effect of the potential to encourage physically active pure recreation participation by designing neighborhoods with good bicycling facilities (as we found to be the case in our empirical analysis when we removed the bicycle ownership variable from the time-use model specification).[35]

---

[35] See Pinjari *et al.* (2007) for a similar finding on residential self selection effects due to bicycle ownership in a commute mode choice model. Other studies that accommodate such self selection effects due to observed factors include, for example, Bagley and Mokhtarian (2002), Bhat and Guo (2007), Cao *et al.,* (2006a), Guo *et al,* (2007), Handy *et al.* (2005), Kitamura *et al.* (1997), Khattak and Rodriguez (2005), and Schwanen and Mokhtarian (2003, 2005b).

Overall, the results pertaining to vehicle ownership emphasize the transportation-public health connection. That is, our analysis suggests that policies and educational campaigns aimed at reducing motorized vehicle ownership and increasing bicycle ownership not only can lead to traffic congestion alleviation, but can also play an important role in improving public health.

An interesting point is in order here regarding the effects of household sociodemographics. In earlier studies of weekend time-use, the presence and number of children has been found to play an important role in adult time-use decisions. For instance, Bhat *et al.* (2006) indicate that individuals in households with children have a high baseline preference for out-of-home recreation and pure recreational pursuits, and a lower preference for in-home leisure activities. They attribute these effects to a stronger need to have a change from caring for children in-home and the propensity to participate with young children in outdoor pursuits. In contrast, we did not find any statistically significant effect of children on time-use patterns on weekdays. This is perhaps because of weekday work- and school-related activities, because of which there is less of a need to have a change from caring for children in-home and less opportunity for joint participation between adults and children.

5.5.2.4 Individual Sociodemographics and Employment Characteristics

Among the individual sociodemographics, the coefficient on the male dummy variable highlights the role of gender in weekday discretionary time-use. Women are more inclined to participate in volunteering activities, while men participate more in internet use, in-home relaxing, in-home recreation, out-of-home meals, and OH physically active recreation. The higher participation of men in relaxing/meal/recreational pursuits also implies, because of the fixed time constraint across all discretionary and maintenance activities, that men participate less in maintenance activity. That is, women have more responsibility for household maintenance activity, a result consistent with the findings of several earlier studies (see, for example, Chen and Mokhtarian, 2006, Gossen and Purvis, 2005, Goulias, 2002, Levinson, 1999, and Srinivasan, 2004).

Several different functional form specifications were attempted for the age-related variables, including a continuous variable, spline variables that allow piece-wise linear effects of age, and age dummy variables. After extensive testing, the best results were obtained using dummy variables for age less than 30 years, age between 30-65 years, and age greater than 65 years. The age variables are introduced in Table 3 with the age category between 30 to 65 years as the base. The results show that young adults (16-29 years) are more likely than other adults to participate in internet browsing and in all discretionary activities other than OH volunteering, OH shopping and physically active recreational pursuits (see Yamamoto and Kitamura, 1999 and Bhat *et al.,* 2006 for similar results). On the other hand, older adults (>65 years) have a higher baseline preference for out-of-home volunteering activities.

The individual mobility-related variables indicate that, in general, individuals with a driver's license have a higher preference (than those without a driver's license) to participate in out-of-home discretionary activities, while those who are physically challenged are less inclined to participate in out-of-home discretionary activities. These are clearly manifestations of enhanced mobility to access activities (in the case of having a driver's license) and mobility constraints (in the case of being physically challenged).

The effect of employment can be discerned from the coefficients on the employment dummy and work duration variables in Table 5.3. Overall, employed individuals have a lower baseline preference (relative to unemployed individuals) for all in-home activities other than relaxing. They also have a higher preference for IH relaxing, OH meal activities, and a lower preference for other OH activities. Further, among those employed, individuals who work longer have a particularly high preference for IH relaxing and a low preference for OH discretionary activities. These results reflect time constraints that make employed individuals (particularly those working long hours) spend more of their non-work time on maintenance activities (the base category), in-home relaxation, and out-of-home meals.

5.5.2.5 Day of Week and Seasonal Variables

The day of week effects reveal the higher inclination of individuals to participate on Fridays (relative to other weekdays) in such discretionary activities as OH shopping, socializing (in-home as well as out-of-home), IH relaxing, IH recreation, OH meals, and OH physically inactive recreation. The results also show the lower preference for pure recreation activities during the summer season, and for all kinds of recreation and IH relaxing activities in the Fall. These seasonal variations need further exploration in future studies. Our attempts to include weather-related factors (such as rainfall and temperature) were not successful.

5.5.2.6 Activity-Travel Environment Attributes

Among the activity-travel environment attributes, the parameters of the employment density variable indicate that individuals residing in high employment density neighborhoods are more likely to spend their leisure time at home than out-of-home. This may possibly be a reflection of traffic congestion and mobility problems in areas with high employment density that leads to higher in-home discretionary activity participation.

The variables representing the opportunities offered by a zone for various types of activities reveal the higher likelihood of participation in OH meal activity (OH physically inactive recreation) in areas with several eat-out centers (physically inactive recreational centers) per square mile. Also, individuals who own one or more bicycles, and live in areas with a high intensity of availability of sports/fitness centers, have a higher baseline preference for OH physically active recreational pursuits. This reflects the interaction effect of owning a bicycle and the presence of sports/fitness centers on OH physically active participation, perhaps because individuals who own a bicycle are also fitness-conscious.

The positive impact of the bicycling facility density variable on OH physically active pure recreation suggests that better bicycling facilities do lead to higher participation rates in physically active pursuits such as walking, biking, or jogging. The coefficient on the interaction of the bicycling facility density variable with household vehicle availability indicates that vehicle ownership moderates down the positive impact of bicycling facilities on OH physically active pure recreation. This finding further adds to the evidence on the association of higher vehicle ownership with physically inactive lifestyles (see Section 5.5.2.3).

5.5.3 Self-Selection Effect

In our model specifications, we explored unobserved heterogeneity in the effects of several variables in the residential and time-investment model components [corresponding to the v_{ql} and η_{qkl} terms in Equation (5.3)]. None of these effects turned out to be statistically significant. However, we did find a statistically significant standard deviation of the unobserved component associated with the bicycling facility density variable that was common to both the residential choice component and the OH physically active recreation baseline utility in the time-investment model component (corresponding to ω_{qkl} in Equation 5.3). The estimated standard deviation was 0.042 with a t-statistic of 2.31. The sign corresponding to this variable in the $\pm\omega_{qkl}z_{qil}$ term of the time-use component was positive. The implication is that there are unobserved individual factors (such as fitness consciousness that cannot be completely captured by such observed variables as bicycle ownership) that make individuals locate in areas with good bicycling facilities and also lead to a high preference for physically active recreation. That is, people who are predisposed to physically active lifestyles tend to self-select themselves into zones with very good bicycling facility density for their residence. If this residential sorting is not accounted for, and the time investment model includes the bicycling facility density variable, the result could be a spurious finding that bicycling facility density *causes* higher participation levels in OH physically active recreation. In our model accommodating self-selection, the impact of bicycling facility density on the OH physically active recreation category was statistically insignificant (see the blank cell corresponding to the OH physically active recreation column for the bicycling facility density variable in Table 5.3). However, when we estimated an independent time-investment model without considering self-selection, the coefficient on bicycling facility density corresponding to the OH physically active recreation activity purpose was positive and highly significant. Clearly, this shows the danger of estimating the effects of the activity-travel environment on activity-travel choices without considering self-selection effects. More generally, it is important to model residential location choice and activity time-use choices in a joint framework to obtain the "true" effects of ATE attributes on activity time-use choices.

5.5.4 Overall Likelihood-based Measures of Fit

The log-likelihood value at convergence for the final joint multinomial logit (MNL)–multiple discrete-continuous extreme value (MDCEV) model is –42759.43. The corresponding value for the independent MNL and MDCEV models with no allowance for residential self selection due to unobserved decision-maker attributes is –42763.84. The likelihood ratio index for testing the presence of residential self selection due to unobserved factors is 8.82, which is larger than the critical chi-square value with 1 degree of freedom at a level of significance greater than 99.5% (the 1 degree of freedom corresponds to the standard deviation of the common error component related to the coefficient on bicycling facility density variable). Further, the log-likelihood value is –44127.17 for the independent MNL and MDCEV model system with (a) equal probability for each of the 236 spatial alternatives in the residential location MNL model and (b) only constants in the baseline preference terms, and the satiation and translation terms, in the activity time-use MDCEV model. The likelihood ratio index for testing the presence of exogenous variable effects and residential self selection effects is 2735.48, which is substantially larger than the critical chi-square value with 88 degrees of freedom at any reasonable level of significance. This clearly underscores the value of the model estimated in this chapter to predict residential location and time-use choices of individuals as a function of relevant exogenous variables and accommodating residential self-selection effects.

5.6 SUMMARY, CONCLUSIONS, AND LIMITATIONS

5.6.1 Summary and Conclusions

This study contributes to both the built environment and travel behavior modeling and activity-based analysis literature by presenting a joint model of residential location choice and activity time-use choices. This modeling system considers a comprehensive set of activity-travel environment (ATE) variables and sociodemographic variables as potential determinants of weekday time-use choices. The model formulation takes the form of a joint mixed Multinomial Logit–Multiple Discrete-Continuous Extreme Value (MNL–MDCEV) structure that controls for the self-selection of individuals into neighborhoods based on observed and unobserved activity time-use preferences. To our knowledge, the analysis in this chapter presents the first instance of the formulation and application of such a unified econometric methodological framework for jointly modeling (a) residential location choice (an unordered multinomial discrete choice variable), and (b) activity time-use choice (a multivariate discrete-continuous choice variable), while also accounting for observed and unobserved heterogeneity in the choice processes. Further, this is the first study that accommodates residential self-selection effects simultaneously in multiple activity participation in time-use decisions. The joint model system is estimated on a sample of 2793 households and individuals residing in Alameda County in the San Francisco Bay Area.

The model results offer several insights regarding individual time-use choices. First, both household/individual sociodemographics, and ATE attributes, are influential in shaping activity time-use choices. However, sociodemographic characteristics are more influential in shaping activity time-use behavior than are the ATE attributes. Second, there are significant observed factors contributing to residential self-selection based on activity time-use preferences. For instance, individuals with bicycles locate themselves into neighborhoods with good bicycling facilities. These same individuals also have a preference for physically active pure recreation pursuits. Ignoring this effect of bicycle ownership would lead to an inflated estimate of the effect of bicycling facility density on time invested in physically active recreation. Similarly, high income households locate in neighborhoods with low employment density and low street block density. Individuals from such high income households also have a preference for out-of-home recreational activities/travel. Thus, ignoring income effects in activity time-use choices can lead to a spuriously estimated negative effect of employment density and street block density on out-of-home recreational activities/travel. Overall, even if there are no common unobserved factors influencing residential choice and activity time-use choices, the results suggest that it behooves the analyst to estimate an independent residential choice model so that any observed demographic factors impacting the sensitivity to ATE attributes in residential choice can be considered in the time-use model. In this way, one can reduce the possibility of "corrupt" inferences regarding the impact of ATE attributes on time-use choices. Of course, another reason to model both residential choice and time-use choice, even in the absence of common unobserved factors, is the fact that ATE attributes impact both these choices. Thus, policy decisions regarding changes in ATE characteristics have to be evaluated in the context of spatial relocations as well as time-use shifts to obtain a comprehensive picture of the changes due to ATE-related policies. Third, people who are predisposed to physically active lifestyles appear to self-select themselves into zones with very good bicycling facility density for their residence. Ignoring such unobserved factors resulted, in our current empirical study, in a spuriously estimated positive effect of the bicycling density variable on physically active recreational activity participation. This result shows the danger of estimating the effects of the activity-travel environment on activity-travel choices without considering self-selection effects due to unobserved factors. Fourth, ATE attributes such as employment density, activity opportunity density/intensity, and bicycling facility density have statistically significant impacts on activity time-use decisions even after controlling for residential sorting effects due to both observed and unobserved factors. Fifth, our results emphasize the transportation-public health connection. That is, our analysis suggests that policies and educational campaigns aimed at reducing motorized vehicle ownership and increasing bicycle ownership, when combined with better provision of bicycling facilities, not only can lead to traffic congestion alleviation, but can also play an important role in improving public health through increased investment of time in physically active pursuits.

To summarize, our results indicate that activity-travel environment attributes are not “completely” exogenous in activity time-use decisions. Households and individuals locate themselves in ATEs that are consistent with their lifestyle preferences, attitudes, and values. In other words, households and individuals make residential location and activity time-use decisions jointly as part of an overall lifestyle package. Nevertheless, the findings from this research suggest that modifying the activity-travel environment can bring about changes in activity time-use patterns.

5.6.2 Limitations

The research in the current chapter can be extended and enhanced by (1) undertaking the analysis at a more disaggregate spatial level of analysis than traffic analysis zones (though such a move also promises to raise some very important computational challenges), and (2) applying the methodology developed in this chapter to richer data sets with attitudinal variables that may further enhance our understanding of the relationship of ATE attributes on activity-travel dimensions.

An important limitation of the modeling structure used in this chapter pertains to the interdependence among (or similarity of) the choice alternatives of the multiple discrete-continuous activity time-use component of the joint model system. That is, the model structure does not recognize that individuals may perceive several alternatives (*i.e.,* the activity types) as similar to each other, and that such similar alternatives (*i.e.,* the activity types) may be more substitutable to each other than to other alternatives. In other words, individuals may perceive the entire choice set of activity time-use alternatives as nests (or subsets, or groups) of similar alternatives, with each nest of alternatives associated with a certain extent of similarity and corresponding substitutability. Such similarity and substitutability of alternatives (*i.e.,* the activity types) can be attributed (to a certain extent) to the presence of common observed sociodemographic factors as well as common unobserved factors affecting the consumption of those alternatives. The magnitude of such common factors contributes to the degree of similarity, interdependence, or substitutability. Econometrically speaking, due to the presence of common unobserved factors affecting the consumption of a subset of alternatives, the stochastic components (or the error terms) associated with the utility expressions of those alternatives (or activity types) may be correlated with each other. Such inter alterative correlations give rise to the interdependence and higher substitution rates among the alternatives (or activity types). In this chapter, the interalternative correlations among the choice alternatives of the multiple discrete-continuous activity time-use component of the joint model system were ignored, and hence the actual substitution rates present in the data may not have been captured.[36] Such a limitation can be overcome by including common error components in the utility expressions of the alternatives (*i.e.,* activity types) to introduce correlations across the utilities of the alternatives. That is, the interalternative correlations can be "*simulated*" by use of common error components. However, including such common error components adds to the number of error components that are already present in the model system (*i.e.,* the v_{ql}, η_{qkl}, and ω_{qkl} terms for $l = 1, 2, \ldots, L$). As the number of error components increase, the parameter estimation may become difficult due to potential identification problems (see walker, 2002). Hence, it is important to develop econometric methods that can capture the interalternative correlations among the alternatives of the multiple discrete-continuous activity time-use model component (*i.e.,* the activity types) without resorting to the use of common error components. The next chapter develops a multiple discrete-continuous nested extreme value (MDCNEV) model that can be used to accommodate the interalternative correlations across the various activity types of the multiple discrete-continuous activity time-use model component of the joint model system developed in this chapter.

[36] The reader will note here that, in this chapter, the interalternative correlations among the choice alternatives of the single discrete residential location choice model component were also ignored. However, several studies in the literature have addressed the issue of interalternative correlations in the case of single discrete residential location choice models. Hence, this discussion is focused on the interalternative correlations in the context of the multiple discrete-continuous activity time-use component of the joint model system.

CHAPTER 6

A MULTIPLE DISCRETE-CONTINOUS NESTED EXTREME VALUE (MDCNEV) MODEL

6.1 BACKGROUND

Several consumer demand choice situations are characterized by multiple discreteness (*i.e.,* the simultaneous choice of one or more alternatives from a set of alternatives that are not mutually exclusive) as opposed to single discreteness (*i.e.,* the choice of a single alternative from a set of mutually exclusive alternatives). In addition, there can be a continuous choice corresponding to the amount of consumption (or the allocation of resources such as time and money) of each chosen alternative, which leads to a multiple discrete-continuous nature of choice situations. Individual activity time-use decisions modeled in the previous chapter can be viewed as a result of multiple discrete-continuous choice making. An individual may participate in different types of activities such as social activities, recreational activities, *etc,* within a given time period such as a day or a week. In addition to the activity participation decisions, the individual makes decisions on the amount of time allocated to each type of activity.

As discussed in the literature review section of the previous chapter, in the recent literature, several important choice situations, including individual activity time-use (Bhat, 2005), household travel expenditures (Rajgopalan and Srinivasan, 2008), and household vehicle ownership and usage (Sen and Bhat, 2006) have been analyzed as multiple discrete-continuous choice situations. A variety of econometric model structures have been used to analyze these choice situations, which can be broadly classified into: (a) multivariate single discrete-continuous modeling frameworks (see for example, Srinivasan and Bhat, 2006), and (b) utility maximization-based modeling frameworks (Hanemann, 1978; Wales and Woodland, 1983; Kim *et al.,* 2002; Phaneuf *et al.,* 2000; and Bhat, 2005 and 2008).

Among the available modeling frameworks, the recently proposed multiple discrete-continuous extreme value (MDCEV) model structure (Bhat, 2005 and 2007) is particularly attractive because of at least two salient features. First, the MDCEV model offers closed-form consumption probability expressions and thus obviates the need for simulation-based estimation. Second, the MDCEV model structure simplifies to the well-known multinomial logit (MNL) model when each (and every) decision maker chooses a single alterative out of the available alternatives in the choice set. Further, the MCEV model is well grounded in utility maximization theory and captures several features of consumer choice making, including the diminishing nature of marginal utility with increasing consumption. Due to these attractive features, the MDCEV model structure has been used in the previous chapter to model the activity time-use component of the joint residential location and activity time-use model system.

An important limitation of the MDCEV model formulation, however, is the neglect of potential interdependence among (or, the similarity of) the choice alternatives. In the context of the joint model developed in the previous chapter, this limitation corresponds to the neglect of the interdependence among (or similarity of) the activity type choice alternatives of the activity time-use model component. This is due to an assumption that the stochastic components (or the error terms) associated with the utility expressions of the activities are assumed to be independent and identically distributed (IID). This assumption of the MDCEV model implies that the utilities of different alternatives are uncorrelated, and is analogous to the IID error terms assumption in the multinomial logit (MNL) model. Such an assumption, however, cannot be justified in several empirical model specifications. The presence of common unobserved factors affecting the utilities of different alternatives leads to a violation of the assumption.

To understand this better, consider the case of individual-level activity participation and time-use decisions. Individuals make decisions on the types of activities (such as in-home physically active recreation, in-home physically inactive recreation, out-of-home physically active recreation, and out-of-home physically inactive recreation) to participate in, and the amount of time to spend in each activity. While making such decision, individuals may perceive several types of activities as similar to each other, and that such activity types may be more substitutable to each other than to other activity types. In other words, individuals may perceive the entire choice set of activity time-use alternatives as nests (or subsets, or groups) containing similar activities, with each nest of alternatives associated with a certain extent of similarity and corresponding substitutability. Such similarity and substitutability of different types of activities can be attributed (to a certain extent) to the presence of common observed sociodemographic factors as well as common unobserved factors affecting the participation of those activities. The magnitude of such common factors contributes to the degree of similarity, interdependence, or substitutability. Econometrically speaking, due to the presence of common unobserved factors affecting the consumption of a subset of activities, the stochastic components (or the error terms) associated with the utility expressions of those activities may be correlated with each other. Such inter alterative correlations give rise to the interdependence and higher substitution rates among subsets of activities.

To understand this further, consider for example, the utilities of in-home and out-of-home physically active activities. There may be several unobserved factors (such as physical fitness orientation) that influence the utilities of both (in-home and out-of-home) physically active activities. Such common unobserved factors lead to correlations, as well as high amount of sensitivity (or substitution), between the two types of activities just identified. In such case, creation of additional opportunities for out-of-home physically active activities may lead to a larger shift in the choices from in-home physically active activity participation to out-of-home physically active activity participation rather than from other types of activities to out-of-home physically active activity participation. That is, one may expect a higher amount of substitution between the two types (in-home and out-of-home) of physically active activities due to correlations between these activities. Neglect of such inter-alternative correlations through a simplifying IID assumption on the utilities can potentially result in a misrepresentation of the substitution patterns among the choice alternatives under consideration. As a result, ignoring the inter-alternative correlations can potentially lead to statistically inferior model fit, biased estimation of model parameters, and distorted policy implications.

Several empirical applications of multiple discrete-continuous choice occasions in the literature have found statistically significant inter-alternative correlations that violate the IID assumption of the MDCEV model. However, in most of the empirical applications to date, the IID assumption is relaxed through a mixed MDCEV (or MMDCEV) model formulation. Such a mixed model formulation employs common random terms (or common error components) across subsets of alternatives to capture the inter-alternative correlations (due to common unobserved factors) among the alternatives belonging to the subsets. The consumption probabilities resulting from the MMDCEV specification are integrals that do not have closed-form expressions unlike in the case of the MDCEV model. This warrants a simulation-based estimation, where the consumption probabilities are evaluated by using simulation-based techniques such as Mote-Carlo simulation or Quasi Monte-Carlo simulation. Simulation-based estimation, however, is computationally intensive and can potentially be infected with technical problems associated with the accuracy of simulation and the identification of parameters (see Walker, 2002). In the context of the model system developed in the previous chapter, including such common error components adds to the number of error components that are already present in the model system (*i.e.,* the v_{ql}, η_{qkl}, and ω_{qkl} terms for $l = 1, 2, \ldots, L$), thus making it more vulnerable to the problems associated with simulation-based estimation. Hence, it is important to develop multiple discrete-continuous model structures that can capture inter-alternative correlations and at the same time offer closed-form expressions for consumption probabilities.

In this chapter, we propose a multiple discrete-continuous nested extreme value (MDCNEV) model that captures inter-alternative correlations among alternatives in mutually exclusive subsets (or nests) of the choice set, and also offers closed-form probability expressions for any consumption pattern. This extension of the MDCEV model is analogous to the nested logit extension of the MNL model. Such a model structure can be used to accommodate the interalternative correlations across the various activity types of the multiple discrete-continuous activity time-use model component of the joint model system developed in the preceding chapter.

To be sure, Bhat (2008) has already identified a general extension of the MDCEV model to a multiple discrete-continuous generalized extreme value (MDCGEV) model that can accommodate very general patterns of correlations by using a GEV error structure and at the same time yield closed-form expressions for consumption probabilities. However, in his paper, it is mentioned that "*the derivation [of the consumption probability expressions] is tedious and the expressions get unwieldy*". Further there is no proof of the existence of closed-form probability expressions from an MDCGEV model. His paper provided expressions for only a specific and simple nested logit error structure with 4 alternatives. In this chapter, in Section 6.2, we prove the existence of closed-form probability expressions, and derive a general and compact form of the expression, for any and all consumption patterns in the case of a general two-level nested MDCNEV error structure. Summary and scope for future work is discussed in Section 6.3.

6.2 THE MDCNEV MODEL: A TWO LEVEL NESTED CASE

In this section, we describe the econometric model structure of the MDCNEV model for a general two-level nested case. Section 6.2.1 draws from Bhat (2008) to describe the basic utility structure and the Kuhn-Tucker (KT) conditions used to solve the consumer demand problem involving multiple discrete-continuous choices. Section 6.2.2 builds on Section 6.2.1 by: (a) explaining the econometric structure for a two-level nested MDCNEV model, and (b) proving the existence of a closed-form consumption probability expression and deriving the compact and general form of the expression for any and all consumption patterns.

6.2.1 Utility Structure and KT Conditions

We consider the following functional form for utility proposed by Bhat (2008):

$$U(\boldsymbol{t}) = \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left\{ \left(\frac{t_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right\};\ \psi_k > 0, \alpha_k \le 1, \gamma_k > 0 \tag{6.1}$$

where $U(\boldsymbol{t})$ is the total utility derived from consuming non-negative (either zero or non-zero) amounts of each of the K alternatives (or goods) available to the decision maker, $\boldsymbol{t}$ is the consumption quantity (Kx1)-vector with elements t_k ($t_k \geq 0$ for all k), and ψ_k , γ_k and α_k are parameters associated with alternative (or good) k.

Stochasticity is introduced into the above utility specification by through a multiplicative random term in the ψ_k parameter of each good as follows:

$$\psi(z_k, \varepsilon_k) = \exp(\beta' z_k + \varepsilon_k) \tag{6.2}$$

where, z_k is a set of attributes characterizing alternative k and the decision-maker, and ε_k captures unobserved characteristics that impact the baseline utility for good k.

From the analyst's perspective, the individual maximizes the random utility subject to a linear budget constraint and non-negativity constraints on t_k :

$$\sum_{k=1}^{K} t_k = T \text{ (where } T \text{ is the total budget) and } t_k \geq 0 \ \forall \ k \ (k = 1, 2, ..., K) \tag{6.3}$$

The analyst can solve for the optimal consumptions by forming the Lagrangian and applying the Kuhn-Tucker (KT) conditions. The Lagrangian function for the problem is:

$$\mathscr{L} = \sum_k \frac{\gamma_k}{\alpha_k} \left[\exp(\beta' z_k + \varepsilon_k)\right] \left\{ \left(\frac{t_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right\} - \lambda \left[\sum_{k=1}^{K} t_k - T \right], \tag{6.4}$$

where λ is the Lagrangian multiplier associated with the budget constraint. The KT first-order conditions for the optimal consumptions $(t_k^*; k = 1, 2, ..., K)$ are given by:

$$\exp(\beta' z_k + \varepsilon_k) \left(\frac{t_k^*}{\gamma_k} + 1 \right)^{\alpha_k - 1} - \lambda = 0, \text{ if } t_k^* > 0, \ (k = 1, 2, ..., K) \tag{6.5}$$

$$\exp(\beta' z_k + \varepsilon_k) \left(\frac{t_k^*}{\gamma_k} + 1 \right)^{\alpha_k - 1} - \lambda < 0, \text{ if } t_k^* = 0, \ (k = 1, 2, \ldots, K)$$

Now, without any loss of generality, designate activity purpose 1 as a purpose to which the individual allocates some non-zero amount of consumption. For the first good, the KT condition may then be written as:

$$\lambda = \exp(\beta' z_k + \varepsilon_k)\left(\frac{t_1^*}{\gamma_1} + 1\right)^{\alpha_1 - 1} \tag{6.6}$$

Substituting for λ from above into Equation (6.5) for the other activity purposes (k = 2,…, K), and taking logarithms, we can rewrite the KT conditions as:

$$V_k + \varepsilon_k = V_1 + \varepsilon_1 \text{ if } t_k^* > 0, \; (k = 2, 3, \ldots, K)$$

$$V_k + \varepsilon_k < V_1 + \varepsilon_1 \text{ if } t_k^* = 0, \; (k = 2, 3, \ldots, K) \tag{6.7}$$

where, $V_k = \beta' z_k + (\alpha_k - 1)\ln\left(\frac{t_k^*}{\gamma_k} + 1\right)$, ($k$ = 1, 2, 3,…, K).

6.2.2 The MDCNEV Model: Econometric Structure and Probability Expressions

The econometric specification (*i.e.,* the error term structure) determines the form of the consumption probability expressions. In the general case, let the joint probability density function of the ε_k terms be $g(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_K)$, and let M alternatives be chosen out of the available K alternatives, and that the consumption amounts of the M goods be $(t_1^*, t_2^*, t_3^*, \ldots, t_M^*)$. The stochastic KT conditions of Equation (6.7) can be used to write the joint probability expression of the consumption pattern as follows (Bhat, 2008):

$$P(t_1^*, t_2^*, t_3^*, \ldots, t_M^*, 0, 0, \ldots, 0) = |J| \int_{\varepsilon_1=-\infty}^{+\infty} \int_{\varepsilon_{M+1}=-\infty}^{V_1 - V_{M+1} + \varepsilon_1} \int_{\varepsilon_{M+2}=-\infty}^{V_1 - V_{M+2} + \varepsilon_1} \cdots \int_{\varepsilon_{K-1}=-\infty}^{V_1 - V_{K-1} + \varepsilon_1} \int_{\varepsilon_K=-\infty}^{V_1 - V_K + \varepsilon_1}$$

$$g(\varepsilon_1, V_1 - V_2 + \varepsilon_1, V_1 - V_3 + \varepsilon_1, \ldots, V_1 - V_M + \varepsilon_1, \varepsilon_{M+1}, \varepsilon_{M+2}, \ldots, \varepsilon_{K-1}, \varepsilon_K) \tag{6.8}$$

$$d\varepsilon_K d\varepsilon_{K-1} \ldots d\varepsilon_{M+2} d\varepsilon_{M+1} d\varepsilon_1 ,$$

where J is the Jacobian whose elements are given by (see Bhat, 2005) $J_{ih} = \frac{\partial[V_1 - V_{i+1} + \varepsilon_1]}{\partial t_{h+1}^*} = \frac{\partial[V_1 - V_{i+1}]}{\partial t_{h+1}^*}$; $i, h = 1, 2, \ldots, M-1$.

In this paper, we rewrite the above probability expression as an integral of the Mth order partial derivative of a K-dimensional joint cumulative distribution of error terms $(\varepsilon_1, \varepsilon_2, .., \varepsilon_K)$:

$$P(t_1^*, \ldots, t_M^*, 0, \ldots, 0) = |J| \int_{\varepsilon_1=-\infty}^{+\infty} \frac{\partial^M}{\partial\varepsilon_1 \ldots \partial\varepsilon_M} F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_K) \Big|_{\varepsilon_2 = V_1 - V_2 + \varepsilon_1, \varepsilon_3 = V_1 - V_3 + \varepsilon_1, \ldots, \varepsilon_K = V_1 - V_k + \varepsilon_1} \, d\varepsilon_1 \tag{6.9}$$

where, $F(\varepsilon_1, \varepsilon_2, ..., \varepsilon_K)$ is the joint cumulative distribution of error terms $(\varepsilon_1, \varepsilon_2, .., \varepsilon_K)$. The reader will note here that the order of the partial derivative in the above expression is equal to the number of chosen alternatives (M), and that the differentials in the partial derivative are with respect to the stochastic utility terms (or the error terms) of each chosen alternative.

For the stochastic components of the utility (or the error terms), we assume a nested extreme value distributed structure that has the following joint cumulative distribution:

$$F(\varepsilon_1, \varepsilon_2, .., \varepsilon_K) = \exp\left[-\sum_{\mathfrak{s}=1}^{S_K} \left\{ \sum_{i \in \mathfrak{s}^{\text{th}} \text{nest}} \exp\left(-\frac{\varepsilon_i}{\theta_{\mathfrak{s}}} \right) \right\}^{\theta_{\mathfrak{s}}} \right] \qquad (6.10)$$

In the above expression, $\mathfrak{s}(=1, 2, ..., S_M, ..., S_K)$ is the index to represent a nest of alternatives, S_K is the total number of nests the K alternatives belong to, and S_M is the total number of nests the chosen *M* alternatives belong to. It is assumed that the nests are mutually exclusive and exhaustive (*i.e.,* each alternative can belong to only one nest and all alternatives are allocated to one of the S_K nests). $\theta_{\mathfrak{s}}(0 < \theta_{\mathfrak{s}} \leq 1; \mathfrak{s} = 1, 2, ..., S_K)$ is the (dis)similarity parameter introduced to induce correlations among the stochastic components of the utilities of alternatives belonging to the $\mathfrak{s}^{th}$ nest[37]. The utilities of alternatives that belong to different nests, however, are still uncorrelated. Thus the nested extreme value distributed error structure allows inter-alternative correlations within mutually exclusive subsets of alternatives.

Using the expression for joint cumulative distribution function of the nested extreme value error structure from Equation (6.10), the expression for $F(\varepsilon_1, \varepsilon_2, .., \varepsilon_K)\big|_{\varepsilon_2 = V_1 - V_2 + \varepsilon_1, ..., \varepsilon_K = V_1 - V_k + \varepsilon_1}$ in Equation (6.9) can be expressed as $\exp\left(-e^{-\varepsilon_1} f\right)$, and the probability expression given in Equation (6.9) can be rewritten as:

---

[37] It may be verified here that when $\theta_{\mathfrak{s}} = 1 \, \forall \, \mathfrak{s}(\mathfrak{s} = 1, 2, ... S_M, ... S_K)$, the cumulative distribution function in Equation (10) simplifies to the cumulative distribution function of independent (or uncorrelated) type-1 extreme value distributed error terms that gives rise to Bhat's (2007) MDCEV model probability expressions.

$$P(t_1^*,...,t_M^*,0,...,0) = |J| \int_{\varepsilon_1=-\infty}^{+\infty} \frac{\partial^M}{\partial\varepsilon_1...\partial\varepsilon_M} \exp\left(-e^{-\varepsilon_1} f\right) \mathrm{d}\varepsilon_1,$$

$$\text{where } f = \sum_{\varsigma=1}^{S_k} \left\{ \left(\sum_{i \in \varsigma^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_\varsigma}} \right)^{\theta_\varsigma} \right\} \tag{6.11}$$

Now, without loss of generality, let $1,2,...,S_M$ be the nests the M chosen alternatives belong to, let $q_1, q_2, ..., q_\varsigma, ..., q_{S_M}$ be the number of chosen alternatives in each of the S_M nests (hence $q_1 + q_2 + ... + q_\varsigma + ... + q_{S_M} = M$), and let $\varepsilon_{1\varsigma}, \varepsilon_{2\varsigma}, ..., \varepsilon_{q\varsigma}$ be the stochastic terms associated with each of the chosen alternatives in the ς^{th} nest. Also, for simplicity in notation, let $F(\varepsilon_1, \varepsilon_2, .., \varepsilon_K)\big|_{\varepsilon_2 = V_1 - V_2 + \varepsilon_1, ..., \varepsilon_K = V_1 - V_k + \varepsilon_1} \left(= \exp\left(-e^{-\varepsilon_1} f\right)\right)$ be represented as F. Using these notational preliminaries, the Mth order partial derivative of the K-dimensional cumulative distribution in Equation (6.11) can be simplified into a product of S_M number of smaller partial derivatives, one for each nest:

$$\frac{\partial^M F}{\partial\varepsilon_1 \partial\varepsilon_2 ... \partial\varepsilon_M} = F \prod_{\varsigma=1}^{S_M} \left(\frac{1}{F} \frac{\partial^{q_\varsigma} F}{\partial\varepsilon_{1\varsigma} \partial\varepsilon_{2\varsigma} ... \partial\varepsilon_{q_\varsigma}} \right) \tag{6.12}$$

The functional form of F, due to the independence of the stochastic terms across different nests, allows the Mth order partial derivative to be separated into such smaller partial derivatives. The reader will note here that the order of each smaller partial derivate in the above expression is equal to the number of chosen alternatives in the nest to which the partial derivate is associated with. From the S_M number of partial derivatives, consider the ${q_\varsigma}^{th}$ order partial derivative for the ς^{th} nest, which can be expanded as follows:

$$\frac{\partial^{q_\varsigma} F}{\partial\varepsilon_{1\varsigma} ... \partial\varepsilon_{q_\varsigma}} = F \left(\prod_{\substack{i \in \varsigma^{\text{th}}\text{nest, and} \\ i \in \{\text{chosen alts}\}}} e^{-\frac{V_1 - V_i}{\theta_\varsigma}} \right) \sum_{r_\varsigma=1}^{q_\varsigma} \left\{ e^{-\varepsilon_1 (q_\varsigma - r_\varsigma + 1)} \left(\sum_{i \in \varsigma^{\text{th}}\text{nest}} e^{-\frac{(V_1 - V_i)}{\theta_\varsigma}} \right)^{(q_\varsigma - r_\varsigma + 1)\theta - q_\varsigma} sum(X_{r_\varsigma}) \right\} \tag{6.13}$$

In the above expression, $sum(X_{r\vartheta})$ is a sum of elements of a row matrix $X_{r\vartheta}$. This matrix takes a form described in Appendix A.

Substitution of Equation (6.13) into Equation (6.12) and the resultant expression into Equation (6.11), and further expansion and algebraic rearrangements (shown in Appendix B) leads to the following expression for consumption probability:

$$P(t_1^*, t_2^*, \ldots t_M^*,\ 0, \ldots, 0) = |J| \left(\prod_{i \in \{\text{chosen alts}\}} e^{-\frac{V_1 - V_i}{\theta_i}} \right) \times$$
$$\times \sum_{r_1=1}^{q_1} \ldots \sum_{r_\vartheta=1}^{q_\vartheta} \ldots \sum_{r_{S_M}=1}^{q_{S_M}} \left\{ \prod_{\vartheta=1}^{S_M} \left(\sum_{i \in \vartheta^{\text{th}}\text{nest}} e^{-\frac{(V_1 - V_i)}{\theta_\vartheta}} \right)^{(q_\vartheta - r_\vartheta + 1)\theta_\vartheta - q_\vartheta} \prod_{\vartheta=1}^{S_M} sum(X_{r\vartheta}) \int_{\varepsilon_1=-\infty}^{+\infty} e^{-\varepsilon_1 \sum_{\vartheta=1}^{S_M} (q_\vartheta - r_\vartheta + 1)} \exp\left(-e^{-\varepsilon_1} f\right) d\varepsilon_1 \right\} \quad (6.14)[38]$$

The integral in the above Equation has the following closed-form expression:

$$I = \int_{\varepsilon_1=-\infty}^{+\infty} e^{-\varepsilon_1 \sum_{\vartheta=1}^{S_M} (q_\vartheta - r_\vartheta + 1)} \exp\left(-e^{-\varepsilon_1} f\right) d\varepsilon_1 = \frac{\left(\sum_{\vartheta=1}^{S_M} (q_\vartheta - r_\vartheta + 1) - 1 \right)!}{f^{\sum_{\vartheta=1}^{S_M} (q_\vartheta - r_\vartheta + 1)}} \quad \text{(proved in Appendix C)}$$

that leads to the following closed-form consumption probability expression for the MDCNEV model:

$$P(t_1^*, t_2^*, \ldots t_M^*,\ 0, \ldots, 0) =$$
$$|J| \left(\prod_{i \in \{\text{chosen alts}\}} e^{-\frac{V_1 - V_i}{\theta_i}} \right) \sum_{r_1=1}^{q_1} \ldots \sum_{r_\vartheta=1}^{q_\vartheta} \ldots \sum_{r_{S_M}=1}^{q_{S_M}} \left\{ \prod_{\vartheta=1}^{S_M} \left(\sum_{i \in \vartheta^{\text{th}}\text{nest}} e^{-\frac{(V_1 - V_i)}{\theta_\vartheta}} \right)^{(q_\vartheta - r_\vartheta + 1)\theta_\vartheta - q_\vartheta} \left(\prod_{\vartheta=1}^{S_M} sum(X_{r\vartheta}) \right) \frac{\left(\sum_{\vartheta=1}^{S_M} (q_\vartheta - r_\vartheta + 1) - 1 \right)!}{f^{\sum_{\vartheta=1}^{S_M} (q_\vartheta - r_\vartheta + 1)}} \right\}$$

(6.15)

After further algebraic rearrangements (details are available with the authors), the above expression simplifies to:

---

[38] In this Equation, θ_i is the dis(similarity) parameter associated with the nest to which alternative '*i*' belongs to. Both θ_i and θ_ϑ are used (at different places) in the expression to represent the same parameters (dissimilarity parameters) for the sake of convenience in notation and representation.

$$
P(t_1^*, t_2^*, \ldots t_M^*,\ 0, \ldots, 0) =
$$

$$
|J| \frac{\displaystyle\prod_{i \in \{\text{chosen alternatives}\}} e^{\frac{V_i}{\theta_i}}}{\displaystyle\prod_{\varsigma=1}^{S_M} \left(\sum_{i \in \varsigma^{\text{th}} \text{nest}} e^{\frac{V_i}{\theta_\varsigma}} \right)^{q_\varsigma}} \sum_{r_1=1}^{q_1} \ldots \sum_{r_\varsigma=1}^{q_\varsigma} \ldots \sum_{r_{S_M}=1}^{q_{S_M}} \left\{ \prod_{\varsigma=1}^{S_M} \left[\frac{\left(\displaystyle\sum_{i \in \varsigma^{\text{th}} \text{nest}} e^{\frac{V_i}{\theta_\varsigma}} \right)^{\theta_\varsigma}}{\displaystyle\sum_{\varsigma=1}^{S_k} \left\{ \left(\sum_{i \in \varsigma^{\text{th}} \text{nest}} e^{\frac{V_i}{\theta_\varsigma}} \right)^{\theta_\varsigma} \right\}} \right]^{q_\varsigma - r_\varsigma + 1} \left(\prod_{\varsigma=1}^{S_M} sum(X_{r\varsigma}) \right) \left(\sum_{\varsigma=1}^{S_M} (q_\varsigma - r_\varsigma + 1) - 1 \right)! \right\}
$$

(6.16)

The general expression above represents the MDCNEV consumption probability for any consumption pattern with a two-level nested extreme value error structure. This expression can be used in the log-likelihood formation and subsequent maximum likelihood estimation of the parameters for any dataset with mutually exclusive groups (or nests) of interdependent multiple discrete-continuous choice alternatives (*i.e.,* mutually exclusive groups of alternatives with correlated utilities). It may be verified that the MDCNEV probability expression in Equation (6.16) simplifies to Bhat's (2008) MDCEV probability expression when each of the utility functions are independent of one another (*i.e.,* $\theta_\varsigma = 1$ and $q_\varsigma = 1 \ \forall \varsigma$, and $S_M = M$). Also, one may verify that the above expression simplifies to the probability expressions derived by Bhat (2008) for a simple nested error structure with four alternatives. Finally, and importantly, it should be noted here that the nested extreme value extension developed in this paper is applicable not only for Bhat's MDCEV model, but also for all Kuhn-Tucker (KT)-based consumer demand model systems involving multiple continuous choices or multiple discrete-continuous choices (see von Haefen and Phaneuf, 2005 for a review of KT-demand model systems).

6.3 SUMMARY AND CONCLUSIONS

The multiple discrete-continuous activity time-use component of the joint model system used in the preceding chapter neglects the potential interdependence among (or, the similarity of) the choice alternatives. However, individuals may perceive the entire choice set of activity time-use alternatives as nests (or subsets, or groups) of similar activities, with each nest of alternatives associated with a certain extent of similarity and corresponding substitutability. Econometrically speaking, due to the presence of common unobserved factors affecting the consumption of a subset of activities, the stochastic components (or the error terms) associated with the utility expressions of those activities may be correlated with each other. Such inter alterative correlations give rise to the interdependence and higher substitution rates among subsets of activities.

Several multiple discrete-continuous analyses in the literature have neglected such inter-alternative correlations through a simplifying IID assumption on the utilities. This assumption can potentially result in a misrepresentation of the substitution patterns among the choice alternatives under consideration, statistically inferior model fit, biased estimation of model parameters, and distorted policy implications. To avoid these problems, several empirical applications of multiple discrete-continuous choice occasions in the literature relaxed the IID assumption through a mixed MDCEV (or MMDCEV) model formulation. Such a mixed model formulation, however, does not lead to closed-from expressions for consumption probabilities and hence warrants a simulation-based estimation.

This chapter developed a multiple discrete-continuous nested extreme value (MDCNEV) model that captures inter-alternative correlations among alternatives in mutually exclusive subsets (or nests) of the choice set. The model formulation results in closed-form probability expressions for any (and all) consumption patterns. This extension of the MDCEV model is an important milestone in the development of multiple discrete-continuous choice models and is analogous to the nested logit extension of the MNL model. Such a model structure can be used to accommodate the interalternative correlations across the various activity types of the multiple discrete-continuous activity time-use model component of the joint model system developed in the preceding chapter.

The MDCNEV model developed in this chapter captures inter-alternative correlations among alternatives in mutually exclusive subsets (or nests) of the choice set. In future research efforts, the model can be enhanced to incorporate more general interalternative correlation patterns such as cross-nested correlations.

CHAPTER 7

POLICY ANALYSIS: BUILT ENVIRONMENT IMPACTS AND THE EFFECT OF RESIDENTIAL SELF-SELECTION

7.1 INTRODUCTION

Chapters 2 through 5 of the dissertation developed multidimensional econometric choice model formulations to model the long-term residential location and short-term activity-travel related decisions in an integrated fashion. Empirical estimations using these model formulations shed light on several important issues surrounding the land-use travel behavior relationship. First, it is found that there are significant observed factors contributing to residential self-selection. Households self select their residential location based on demographic characteristics such as auto and bicycle ownership, income, household size, and race. Second, the presence of common unobserved heterogeneity across residential location choices and activity-travel choices indicated the presence of significant unobserved factors contributing to residential self-selection. Third, the built environment attributes have statistically significant impacts on commute mode choice even after controlling for residential self-selection. However, these findings from the model estimation results have not been corroborated with policy simulations. Further, the following questions have not been answered yet in the dissertation:

(1) Although the built environment impacts have come out to be statistically significant even after controlling for residential self-selection, do these impacts translate into discernible changes in travel behavior?

(2) How much does the built environment matter in comparison with the sociodemographics?

(3) How best to implement built environment policies?

(4) Is the effect of residential self-selection large enough to warrant the need to control for it by estimating the sophisticated models developed in this dissertation? If not, it would be better to use simpler and parsimonious model formulations.

This chapter attempts to answer these questions by undertaking several policy analyses. Specifically, Section 7.2 presents policy applications to answer the first three questions, and Section 7.3 presents a policy application to answer the fourth question. Section 7.4 provides a summary of the policy simulations and identifies additional policy questions that can be answered by using the models developed in this dissertation.

7.2 IMPACTS OF THE BUILT ENVIRONMENT

The short-term activity/travel choice components of the joint model systems estimated in the previous chapters can be applied to predict the short-term changes in activity-travel patterns due to changes in sociodemographic characteristics and built environment attributes. In this section, we present such policy simulations in the contexts of activity participation and time-use behavior (Section 7.2.1) and bicycle ownership (Section 7.2.2) to answer the first three questions raised above.

7.2.1 Impacts of the Built Environment on Time-use Behavior

This section presents a policy analysis using the time-use component of the joint model estimated in Chapter 5. The model estimates were used to quantify the effects on aggregate time-use patterns of changes in two sociodemographic characteristics and two built environment attributes. The sociodemographic changes correspond to a decrease in household vehicle ownership for each individual by one (except for zero car households whose car ownership level is left unchanged), and an increase in household bicycle ownership for each individual by one. The built environment attribute changes include a ten-fold increase in physically inactive recreation centers per square mile in each TAZ, and a ten-fold increase in bicycling facility density in each TAZ. To examine the effects of each of these changes, we computed the percentage change in the aggregate time investment in the five recreational activity categories (*i.e.,* IH recreation, OH physically active pure recreation, OH physically inactive pure recreation, OH physically active recreation, and OH physically inactive recreation) from before- to after- the change.[39] Table 7.1 presents the results based on the policy simulations for 150 individuals.

Table 7.1. Impact of Change in Built Environment Attributes and Sociodemographic Characteristics

	Predicted % change in aggregate time use in recreational activity purposes			
	Household vehicle ownership decreased by 1	**Household bicycle ownership increased by 1**	**Zonal-level physically inactive recreation centers per square mile increased ten-fold**	**Zonal-level bicycling facility density increased ten-fold**
IH Recreation	14.48	--	--	--
OH Physically Active Pure Recreation	31.47	22.03	--	17.38
OH Physically Inactive Pure Recreation	--	--	--	--
OH Physically Active Recreation	18.07	--	--	--
OH Physically Inactive Recreation	--	--	2.89	--

A "--" entry in the table indicates changes less than 1%. Four important observations may be made from this table. First, and as indicated earlier in Chapter 5, policies and educational campaigns aimed at reducing motorized vehicle ownership and increasing bicycle ownership not only can lead to traffic congestion alleviation, but can also play an important role in improving public health (as evidenced by the increase in overall time-use in the OH physically active recreation categories in the table). Second, the activity-travel environment can be used as a tool to influence individual activity time-use patterns. In the table, this is reflected in the change in OH physically inactive recreation time due to the change in the number of physically inactive recreation centers

[39]We do not show the effects on other activity purposes in the table to reduce clutter and because these other effects are rather small. Also, to keep the constrained optimization-based prediction process manageable (see Bhat, 2005 for the prediction process details), we randomly chose 150 individuals from the estimation sample for the prediction analysis.

per square mile, and the change in OH physically active pure recreation time due to the change in bicycling facility density. Third, while the results show evidence that the built environment can be engineered to influence activity time-use, the results also show that the ability to do so is very limited. Specifically, the changes in activity time-use are highly inelastic to changes in built environment attributes. Fourth, while the percentage changes in activity time-use due to changes in sociodemgraphics and changes in built environment attributes are not strictly comparable (because the sociodemographic variables correspond to a change by 1 to ordinal variables, while the built environment attributes correspond to a 1000% increase to continuous variables), there is an indication that sociodemographics play a far more dominant role in determining activity time-use than do built environment attributes.

7.2.2 Impacts of the Built Environment on Bicycle Ownership

In this section, the coefficient estimates of the bicycle ownership component of the heterogeneous-joint model system developed in Chapter 4 have been applied to predict the changes (*i.e.,* the aggregate-level elasticities) in bicycle ownership due to changes in socio-demographic characteristics and built environment attributes. The impact of the built environment variables was computed in two ways: (1) by computing the elasticity effect of the binary bicycle friendly residential neighborhood type variable[40], and (2) by computing the elasticity effect due to a 25% change in the built environment attributes used to define the residential neighborhood type variable.

Table 7.2 shows the aggregate level elasticity effects on the expected bicycle ownership levels in the estimation sample (of about 5000 households). Several observations can be made from the table. First, the elasticity effects of the socio-demographic variables are higher than those of either the residential neighborhood type variable or the built environment attributes. Second, the elasticity effects of the built environment attributes are smaller compared to that of the neighborhood type variable. While the elasticity effect of the built environment variables represents the effect of addressing individual attributes of a neighborhood, the elasticity effect of the

[40] See Chapter 4 for details on the binary bicycle friendly residential neighborhood type variable.

neighborhood type variable represents an overall effect of changes in all of the built environment variables. This indicates that built environment policies may be more effective when used in combination (*i.e.,* for example, increase the bicycling facilities in the neighborhood, and also increase the connectivity of the bicycle route network to other neighborhoods) rather than modifying individual elements (such as increase only the bicycle facilities) of a neighborhood. Third, although the elasticity effects of the built environment attributes are smaller in magnitude compared to those of the sociodemographics, the magnitudes are not negligible and indicate discernible impacts of the built environment on bicycle ownership levels.

TABLE 7.2 Elasticity Effects of Variables in the Bicycle Ownership Component of the Heterogeneous-Joint Model of Chapter 4

Sociodemographic variables in the bicycle ownership component of the heterogeneous-joint model	**Elasticity Effect (%)**
Number of active adults in the household	13.04
Number of children (of age < 5 years) in the household	17.95
Number of children (of age between 5 and 16) in the household	47.47
Number of students in the household	13.82
Single person household	-12.65
Age of householder greater than 60 years	-25.10
Householder is male	5.20
Caucasian household	24.10
Household annual Income in 10,000s of dollars	4.48
Household lives in a single family dwelling unit	16.53
Own household	11.57
Residential neighborhood type variable	7.50
Built Environment variables used to define the neighborhood type	
Bicycle lane density (mileage per square mile)	1.22
Number of zones accessible from the home zone by bicycle	1.27
Street block density (mileage per square mile)	1.42
Number of physically active and natural recreational centers in the zone	1.01

7.3 THE IMPACT OF RESIDENTIAL SELF-SELECTION

In this section, we present a policy simulation analysis to assess if the impact of residential self-selection is large enough to warrant the need for controlling it. More specifically, we test the impact of self-selection effects due to unobserved factors that require advanced (*i.e.,* joint/integrated) model formulations to control for.[41]

To address this issue of residential self-selection due to unobserved factors, we utilized the time-use model component that was estimated jointly with the residential location choice model (presented in Chapter 5), and a time-use model that was estimated independently. Let us label the former model as the time-use model that considers residential self-selection, and the latter model as the time-use model that ignores residential self-selection. The model that considers self-selection yielded a statistically significant common unobserved heterogeneity in the impact of the built environment variable "bicycle facility density" across residential location and physically active out-of-home recreational activity (such as playing outdoor sports, exercising at gym, *etc.*) participation choices. This indicated statistically significant self-selection effects due to unobserved factors in the impact of bicycle facility density on physically active out-of-home recreational activity participation and time allocation. That is, people who are predisposed to physically active lifestyles (which are usually unobserved in typical travel survey data) appear to self-select themselves into zones with very good bicycling facility density for their residence. After controlling for such unobserved self-selection effects, the coefficient of the variable "bicycle facility density" on out-of-home physically active activity participation was statistically insignificant. Ignoring the unobserved self-selection effects, however (by estimating a time-use model without jointly considering residential location choice), resulted in a spuriously estimated statistically significant positive coefficient of the bicycling facility density variable on physically active out-of-home recreational activity participation. The question is "how much is the spuriously

[41] We do not test the magnitude of self-selection due to observed sociodemographic factors, because such effects can be accommodated by simply including the sociodemographic attributes as explanatory variables in the activity-travel models. That is, there is no need of advanced (*i.e.,* joint/integrated) model formulations to accommodate self-selection due to observed factors.

estimated effect due to residential self-selection?" Is it large enough to warrant the need to consider and disentangle this self-selection effect from the "true" causal impact of bicycle facility density on physically active out-of-home recreational activity participation?

Table 7.3 presents the results of policy simulation analysis to answer the above raised questions. The table presents the predictions (on 150 randomly chosen individuals from the estimation data) of four different measures of time-use in out-of-home physically active recreational activities. The predictions of the time-use model that considers self-selection are shown in column 2, and those of the time-use model that ignores self-selection are shown in column 3.

Table 7.3 Policy Analysis to Test the Impact of Residential Self-Selection Effects

Measure of time-use in physically active out-of-home recreational activities	**Predictions of…**		**Over prediction of the time-use model that ignores self-selection (with respect to the predictions of the model that considers self-selection)**
	The time-use model that considers self-selection	**the time-use model that ignores self-selection**	
Number of individuals undertaking the activity	12	13	8.33 %
Total amount of time spent by all 150 individuals in the activity	18 hours	20 hours	11.11 %
Average amount of time spent (per individual in the sample)	7.2 minutes	8 minutes	11.11 %
Average amount of time spent (per individual participating in the activity)	83 minutes	100 minutes	20.48 %

The analysis results show that the model that ignored residential self-selection over predicts the time-use in physically active out-of-home recreational activities. The over prediction can be attributed to the neglect of residential self-selection while estimating the relationship between bicycle facility density and participation in physically active out-of-home recreational activities. The extent of over prediction in the number of individuals undertaking the activity is 8.33 percent. On the other hand, the extent of over prediction in the time allocated to the activity is 11.11 percent for all individuals and is much higher (20.48 percent) for those individuals who were predicted to participate in the activity. These over predictions suggest a possibility of large errors in the assessment of the impact of built environment attributes at an aggregate level in the population. Over all, the results suggest a significant over prediction in the out-of-home physically active activity participation behavior by the model that ignored residential self-selection. These findings substantiate the findings from the empirical estimation results of chapter 5, and indicate that the neglect of residential self-selection due to unobserved factors can result in non-negligible amounts of spuriously estimated impacts of the built environment.

7.4 SUMMARY AND DISCUSSION

7.4.1 Summary

Gathering all the evidence from the above presented policy analyses as well as the model estimations of the previous chapters, the questions raised in the beginning of the chapter can be answered as below.

(1) Although the built environment impacts have come out to be statistically significant even after controlling for self-selection effects, do these impacts translate into discernible changes in travel behavior?

The direct influence of the built environment changes on the time-use behavior patterns may be very limited (see Section 7.2.1). However, such policies appear to have a discernible direct impact on bicycle ownership (see Section 7.2.2). Further, given that changes in bicycle ownership appear to have large impacts on time-use behavior (See Section 7.2.1), the built environment polices may have a discernible indirect impact on time-use behavior as well. The question remains to be answered, in the context of time-use behavior, by undertaking a comprehensive analysis of the direct and indirect (through bicycle ownership and vehicle ownership) impacts of the built environment. This question remains to be answered in the mode choice[42] and auto ownership contexts as well. Further, extensive policy analyses should be carried out with different built environment attributes to determine which neighborhood attribute(s) matter the most for which activity-travel choice dimension. Perhaps, as suggested by the analysis in Section 7.2.2, several built environment attributes together may have considerably discernible influence on most aspects of activity-travel behavior. In such case, it is important to find out the combination of policies and the means of implementation that would best serve the purpose.

(2) How much does the built environment matter in comparison with the sociodemographics?

Sociodemographics appear to play a far more dominant role than does the built environment in shaping activity and travel behavior aspects such as activity participation, time-use, and bicycle ownership.

(3) How best to implement built environment policies?

Built environment changes should be implemented in combination (rather than in isolation) for achieving synergistic results. Further research needs to understand what combination of policies would yield desirable results.

(4) Is the impact of residential self-selection large enough to warrant the need to control for it by estimating the sophisticated models developed in this dissertation?

Yes, neglecting the residential self-selection effects due to unobserved factors (when present) may lead to large errors in the assessment of the impact of built environment attributes at an aggregate level. Hence, it is important to estimate the joint residential location and activity-travel choice models developed in this dissertation. At the least, the joint models should be estimated to test the presence of self-selection effects.

[42] The integrated model developed in Chapter 3 can be used to answer exactly these types of questions that consider direct and indirect impacts to assess overall impacts.

7.4.2 Discussion

The model estimation results from the previous chapters and the policy analyses of this chapter helped answer the questions raised in the beginning of the chapter to a certain extent. However, more analyses are required from different contexts (of geography as well as the dimensions of activity-travel behavior) to accumulate the evidence and to come to a consensus. Further, the following questions remain unanswered.

(1) What is the total influence of residential self-selection due to all observed and unobserved factors in the context of different activity-travel choice dimensions?
(2) What are the direct and indirect (through the influence on auto and bicycle ownership) impacts of changes in the built environment on activity-travel behavior?
(3) What is the impact of the built environment on long-term residential location choices?
(4) How do the short-term and long-term impacts of built environment on the activity-travel patterns compare?
(5) How long does it take a built environment change to influence activity-travel behavior change?

The models developed in this dissertation can be used to answer the first four questions to a certain extent, but different methods need to be developed to address the fifth question.

CHAPTER 8

SUMMARY, CONCLUSIONS, AND FUTURE WORK

8.1 INTRODUCTION

Until about past three decades or so, the focus of transportation planning was to provide adequate transportation infrastructure supply to meet the mobility needs of the population. In such a supply-oriented, long-term, planning process, the role of travel demand models was to provide basic, aggregate-level, forecasts of the demand for travel. Over the past three decades, however, several signs of non-sustainable development, including suburban sprawl, auto dependence, traffic congestion, fossil fuel dependence, and environmental and global climate changes, have become increasingly apparent. As a result, the focus of transportation planning has expanded to include the objective of promoting sustainability.

Contemporary land-use and transportation planning, and policy efforts toward sustainability include, for example, integrated land-use and multimodal transportation planning, travel demand management, congestion pricing and high occupancy vehicle lane provision, mixed land-use and transit oriented development, and non-motorized travel promotion. These planning strategies and policy efforts are aimed at influencing individual-level daily activity and travel behavior to control overall travel demand, and to mitigate suburban sprawl, auto dependence, and traffic congestion. Consequently, the role of travel demand models has expanded to understanding the individual-level (as opposed to aggregate-level) behavioral responses to, and assessing the effectiveness of, these policies and corresponding infrastructure investments. This is evidenced in the evolution of the travel demand modeling field along three distinct directions: (a) Activity-based travel demand modeling, (b) Built environment and travel behavior modeling, and (c) Integrated land-use – transportation modeling. While the emergence of the activity-based approach may be attributed to the need to understand individual-level behavioral responses to travel demand management policies, the research on built environment and travel behavior has been driven by the need to accurately estimate the impact of built environment policies (such as smart growth strategies) on travel behavior. The integrated land use-travel demand modeling has emerged because of efforts to examine the impacts of land-use and transportation planning investments and policy actions on both land-use and transportation patterns within a unified behavioral and policy analysis framework. The three fields of research, however, have progressed in a rather disjoint fashion. That is, what is important in one field of research has not been considered in the other two research fields.

In view of this discussion, the overarching theme and goal of this dissertation was to contribute toward the research needs that are at the intersection of the three streams of research identified above and bring the three research areas together into a unified research framework. This is achieved by the simultaneous consideration of the following three aspects, each of which is of high importance in each direction of research identified above:

(1) Activity-based and tour-based approach to travel behavior analysis (by considering tour-based commute mode choice and activity participation and time-use behavior),

(2) Consideration of a comprehensive set of built environment attributes and corresponding residential self-selection effects, and

(3) Integrated modeling of long-term land-use related choices and medium- and short-term travel-related choices.

To this end, the specific substantive and methodological contributions of this dissertation are summarized in the subsequent section (Section 8.2). Section 8.3 summarizes the important empirical results, and Section 8.4 highlights the limitations of this research, and identifies directions for future research.

8.2 SUMMARY OF CONTRIBUTIONS

8.2.1 Integrated Models of Multidimensional Choice Processes

The dissertation formulated a series of integrated models of residential location choice and activity-travel behavior choices. These models are used to understand and disentangle the "causal" relationships and "spurious" associative correlations (or the endogeneity effects, or the self-selection effects, or the feedback effects) between long-term land-use related choices (such as the residential location choice), medium-term auto ownership and bicycle ownership choices, and short-term daily activity- and travel-related choices (such as the travel mode choice, and activity participation and time-use behavior) of households and individuals.

From a substantive viewpoint, this effort realizes the vision of modeling long-term, medium-term, and short-term choices of individuals and households as a lifestyle package (or a bundle of decisions) and represents a movement toward "truly" integrated land-use travel demand models. Although there is a rich body of literature devoted to the integrated modeling of location choices and travel choices, much of the work in the past has been limited by the analytical and computational complexity associated with estimating truly simultaneous equations systems that include a multitude of discrete and discrete-continuous choice variables. Further, to our knowledge, this is the first instance of the development of multi-dimensional choice model systems that can capture both causal and associative relationships between:

(a) Two unordered multinomial discrete choice variables (*i.e.,* the joint residential location choice and commute mode choice model developed in Chapter 2),

(b) Several unordered and ordered discrete choice variables (*i.e.,* the multi-dimensional choice model of residential location, auto ownership, bicycle ownership, and tour-based commute mode choice, developed in Chapter 3),

(c) An unordered multinomial choice variable and a set of multiple discrete-continuous choice variables (*i.e.,* the joint residential location choice and activity time-use behavior model, developed in Chapter 5).

In view of the methodological and behavioral inadequacies associated with the multidimensional multinomial (or deeply nested) logit modeling approaches adopted in the literature (see Chapter 2 for a critique on these methodologies), this dissertation uses the mixed multidimensional choice modeling approach to model multiple choice dimensions in an integrated fashion. The mixed multidimensional choice modeling approach is a powerful alterative to develop highly integrated multidimensional choice model systems. Such integrated model systems can accommodate the multitude of relationships along several choice dimensions in a simultaneous equations framework.

8.2.2 Self-Selection Effects and Unobserved Heterogeneity

The dissertation contributes toward accommodating and understanding self-selection effects in models of activity-travel behavior. The three types of self-selection effects addressed in this dissertation include: (a) Residential self-selection, (b) Endogeneity of auto ownership preferences, and (c) Endogeneity of bicycle ownership preferences. A notable contribution is in the context of accommodating, understanding and quantifying residential self-selection. Specifically, residential self-selection effects due to both observed and unobserved factors are accommodated in the contexts of household-level auto ownership and bicycle ownership decisions, individual-level trip-based and tour-based commute mode choice decisions, and individual-level activity participation and time-use behavior.

To our knowledge, from a substantive standpoint, this dissertation represents the first attempt to account for residential self-selection effects in: (a) "tour-based" mode choice models, and (b) "activity-based" models of activity participation and time-use behavior in multiple activities. From a methodological standpoint, this study represents the first instance of the incorporation of residential self-selection effects in multinomial logit and multiple discrete-continuous choice models of activity-travel behavior.

A notable contribution is the simultaneous incorporation of residential self-selection effects in multiple choice processes (auto ownership, bicycle ownership, and commute mode choice) while also accommodating for other unobserved effects such as endogeneity of auto and bicycle ownership, common unobserved heterogeneity between auto ownership and bicycle ownership, and unobserved heterogeneity in the impact of neighborhood and transportation system attributes (see Chapter 3). Such a simultaneous incorporation of all the afore-mentioned unobserved effects underscored the potential estimation bias (and the resulting distortion of policy implications) in all applications of the previous work in this area, which has, at the best, attempted to accommodate these unobserved effects in a piecemeal fashion (*i.e.,* one at a time, while ignoring all other effects).

Another important contribution in this area is the recognition and accommodation of heterogeneity in unobserved residential self-selection effects. Heterogeneity is incorporated along three different dimensions: (1) Heterogeneity across neighborhood characteristics, (2) Heterogeneity across choice alternatives, and (3) Heterogeneity across sociodemographic segments. Along the first dimension, as pioneered by Bhat and Guo (2007), the analyst can accommodate differential residential self-selection effects specific to each neighborhood attribute and the corresponding built environment policy. Along the second direction, the analyst can incorporate differential residential self-selection effects across choice alternatives. Along the second direction, to our knowledge, this is the first study to accommodate differential residential self-selection effects across various sociodemographic segments of the population (see Chapter 4).

Overall, the dissertation offers a highly detailed and intuitive dissection of the various unobserved effects, including residential self-selection, endogeneity, unobserved heterogeneity, and common unobserved heterogeneity in the context of residential location and activity-travel choice modeling. All of these effects were considered without loosing sight of the activity-based and integrated land-use travel demand modeling approaches.

8.2.3 The Multiple Discrete-Continuous Nested Extreme Value (MDCNEV) Model

Another significant methodological contribution of this dissertation is the econometric formulation of the multiple discrete-continuous nested extreme value model. The model can accommodate interalternative correlation patterns and flexible substitution patterns across mutually exclusive subsets (or nests) of alternatives in multiple discrete-continuous choice models. The salient feature of the model is that it offers closed form probability expressions for any (and all) multiple discrete-continuous choice patterns while accommodating the inter-alternative correlations. This development offers a powerful alternative to the currently used mixed modeling approach that requires simulation-based estimation to accommodate inter-alternative correlations.

In the context of the joint residential location and activity time-use choice model developed in Chapter 5, the MDCNEV model structure can be used to replace the MDCEV structure used to model the activity participation and time-use choice component. More specifically, such a model structure can be used to accommodate the interalternative correlations across the various activity types of the multiple discrete-continuous activity time-use model component of the joint model system developed in Chapter 5. This extension of the MDCEV model is analogous to the nested logit extension of the multinomial logit model and constitutes an important state-of-the-art development in the field of choice modeling.

8.3 EMPIRICAL FINDINGS AND POLICY IMPLICATIONS

The empirical analyses undertaken in this dissertation represent compelling applications of the modeling methodologies with real world choice data. All of the empirical analyses were undertaken in the context of the choices made by the residents of the San Francisco Bay Area. The 2000 Bay Area Activity-Travel Survey data along with several other sources of secondary data (including land-use/demographic coverage data, zone-to-zone network level-of-service data, GIS layers of highways, local roads, bicycle facilities, businesses, schools, parks and gardens, Census demographic data, and Public Use Micro Sample data) were used in the empirical analysis. These data were used to develop comprehensive databases of (1) residential location, auto and bicycle ownership, and activity-travel choices, (2) and an extensive set of natural, built, socio-demographic and travel environment attributes that were used in model estimations and policy analyses. This section summarizes the important empirical findings from, and policy implications of, all the empirical analyses undertaken in the dissertation in an itemized fashion. The first two points focus on the results and implications specific to integrated modeling of multiple choice processes, the next set of points (from 3 through 10) focus on residential self-selection effects, and the last set of points focuses on and built environment impacts.

8.3.1 Empirical Findings and Policy Implications: Integrated Modeling of Multiple Choice Dimensions

(1) The empirical results provide a compelling support to the development and use of integrated models of multidimensional choice processes. Almost all of the hypotheses regarding the relationships between the long-term residential location and medium/short-term activity-travel choices (such as auto and bicycle ownership, activity participation and time-use, and mode choice) were supported by the empirical results and policy simulations. The relationships are characterized by a mix of causality and spurious associations. The spurious associations include a series of behavioral aspects that are critical to simultaneously modeling multiple choice dimensions. These include endogeneity effects (where any one dimension is not exogenous to another, but is really endogenous to the system as a whole), residential self-selection effects due to observed and unobserved factors (where households locate in neighborhoods based on lifestyle and mobility preferences), unobserved heterogeneity (where households and individuals show significant variation in sensitivity to explanatory variables), and correlated error structures (where common unobserved factors significantly and simultaneously impact multiple choice dimensions).

(2) The findings of this dissertation have important implications, both from a modeling perspective and a policy perspective. From a modeling perspective, it is clear that travel demand model systems (such as activity-based microsimulation models) should increasingly move towards the incorporation of integrated models of location and travel choices similar to that developed in this dissertation. The incorporation of such model systems allows one to account for the various behavioral phenomena and aspects of interest outlined earlier. Unfortunately, most model systems in place today treat multi-dimensional choice processes as a series of sequential independent decisions or attempt to connect the various decisions through the use of deeply nested logit models that are quite restrictive with regard to their ability to reflect the various aspects of behavior highlighted in this paper. Given recent advances in analytical and

computational methods and the greater availability of comprehensive data sets that include both land use and travel choices, the time is ripe to begin implementing models of the type developed in this dissertation in future model systems. From a policy perspective, *not* using a model system to jointly model multiple choice dimensions in a unified framework could lead to erroneous estimates or forecasts of the impacts of changes in built environment attributes on location and travel choices of households and individuals. To provide credible policy information, it is important to capture all these effects simultaneously, rather than in a piecemeal fashion.

8.3.2 Empirical Findings and Policy Implications: Residential Self-Selection Effects

(3) The empirical estimations highlight the presence of significant amount of residential self-selection effects due to observed sociodemographic variables. Households self select their residential location based on demographic characteristics such as auto and bicycle ownership preferences, income, household structure (household size, presence of kids, *etc.*), race, housing type, and house ownership. Such self-selection effects due to observed sociodemographics have been found in all the models of activity-travel behavior, including auto ownership, bicycle ownership, mode choice, and activity time-use behavior considered in this dissertation. From a policy perspective, ignoring sociodemographic effects in activity-travel behavior models (as well as in auto and bicycle ownership models) can result in a "corrupt" assessment of the impact of built environment policies on individuals' activity-travel behavior.

(4) Significant magnitude of unobserved factors (such as predisposed lifestyle preferences, values, and attitudes such as environmental friendly nature, and physical fitness consciousness) contribute to residential self-selection effects. Ignoring such unobserved factors resulted, for example, in a spuriously estimated effect of the "bicycling facility density" variable on participation and the time spent in physically active out-of-home recreational (such as outdoor sports, exercising in gym, *etc.*). The extent of the influence of unobserved self-selection effects on the impact of this variable is large enough to underscore the importance of capturing such effects through a joint residential location and time-use model estimation. Other policy

variables affected by unobserved self-selection effects include the "total (auto)commute time of all commuters in the household". Unobserved self-selection effects distorted the impact of this variable on mode choice (see Chapters 2 and 3) as well as on bicycle ownership (see Chapter 3). Further investigations are needed to understand the type of unobserved factors contributing to self-selection effects in the context of "bicycling facility density" and "total (auto)commute time of all commuters in the household". In addition, it would be useful to incorporate attitudes, lifestyle preferences, and subjective perceptions on the built environment in activity-travel behavior modeling to better understand and incorporate residential self-selection effects.

(5) Based on the evidence from all empirical estimations, it appears that the self-selection effects due to observed socio-demographic factors may be larger (in the number of factors contributing to, and the extent of, self-selection) than the self-selection effects due to unobserved factors.

(6) Ignoring residential self-selection effects may not always lead to an overestimation of the impact of built environment impacts. In a particular empirical context (in Chapter 4), ignoring self-selection due to sociodemographic attributes such as number of children, housing unit type, and house ownership resulted in an underestimation of the impact of bicycling-friendly neighborhoods on bicycle ownership.

(7) Self-selection effects are specific to each built environment attribute (and the corresponding policy) and each choice alternative under consideration. Thus, it is important to incorporate self-selection effects separately for each built environment attribute and choice alternative under consideration, rather than estimating a common self-selection effect to all built environment attributes and all choice alternatives. The nested logit modeling approach, for example, muddles all self-selection effects into one log-sum variable. The mixed modeling approach used in this dissertation clearly disentangles the different self-selection effects as well as other unobserved effects.

(8) No significant heterogeneity was found in the extent of unobserved residential self-selection effects in a particular empirical context (*i.e.,* in the context of the impact of

bicycle-friendly neighborhoods on bicycle ownership[43]). This is perhaps because of the absence of any unobserved self-selection effects to begin with. However, it is important to, at the least, test for the presence of heterogeneity in self-selection across sociodemographic segments in different (geographical and activity-travel choice dimension) empirical contexts.

(9) Even if there are no unobserved residential self-selection effects, it behooves the analyst to estimate an independent residential choice model so that any observed demographic factors influencing the sensitivity to built environment attributes in residential choice can be considered in the activity-travel choice models. In this way, one can reduce the possibility of "corrupt" inferences regarding the impact of built environment attributes on activity-travel choices. Of course, another reason to model both residential choice and activity-travel choices, even in the absence of common unobserved factors, is the fact that built environment attributes impact both these choices. Thus, policy decisions regarding changes in built environment characteristics have to be evaluated in the context of spatial relocations as well as activity-travel behavior shifts to obtain a comprehensive picture of the changes due to built environment policies.

(10) Residential self-selection effects due to unobserved factors result in common unobserved heterogeneity across residential location choice model and the activity/travel behavior model of interest. In addition to residential self-selection effects, other types of common unobserved heterogeneity may exist, including the endogeneity of auto ownership and bicycle ownership choices in activity-travel behavior models. This is evidenced by the significant endogeneity of auto ownership and bicycle ownership in the context of commute tour mode choice, and the common unobserved heterogeneity between auto ownership and bicycle ownership propensity equations (in Chapter 3).

[43] In this context, it was necessary to go beyond simple log-likelihood test results to conclude the absence of demographic heterogeneity in unobserved residential self-selection. While the log-likelihood improvement suggested a presence of demographic heterogeneity, policy simulations indicated otherwise. Perhaps, log-likelihood improvement should not be the sole criterion to test alternative hypotheses.

8.3.3 Empirical Findings and Policy Implications: Built Environment Impacts

(11) Built environment attributes are not truly exogenous in travel-related choice decisions. Households and individuals locate (or self-select) themselves in built environments that are consistent with their lifestyle preferences, attitudes, and values. In other words, households and individuals make residential location, auto ownership, bicycle ownership, and activity-travel decisions jointly as part of an overall lifestyle package. Nevertheless, the findings in this dissertation suggest that modifying the built (or activity-travel) environment can bring about changes in activity and travel related behavior. This is evidenced by the statistical significance of these attributes in the activity-travel models even after controlling for residential self-selection and other unobserved effects. As a quick reference, a summary of the built environment effects gathered from all the empirical estimations undertaken for this dissertation is presented in Table 8.1.

(12) Although the estimation results show statistically significant impacts of the built environment attributes on various activity-travel choices, the policy simulations indicate mixed results on the extent of the built environment impacts. The direct influence of the built environment changes on time-use behavior appears to be very limited. On the other hand, such policies appear to have a discernible direct impact on bicycle ownership.

In the context of time-use behavior, given that changes in bicycle ownership appear to have large impacts on time-use behavior and that built environment policies appear to have a discernible direct impact on bicycle ownership, the built environment polices may have a discernible indirect impact on time-use behavior. Undertaking a comprehensive analysis of the direct and indirect (through bicycle ownership and vehicle ownership) impacts of the built environment on time-use behavior may shed more light on this issue.

Table 8.1 Summary of the Impacts of Activity-Travel Environment on Residential Location and Activity-Travel Choices

Variables	Residential Location choice	Auto Ownership	Bicycle Ownership	Mode Choice	Activity Time-Use
Residence-end zonal size and density					
Logarithm of number of households in zone	+				
Household/population density	- for HHs w/ kids, seniors, high income	70% -			
Employment density	- for HHs w/ kids, high income, Caucasians	-		64% + on walk/bike	+ on internet use, IHSocial/relaxing, recreation
Residence-end zonal land-use structure					
Fraction of residential land area	-				
Fraction of commercial land area	-				
Land-use mix	-			+ on transit	
Residence-end demographic composition					
Absolute difference between zonal median income and household income	+				
Absolute difference between zonal average household size and household size	+				
Zonal-median housing value	+				
Fraction of a particular racial population	+ for HH of that race				
Residence-end zonal activity opportunities					
Schools	+				
Physically active recreation centers	+		+		+ on out-of-home physically active
Natural recreational centers	+		+		
Physically inactive recreation centers					+ on out-of-home physically
Residence-end transportation measures					
Bicycling facility density	+		+	+ on walk/bike	+ on physically active recreation
Highway density	-	+			

Table 8.1 (Continued) Summary of the Impacts of Activity-Travel Environment on Residential Location and Activity-Travel Choices

Variables	Residential Location choice	Auto Ownership	Bicycle Ownership	Mode Choice	Activity Time-Use
Residence-end transportation measures					
Street block density	-	99.8% -	+	+ on walk	
No. of zones accessible in 30 min by transit	+	67% -		+ on transit	
No. of zones within 6 miles on bike route		-	+		
Work-end zonal transportation measures					
Bicycle facility density				+ on bike	
Average bicycling facility density of HH commuter employment locations		-			
Average street block density of the HH commuter employment locations		93% -			
No. of zones accessible in 30 min by transit				+ on transit	
Work-end zonal density measures					
Household density				+ on transit/walk	
Employment density				+ on transit/walk	
Household-level commute LOS variables					
Total commute time (by auto) of all commuters in HH	- for large share of population	+	- for large share of population	+ for large share of population	
Total commute cost by (auto) of all commutes in HH	- for low income HHs	- for low income			
No. of HH commuters w/ home and work zones connected by transit in 30 minutes	+	-			
Number of HH commuters w/ home and work zones connected by bike in 6miles	+	-			
Individual-level commute LOS					
Travel times and costs				- on all modes	

Index: - Negative impact, + Positive impact

Note: The entry “70% -” indicates that the coefficient varies in the population (due to unobserved heterogeneity) with 70% of the impact being negative. Similarly, the entry “- for large share of population” indicates that the coefficient (or the impact) varies in the population (due to unobserved heterogeneity) with more than 99% of the impact being negative.

This question remains to be answered in the mode choice and auto ownership contexts as well[44]. Further, extensive policy analyses should be carried out with different built environment attributes to determine which neighborhood attribute(s) matter the most for each dimension of activity-travel behavior.

(13) Several built environment attributes together may have a considerably discernible influence on most aspects of activity-travel behavior. That is, built environment policies should be implemented in combination (for example, increase the bicycling facilities in the neighborhood, and also increase the connectivity of the bicycle route network to other neighborhoods), rather than in isolation, for achieving synergistic results. On similar lines, land-use transportation policies need to modify the land-use and transportation characteristics of both work and residence places in order to achieve desirable shifts in travel behavior.[45] Overall, a comprehensive and synergistic package of policies and programs is needed. For example, some empirical results suggested that policies and educational campaigns aimed at reducing motorized vehicle ownership and increasing bicycle ownership, when combined with better provision of bicycling facilities, not only can lead to traffic congestion alleviation, but can also play an important role in improving public health through increased investment of time in physically active pursuits. These findings emphasize the transportation-public connection as well as underscore the importance of creative combinations of policies for synergistic impacts. Further research needs to investigate the combinations of policies and the means of implementation that would best serve the purpose.

(14) Land-use density variables act as proxies to transportation network and modal accessibility attributes. Exclusion of the latter can result in an over assessment of the impact of the former. Broadly speaking, neglect of a set of built environment attributes would result in an over assessment of the impact of the built environment

[44] The integrated model developed in Chapter 3 can be used to answer exactly these types of questions that consider direct and indirect impacts to assess overall impacts.

[45] The neglect of work-end transportation network and land-use density attributes resulted in an over assessment of the impact of residence-end land-use and transportation system variables on several travel-related choices such as auto ownership and mode choice.

attributes included in the model. Thus, it is important to consider a comprehensive set of built environment factors.

(15) Despite all the above-discussed findings in the context of built environment, sociodemographics appear to play a far more dominant role than does the built environment in shaping auto and bicycle ownership and activity and travel behavior. This evidence, combined with the finding that residential self-selection effects due to sociodemographic attributes are prevalent, underscores the importance of sociodemographics in models of activity-travel behavior. It is important to control for the impacts of sociodemographic factors while assessing the impacts of policy sensitive variables. Broadly speaking, it is important to first (and foremost) carefully identify and specify all sociodemographic and observable factors in models of activity-travel behavior before moving on to sophisticated model specifications to capture unobserved effects. However, one must recognize that it may not always be possible to observe and directly specify all factors in a model of complex human choice behavior. For example, estimation results in the dissertation show significant unobserved heterogeneity in the impact of "total (auto)commute time of all commuters in the household", among other attributes. In fact, the estimated unobserved variations in the impact of this commute time variable are large enough to warrant an investigation of the sources of this variation.

8.4 LIMITATIONS AND FUTURE WORK

Admittedly, this dissertation has several limitations. These limitations form a basis for several research extensions. The limitations and future research directions are identified in this section.

8.4.1 Attitudes and Perceptions

Individuals' attitudes and preferences pertaining to their activity-travel choices, and their perceptions of neighborhood built environment attributes are not considered in this dissertation. Several earlier studies (Kitamura *et al.,* 1997; Schwanen and Mokhtarian 2005; and Cao 2006) have shown that attitudes and subjective perceptions can play an important role in travel behavior choices. However, most activity-travel surveys do not collect such information. Further, it is difficult to elicit and quantify such subjective perceptions.

The methods presented in this dissertation are general and are readily applicable for use with attitude and perception data, should such data become available. It is an important avenue to use these methodologies with datasets that have attitudinal information. It is possible that controlling for attitudinal variables can serve to lessen the impact of self-selection effects (Schwanen and Mokhtarian 2005, Kitamura *et al.,* 1997, Handy *et al.,* 2005, and Cao *et al.,* 2006a). Alternatively, it is also possible that additional residential self-selection effects will be found due to common (across residential location and activity-travel behavior choices) unmeasured variables such as attitudes toward walking and/or driving and the subjective perceptions of the neighborhood activity-travel environment.

8.4.2 Longitudinal Data and Longitudinal Studies

This dissertation utilizes cross-sectional data for the analysis of long-term, medium-term, and shot-term choices. The data does not provide information on when people moved and when they acquired their vehicles and bicycles, and when they made their mode choice and activity time-use choice preferences. To the extent that the shorter-term choices may have been made before the long-term choices, one can not clearly decipher the impact of the current neighborhood type on bicycle ownership. To make things worse, even if the data was available on when the choices were made, not all the choices would have been made at a single point in time when the household was at a particular neighborhood. However, when one views all these choices together as a “package” or a “bundle” of choices made by households over a larger timeframe, the current choices observed in the data can be treated as a result of households’ overall lifestyle preferences at that point in time. The joint modeling framework captures the spirit of modeling these choices as a bundle of choices made as part of a lifestyle package. Nevertheless, longitudinal studies with longitudinal data will enhance our understanding of the causal and spurious relationships between the several choice dimensions under consideration. Longitudinal studies have a stronger "time precedence" (*i.e.,* cause preceding the effect; see Mokhtarian and Cao, 2008) and can incorporate dynamics (temporal variations in choice behavior) and state dependence (influence of past choices on current choices) to better explain the link between built environment and travel behavior.

Longitudinal studies require longitudinal data with information on the choices made (along with the neighborhood activity-travel environmental characteristics) over multiple time periods. Cao (2006) recommends that the data on attitudinal preferences and subjective perceptions of the built environment should also be collected over different time periods. Such data collection efforts (and, in fact, the research design itself) can be time consuming and expensive, however, with potentially high benefits of better understanding of the intricate relationships between longer-term and shorter-term choices and the built environment and activity-travel behavior (including residential self-selection effects). Such studies can be used not only to answer whether and how much does built environment impact activity-travel behavior, but also to answer important questions such as how long does it take to change individuals' and households' activity-travel behavior and preferences.

Several studies in the past have attempted to mimic the longitudinal studies using before- and after- residential move data (Krizek, 2003), and quasi-longitudinal data (Cao *et al.,* 2007), and a more recent research project in Australia labeled as the "RESIDential Environments project" (or the RESIDE project) aims a "truly" longitudinal study. It would be an interesting avenue of further research to enhance the modes developed in this dissertation to be used with such rich sources of longitudinal data.

Finally, on a tangential note, this dissertation and most of the earlier studies rely upon the statistical association between revealed choices as a means to assess the cause-and-effect relationship between the corresponding decisions. While such revealed choice data provides information on the observed choice outcomes of decision-makers, it does not provide insights into the underlying behavioral processes that lead to those decisions (Ye *et al.,* 2007). In order to clearly understand the underlying the behavioral mechanism, detailed data on behavioral processes and decision sequences is needed. In this context, the longitudinal research designs may benefit from collecting data on the behavioral processes (such as eliciting information on what prompts the change in residential locations, auto ownership, etc).

8.4.3 Spatial Representation and Correlation

This dissertation employs a traffic analysis zone (TAZ)-based spatial representation for residential location choice modeling as well as for measuring the attributes of the built environment. The choice of traffic analysis zones in such models has been the subject of debate for many years now; especially so in the case of location choice modeling. The research in this dissertation can be extended and enhanced by undertaking the analysis at a more disaggregate spatial level of analysis than traffic analysis zones (though such a move also promises to raise some very important computational challenges). In another direction, the neighborhood type definition used in Chapter 4 can be employed to disaggregate spatial units without encountering the computational challenges. In general, further research should explore the sensitivity of model results to using alternative aggregations of spatial units for representing built environment and location choice processes. An important conceptual and modeling consideration in such studies relates to: (1) quantifying the perception of geographical space, (2) definition of residential location alternatives, and (3) the consideration of residential location alternatives by individuals.

In addition to exploring alternative spatial configurations, it would be useful to accommodate spatial correlations (*i.e.,* correlations among the residential location spatial choice alternatives) along with residential self-selection effects.

8.4.4 Endogeneity of Work Location and Other Choices

The residential location, auto ownership, bicycle ownership, mode choice, and activity-travel behavior analyses undertaken in this dissertation assume the individual employment locations as exogenous (*i.e.,* fixed and given). However, the residential location and work location decisions may be joint in nature (see Siegel, 1975, Simpson, 1980, Waddell, 1993a, and Merriman, 1994), and hence the commute travel times and travel costs may be endogenous to each of the choice dimensions considered in the dissertation. It is an important future research avenue to incorporate the joint nature of residential and workplace decisions in all the multidimensional models estimated in this dissertation.

On similar lines, the housing cost variable (see Guevara and Ben-Akiva, 2006, and de Palma *et al.,* 2005) and other residential choices such as housing unit type and house ownership (*i.e.,* own or rent) decisions may be endogenous to residential location choice model. Future research efforts should consider the impact of these endogeneity issues while developing the multidimensional models of residential location and activity-travel behavior choices.

8.4.5 Larger Integrated Model Systems

The multidimensional model systems presented in this dissertation can be extended to include additional choice dimensions and endogenous variables. However, the inclusion of additional endogenous variables will add computational burden. Nevertheless, future research efforts should focus on further expanding the joint model systems presented in this dissertation on developing multi-dimensional choice models that incorporate other variables of interest. For example, the joint residential location choice and activity participation and time-use behavior model can be extended to include the endogeneity of auto ownership and bicycle ownership decisions. Such a model system can be further extended to include travel mode choice, time-of-day choice, and travel destination choice of non-work activities. Development of such highly integrated model systems will serve the following purposes: (1) Integration of activity-based analysis (*i.e.,* activity participation and time-use analysis) with travel behavior analysis (mode choice, time-of-day, and location analysis), and (2) The accommodation of residential self-selection effects in such dimensions as destination choice.

8.4.6 Policy Scenario Analysis and Transferability of Results

The models developed in this dissertation can be used for evaluating a variety of policy questions (see Chapter 7). An important and immediate next step to this dissertation is to extend the work of Chapter 7 and undertake extensive policy analysis exercises that can throw more light on the influence of the built environment on various choice dimensions including residential location choices and travel-related choices. More specifically, it would be useful to undertake policy analyses to answer the questions identified at the end of Chapter 7. Another important avenue of research is to test if the empirical findings of the current model applications to the San Francisco Bay Area can be generalized to other geographical contexts.

8.4.7 Residential Self-Selection and Latent Residential Demand

The residential self-selection analyses in this dissertation assume that there is no mismatch between the residential preferences of households and their current residential locations. However, there may be an undersupply (in comparison with the demand) of neighborhoods that promote non-auto travel. In such cases, building neighborhoods that promote alternatives to auto use would lead to a reduction in auto use in the population even if the only process at work is residential self-selection (Bhat and Guo, 2007). It would be interesting to consider the issues of demand-supply mismatch and residential self-selection, in order to quantify the extent to which the latent residential self-selection (or the unmet residential demand) can be accommodated by providing the adequate supply of appropriate neighborhoods (See Crane 2000, and Schwanen and Mokhtarian, 2005b).

APPENDIX A

For $r_s = 1$, $X_{rs} = \{1\}$.

For $r_s = 2$, $X_{rs} = \left\{ \frac{(q_s - 1)(1-\theta_s)}{\theta_s} + \frac{(q_s - 2)(1-\theta_s)}{\theta_s} + \ldots + \frac{2(1-\theta_s)}{\theta_s} + \frac{1(1-\theta_s)}{\theta_s} \right\}$.

For $r_s = 3, 4, \ldots, q_s$, X_{rs} is a matrix of size $\begin{bmatrix} q_s - 2 \\ r - 2 \end{bmatrix}$ which is formed as described below:

Consider the following row matrices A_{q_s} and A_{rs} (with the elements arranged in the descending order, and of size $q_s - 1$ and $r_s - 2$, respectively):

$$A_{q_s} = \left\{ \frac{(q_s - 1)(1-\theta_s)}{\theta_s}, \frac{(q_s - 2)(1-\theta_s)}{\theta_s}, \frac{(q_s - 3)(1-\theta_s)}{\theta_s}, \ldots, \frac{3(1-\theta_s)}{\theta_s}, \frac{2(1-\theta_s)}{\theta_s}, \frac{1(1-\theta_s)}{\theta_s} \right\}$$

$$A_{rs} = \left\{ r_s - 2, r_s - 3, r_s - 4, \ldots, 3, 2, 1 \right\}.$$

Choose any $r_s - 2$ elements (other than the last element, $\frac{1-\theta_s}{\theta_s}$) of the matrix A_{q_s} and arrange them in the descending order into another matrix A_{iq_s}. Note that we can form $\begin{bmatrix} q_s - 2 \\ r_s - 2 \end{bmatrix}$ number of such matrices. Subsequently, form another matrix $A_{irq_s} = A_{iq_s} + A_{rs}$.

Of the remaining elements in the A_{q_s} matrix, discard the elements that are larger than or equal to the smallest element of the A_{iq_s} matrix, and store the remaining elements into another matrix labeled B_{irq_s}. Now, an element of X_{rs} (*i.e.*, x_{irq_s}) is formed by performing the following operation: $x_{irq_s} = \text{Product}(A_{irq_s}) \times \text{Sum}(B_{irq_s})$; that is, by multiplying the product of all elements of the matrix A_{irq_s} with the sum of all elements of the matrix B_{irq_s}.

Note that the number of such elements of the matrix X_{rs} is equal to $\begin{bmatrix} q_s - 2 \\ r_s - 2 \end{bmatrix}$.

APPENDIX B

From Equation (6.13), the $q_{\ell}{}^{th}$ order partial derivative for the ℓ^{th} nest is:

$$\frac{\partial^{q_\ell} F}{\partial\varepsilon_{1\ell}\ldots\partial\varepsilon_{q_\ell}} = F\left(\prod_{\substack{i\in\ell^{\text{th}}\text{nest, and}\\ i\in\{\text{chosen alts}\}}} e^{-\frac{V_1-V_i}{\theta_\ell}}\right)\sum_{r_\ell=1}^{q_\ell}\left\{e^{-\varepsilon_1(q_\ell-r_\ell+1)}\left(\sum_{i\in\ell^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_\ell}}\right)^{(q_\ell-r_\ell+1)\theta-q_\ell} sum(X_{r_\ell})\right\} \quad \text{(B1)}$$

Equation (B1) can be used to expand the M$^{\text{th}}$ order partial derivative of Equation (6.12) as follows:

$$\frac{\partial^M F}{\partial\varepsilon_1\partial\varepsilon_2\ldots\partial\varepsilon_M} =$$

$$= F^{(1-S_M)}\prod_{\ell=1}^{S_M}\left[F\left(\prod_{\substack{i\in\ell^{\text{th}}\text{nest, and}\\ i\in\{\text{chosen alts}\}}} e^{-\frac{V_1-V_i}{\theta_\ell}}\right)\sum_{r_\ell=1}^{q_\ell}\left\{e^{-\varepsilon_1(q_\ell-r_\ell+1)}\left(\sum_{i\in\ell^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_\ell}}\right)^{(q_\ell-r_\ell+1)\theta-q_\ell} sum(X_{r\ell})\right\}\right]$$

$$= F\prod_{\ell=1}^{S_M}\left[\left(\prod_{\substack{i\in\ell^{\text{th}}\text{nest, and}\\ i\in\{\text{chosen alts}\}}} e^{-\frac{V_1-V_i}{\theta_\ell}}\right)\sum_{r_\ell=1}^{q_\ell}\left\{e^{-\varepsilon_1(q_\ell-r_\ell+1)}\left(\sum_{i\in\ell^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_\ell}}\right)^{(q_\ell-r_\ell+1)\theta-q_\ell} sum(X_{r_\ell})\right\}\right]$$

$$= \exp\left(-e^{-\varepsilon_1}f\right)\prod_{\ell=1}^{S_M}\left[\left(\prod_{\substack{i\in\ell^{\text{th}}\text{nest, and}\\ i\in\{\text{chosen alts}\}}} e^{-\frac{V_1-V_i}{\theta_\ell}}\right)\sum_{r_\ell=1}^{q_\ell}\left\{e^{-\varepsilon_1(q_\ell-r_\ell+1)}\left(\sum_{i\in\ell^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_\ell}}\right)^{(q_\ell-r_\ell+1)\theta-q_\ell} sum(X_{r_\ell})\right\}\right] \quad \text{(B2)}$$

The above expression for the M$^{\text{th}}$ order partial derivative can be substituted into the probability expression of Equation (6.11) as follows:

$$P(t_1^*, t_2^*, \ldots, t_M^*, 0, \ldots, 0) = |J| \times$$

$$\int_{\varepsilon_1=-\infty}^{+\infty} \exp\left(-e^{-\varepsilon_1}f\right)\prod_{\ell=1}^{S_M}\left[\left(\prod_{\substack{i\in\ell^{\text{th}}\text{nest, and}\\ i\in\{\text{chosen alts}\}}} e^{-\frac{V_1-V_i}{\theta_\ell}}\right)\sum_{r_\ell=1}^{q_\ell}\left\{e^{-\varepsilon_1(q_\ell-r_\ell+1)}\left(\sum_{i\in\ell^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_\ell}}\right)^{(q_\ell-r_\ell+1)\theta-q_\ell} sum(X_{r_\ell})\right\}\right] d\varepsilon_1 \quad \text{(B3)}$$

The probability expression in Equation (B3) can be rewritten as follows:

$$P(t_1^*, t_2^*, \ldots, t_M^*, 0, \ldots, 0)$$

$$= |J| \int_{\varepsilon_1=-\infty}^{+\infty} \exp\left(-e^{-\varepsilon_1}f\right)\left(\prod_{i\in\{\text{chosen alts}\}} e^{-\frac{V_1-V_i}{\theta_i}}\right)\prod_{\ell=1}^{S_M}\left[\sum_{r_\ell=1}^{q_\ell}\left\{e^{-\varepsilon_1(q_\ell-r_\ell+1)}\left(\sum_{i\in\ell^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_\ell}}\right)^{(q_\ell-r_\ell+1)\theta_\ell-q_\ell} sum(X_{r_\ell})\right\}\right] d\varepsilon_1$$

$$=|J|\left(\prod_{i\in\{\text{chosen alts}\}} e^{-\frac{V_1-V_i}{\theta_i}}\right)\int_{\varepsilon_1=-\infty}^{+\infty} \exp\left(-e^{-\varepsilon_1} f\right)\prod_{s=1}^{S_M}\left[\sum_{r_s=1}^{q_s}\left\{e^{-\varepsilon_1(q_s-r_s+1)}\left(\sum_{i\in s^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_s}}\right)^{(q_s-r_s+1)\theta_s-q_s} sum(X_{r_s})\right\}\right] d\varepsilon_1,$$

(B4)

The reader will observe that the expression in Equation (B4) involves a product over all nests (*i.e.,* over $s = 1,2,\ldots,S_M$) of summations over all alternatives in a nest (*i.e.,* over $r_s = 1,2,\ldots,q_s$). This product of summations can be expressed as a summation of products and the Equation (B4) can be rewritten as follows:

$$P(t_1^*, t_2^*, \ldots,\ t_M^*, 0, \ldots, 0) = |J| \times$$

$$\times\left(\prod_{i\in\{\text{chosen alts}\}} e^{-\frac{V_1-V_i}{\theta_i}}\right)\int_{\varepsilon_1=-\infty}^{+\infty} \exp\left(-e^{-\varepsilon_1} f\right)\left[\sum_{r_1=1}^{q_1}\cdots\sum_{r_{S_M}=1}^{q_{S_M}}\left\{\left(\prod_{s=1}^{S_M} e^{-\varepsilon_1(q_s-r_s+1)}\right)\prod_{s=1}^{S_M}\left(\sum_{i=1}^{q_s} e^{-\frac{(V_1-V_i)}{\theta_s}}\right)^{(q_s-r_s+1)\theta_s-q_s}\prod_{s=1}^{S_M} sum(X_{r_s})\right\}\right] d\varepsilon_1$$

(B5)

In the above expression, all the terms containing ε_1 are moved to the right corner and the other terms are moved out of the integral as follows:

$$P(t_1^*, t_2^*, \ldots,\ t_M^*, 0, \ldots, 0) = |J| \times$$

$$\times\left(\prod_{i\in\{\text{chosen alts}\}} e^{-\frac{V_1-V_i}{\theta_i}}\right)\int_{\varepsilon_1=-\infty}^{+\infty}\sum_{r_1=1}^{q_1}\cdots\sum_{r_{S_M}=1}^{q_{S_M}}\left\{\prod_{s=1}^{S_M}\left(\sum_{i\in s^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_s}}\right)^{(q_s-r_s+1)\theta_s-q_s}\prod_{s=1}^{S_M} sum(X_{r_s})\left(\prod_{s=1}^{S_M} e^{-\varepsilon_1(q_s-r_s+1)}\right)\exp\left(-e^{-\varepsilon_1} f\right) d\varepsilon_1\right\}$$

$$=|J|\times$$

$$\times\left(\prod_{i\in\{\text{chosen alts}\}} e^{-\frac{V_1-V_i}{\theta_i}}\right)\sum_{r_1=1}^{q_1}\cdots\sum_{r_{S_M}=1}^{q_{S_M}}\left\{\prod_{s=1}^{S_M}\left(\sum_{i\in s^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_s}}\right)^{(q_s-r_s+1)\theta_s-q_s}\prod_{s=1}^{S_M} sum(X_{r_s})\int_{\varepsilon_1=-\infty}^{+\infty}\left(\prod_{s=1}^{S_M} e^{-\varepsilon_1(q_s-r_s+1)}\right)\exp\left(-e^{-\varepsilon_1} f\right) d\varepsilon_1\right\}$$

(B6)

Finally, the product of exponentials in the integral above is expressed as a single exponential, and the probability expression is as below:

$$P(t_1^*, t_2^*, \ldots,\ t_M^*, 0, \ldots, 0) = |J| \times$$

$$\times\left(\prod_{i\in\{\text{chosen alts}\}} e^{-\frac{V_1-V_i}{\theta_i}}\right)\sum_{r_1=1}^{q_1}\cdots\sum_{r_{S_M}=1}^{q_{S_M}}\left\{\prod_{s=1}^{S_M}\left(\sum_{i\in s^{\text{th}}\text{nest}} e^{-\frac{(V_1-V_i)}{\theta_s}}\right)^{(q_s-r_s+1)\theta_s-q_s}\prod_{s=1}^{S_M} sum(X_{r_s})\int_{\varepsilon_1=-\infty}^{+\infty} e^{-\varepsilon_1\sum_{s=1}^{S_M}(q_s-r_s+1)}\exp\left(-e^{-\varepsilon_1} f\right) d\varepsilon_1\right\}$$

(B7)

This expression is the same as the consumption probability expression given in Equation (14) of the text. The expression includes an integral that has a closed-from solution (as proved in the Appendix C below).

APPENDIX C

Consider the integral $I = \int\limits_{\varepsilon_1=-\infty}^{+\infty} e^{-\varepsilon_1 \sum\limits_{\text{ð}=1}^{S_M}(q_{\text{ð}} - r_{\text{ð}} + 1)} \exp\left(-e^{-\varepsilon_1} f\right) d\varepsilon_1$ from Equations (6.14) or (B7).

The above integral is in the following form:

$$I = \int\limits_{\varepsilon_1=-\infty}^{+\infty} \left(e^{-\varepsilon_1}\right)^n \exp\left(-e^{-\varepsilon_1} f\right) d\varepsilon_1, \quad \text{where } n = \sum_{\text{ð}=1}^{S_M}(q_{\text{ð}} - r_{\text{ð}} + 1) \tag{C1}$$

Now, let $e^{-\varepsilon_1} = x$, then $d\varepsilon_1 = {-dx}/{e^{-\varepsilon_1}}$.

Using the above substitution, the integral can be expressed as:

$$\begin{aligned} I &= -\int\limits_{x=+\infty}^{0} x^{n-1} \exp\left(-xf\right) dx \\ &= \int\limits_{x=+\infty}^{0} \frac{x^{n-1}}{f} d\left(e^{-xf}\right) \end{aligned} \tag{C2}$$

Applying integration by parts, the above integral can be simplified as follows:

$$\begin{aligned} I &= \left| \frac{x^{n-1}}{f} e^{-xf} - \int \frac{(n-1)x^{n-2}}{f} e^{-xf} dx \right|_{x=+\infty}^{0} \\ &= \left| \frac{x^{n-1}}{f} e^{-xf} + \left(\frac{(n-1)x^{n-2}}{f^2} e^{-xf} - \int \frac{(n-1)(n-2)x^{n-3}}{f^2} e^{-xf} dx \right) \right|_{x=+\infty}^{0} \\ &= \left| \frac{x^{n-1}}{f} e^{-xf} + \frac{(n-1)x^{n-2}}{f^2} e^{-xf} + \left(\frac{(n-1)(n-2)x^{n-3}}{f^3} e^{-xf} - \int \frac{(n-1)(n-2)(n-3)x^{n-4}}{f^3} e^{-xf} dx \right) \right|_{x=+\infty}^{0} \\ &\vdots \\ &= \left| e^{-xf} \left(\frac{x^{n-1}}{f} + \frac{(n-1)x^{n-2}}{f^2} + \frac{(n-1)(n-2)x^{n-3}}{f^3} + \ldots + \frac{(n-1)(n-2)\ldots 2\, x^1}{f^{n-1}} + \frac{(n-1)!\, x^0}{f^n} \right) \right|_{x=+\infty}^{0} \\ &= \frac{(n-1)!}{f^n} \\ \text{or} \quad I &= \frac{\left(\sum\limits_{\text{ð}=1}^{S_M}(q_{\text{ð}} - r_{\text{ð}} + 1) - 1 \right)!}{f^{\sum\limits_{\text{ð}=1}^{S_M}(q_{\text{ð}} - r_{\text{ð}} + 1)}} \end{aligned} \tag{C3}$$

REFERENCES

Abraham, J.E., and J.D. Hunt (1997) Specification and Estimation of Nested Logit Model of Home, Workplaces, and Commuter Mode Choices by Multiple-Worker Households. Transportation Research Record 1606, pp. 17-24.

Amemiya, T. (1974) Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica*, 42, pp. 999-1012.

American Planning Association (APA). (2002) *Planning for Smart Growth: 2002 State of the States*. Available at http://www.planning.org/growingsmart/states2002.htm.

Anas, A. (1981) The Estimation of Multinomial Logit Models of Joint Location and Travel Mode Choice from Aggregated Data. Journal of Regional Science 21(2), pp. 223-242.

Anas, A. (1995) Capitalization of Urban Travel Improvements into Residential and Commercial Real Estate: Simulations with a Unified Model of Housing, Travel Mode and Shopping Choices. Journal of Regional Science 35(3), pp. 351-375.

Anas, A., and L.S. Duann (1985) Dynamic Forecasting of Travel Demand, Residential Location, and Land Development: Policy Simulations with the Chicago Area Transportation/Land-Use Analysis System. Papers in Regional Science Association 56, pp. 38-58.

Axhausen, K. and Gärling, T. (1992) Activity-based approaches to travel analysis: conceptual frameworks, models and research problems. Transport Reviews, 12, 324-341.

Badoe, D. A. and Miller, E. J. (2000). Transportation–land-use interaction: empirical findings in North America, and their implications for modeling. *Transportation Research D*, Vol 5, issue 4, pp. 235-263.

Bagley, M. N. and P.L. Mokhtarian (2002) The Impact of Residential Neighborhood Type on Travel Behavior: A Structural Equations Modeling Approach. Annals of Regional Science, 36(2), 2002, pp. 279-297.

Baltes M.R. (1996) Factors Influencing Nondiscretionary Work Trips by Bicycle Determined from 1990 U.S. Census Metropolitan Statistical Area Data, Transportation Research Record 1538, pp. 96-101.

Beck, M.J.H. and L. H. Immers (1994) Bicycle Ownership and Use in Amsterdam. Transportation Research Record 1441, pp. 141-148.

Ben-Akiva, M., and A. dePalma (1986) Analysis of a Dynamic Residential Location Choice Model with Transactions Costs. Journal of Regional Science 26, pp. 321-341.

Ben-Akiva, M., and J.L. Bowman (1995) Activity-Based Disaggregate Travel Demand Model System with Daily Activity Schedules, Massachusetts Institute of Technology.

Ben-Akiva, M., and J.L. Bowman (1998) Integration of an Activity-based Model System and a Residential Location Model. Urban Studies 35(7), pp. 1131-1153.

Ben-Akiva, M., and S.R. Lerman (1974) Some Estimation Results of a Simultaneous Model of Auto Ownership and Mode Choice to Work. Transportation 3(4), pp. 357-376.

Ben-Akiva, M., and S.R. Lerman, and D. Damm (1980) Understanding, Prediction, and Evaluation of Transportation Related Consumer Behavior, MIT Center for Transportation Studies.

Berry, S., Levinsohn, J. and A. Pakes (1995) Automobile Prices in Market Equilibrium. *Econometrica*, Vol 63, pp. 841-889.

Bhat, C.R. (1998) A Model of Post-Home Arrival Activity Participation Behavior. *Transportation Research Part B*, Vol. 32, No. 6, pp. 387-400.

Bhat, C. R. (2003) Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. Transportation Research Part B, 37(9), 837-855.

Bhat, C.R. (2005) A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions. Transportation Research Part B, 39(8), 679-707.

Bhat, C.R. (2008) The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions. *Transportation Research Part B*, Vol 42, pp. 274-303.

Bhat, C. R. and Gossen, R. (2004) A Mixed Multinomial Logit Model Analysis of Weekend Recreational Episode Type Choice. Transportation Research Part B, 38(9), 767-787.

Bhat, C. R. and Lockwood, A. (2004) On Distinguishing Between Physically Active and Physically Passive Episodes and Between Travel and Activity Episodes: An Analysis of Weekend Recreational Participation in the San Francisco Bay Area. Transportation Research Part A, 38(8), 573-592.

Bhat, C.R., and Koppelman, F.S. (1999) A Retrospective and Prospective Survey of Time-Use Research. Transportation, 26(2), 119-139.

Bhat, C.R. and Misra, R. (1999) Discretionary activity time allocation of individuals between in-home and out-of-home and between weekdays and weekends. Transportation 26(2), 193-229.

Bhat, C.R. and Srinivasan, S. (2005). A multidimensional mixed ordered-response model for analyzing weekend activity participation. *Transportation Research Part B*, 39(3), 255-278.

Bhat, C.R. and S. Sen (2006). Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. *Transportation Research Part B*, 40(1), 35-53.

Bhat, C.R., and J.Y. Guo (2007) A Comprehensive Analysis of Built Environment Characteristics on Household Residential Choice and Auto Ownership Levels. Transportation Research 41B(5), pp. 506-526.

Bhat, C. R., Govindarajan, A., and Pulugurta, V. (1998) Disaggregate Attraction-End Choice Modeling. Transportation Research Record, 1645, 60-68.

Bhat, C. R., Guo, J. Y., Srinivasan, S., and Sivakumar, A. (2004) Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns. Transportation Research Record, 1894, 57-66.

Bhat, C.R., J. Guo, S. Srinivasan, A. Sivakumar, A. Pinjari, and N. Eluru (2004) Activity-Based Travel-Demand Modeling for Metropolitan Areas in Texas: Representation and Analysis Frameworks for Population Updating and Land-Use Forecasting, Research Report 4080-6, prepared for the Texas Department of Transportation, Center for Transportation Research, The University of Texas at Austin, November 2004.

Bhat, C.R., Srinivasan, S., and Sen, S. (2006) A Joint Model for the Perfect and Imperfect Substitute Goods Case: Application to Activity Time-Use Decisions. Transportation Research Part B, 40(10), 827-850.

Birch, E.L. (2005) Who Lives Downtown? Brookings Institution Metropolitan Policy Program. November 2005.

Boarnet , M. G., and Sarmiento, S. (1998) Can Land-use Policy Really Affect Travel Behavior? A Study of the Link Between Non-work Travel and Land-Use Characteristics. Urban Studies, 35(7), 1155-1169.

Boarnet, M. G. and Crane, R. (2001) The Influence of Land Use on Travel Behavior: Specification and Estimation Strategies. Transportation Research Part A, 35, 823-845.

Boarnet, M.G., and S. Sarmiento (1998) Can Land-use Policy Really Affect Travel Behavior? A Study of the Link between Non-work Travel and Land-Use Characteristics. Urban Studies 35(7), pp. 1155-1169.

Bockstael, Nancy, W. Michael Hanemann, and Catherine L. Kling (1987) 'Estimating the value of water quality improvements in a recreation demand framework.' Water Resources Research 23(5), 951–960.

Cao, X. (2006). The Causal Relationship between the Built Environment and Personal Travel Choice: Evidence from Northern California. Doctoral Dissertation, University of California, Davis.

Cao, X., Mokhtarian, P. L., and Handy, S. L. (2007) Do Changes in Neighborhood Characteristics Lead to Changes in Travel Behavior? A Structural Equations Modeling Approach. Forthcoming, Transportation.

Cao, X., P.L. Mokhtarian, and S.L. Handy (2005) The Impacts of the Built Environment and Residential Self-selection on Nonwork Travel: A Seemingly Unrelated Regression Approach. Paper 06-1595 on the Transportation Research Board 85th Annual Meeting CD-ROM. Washington, DC.

Cao, X., Mokhtarian, P. L., and Handy, S. L. (2006a) Examining the impacts of residential self-selection on travel behavior: Methodologies and empirical findings. Research Report UCD-ITS-RR-06-18, Institute of Transportation Studies, University of California, Davis, November.

Cao, X., S.L. Handy, and P.L. Mokhtarian (2006b) The Influences of the Built Environment and Residential Self-selection on Pedestrian Behavior: Evidence from Austin, TX. Transportation 33(1), pp. 1-20.

Carpenter, S. and P.M. Jones (1983) Recent Advances in Travel Demand Analysis. Gower, Aldershot, England.

Cervero R. and M. Duncan (2003) Walking, Bicycling, and Urban Landscapes: Evidence from the San Francisco Bay Area. Final Summary Report, University of California Transportation Center, University of California, Berkeley.

Cervero R., and Duncan, M. (2002) Residential Self Selection and Rail Commuting: A Nested Logit Analysis. Working Paper, University of California Transportation Center, Berkeley, CA. Available at http://www.uctc.net/papers/604.pdf

Cervero, R. (1989) America's Suburban Centers: The Land Use-Transportation Link. Unwin-Hyman, Boston.

Cervero, R. (2002) Built Environments and Mode Choice: Toward a Normative Framework. Transportation Research Part D, 7(4), 265-284.

Cervero, R. and K. Kockelman (1997) Travel Demand and the 3Ds: Density, Diversity and Design. Transportation Research Part D, 2, pp. 199-219.

Cervero, R., and K.Wu (1997) Influences of land use environments on commuting choices: An analysis of large U.S. metropolitan areas using the 1985 American Housing Survey. Working Paper 683, 1997.University of California at Berkeley.

Cervero, R., and M. Duncan (2003) Walking, Bicycling, and Urban Landscapes: Evidence from the San Francisco Bay Area. American Journal of Public Health Vol. 93, No. 9, 2003, pp. 1478-1483.

Cervero, R., N. Creedman, M. Pohan, and M. Pai (2002) City CarShare: Assessment of Short-Term Travel-behavior Impacts. Report prepared for the Department of Transportation & Parking, City of San Francisco, CA.

Chatman, D.G. (2005) How the Built Environment Influences Non-work Travel: Theoretical and Empirical Essays. Dissertation submitted in fulfillment of the Ph.D. in Urban Planning, University of California, Los Angeles.

Chib, S. and E. Greenberg. (1998). Analysis of Multivariate Probit Models. *Biometrika* 85(2), 347–361.

Chib, S., P. B. Seetharaman, and A. Strijnev. (2004). Analysis of Multi-Category Purchase Incidence Decisions Using IRI Market Basket Data. *Advances in Econometrics* 16, 55–90.

Chib, S., P. B. Seetharaman, and A. Strijnev. (2005). Joint Modeling of Multiple Category Incidence and Private Label Choice, Accounting for No Category Incidence and Panel Heterogeneity. Working Paper, School of Management, University at Buffalo.

Chen, C. and Mokhtarian, P.L. (2006) Tradeoffs between time allocations to maintenance activities/travel and discretionary activities/travel Transportation. 33(3), 223-240.

Cirillo, C., and K.W. Axhausen (2002) Mode Choice of Complex Tours. Proceedings of the European Transport Conference, Cambridge.

Clark, W.A.V., and W.F.J. Van Lierop (1986) Residential Mobility and Household Location Modeling. Handbook of Urban and Regional Economics-I, pp. 97-132.

Copperman, R. and Bhat, C.R. (2007) An Analysis of the Determinants of Children's Weekend Physical Activity Participation. Transportation, 34(1), 67-87.

Cox, D. R. (1972). The Analysis of Multivariate Binary Data, Applied Statistics. *Journal of the Royal Statistical Society, Series C* 21(2), 113–120.

Crane, R (2000) The Influence of Urban Form on Travel: An Interpretive Review. Journal of Planning Literature, Vol. 15, No. 1, pp. 3-23.

Crane, R., and R. Crepeau (1998) Does neighborhood design influence travel? A behavioral analysis of travel diary and GIS data. Transportation Research Part D, Vol. 3, No. 4.

Daly, A. (1982) Estimating Choice Models Containing Attraction Variables. Transportation Research Part B, 16, 5-15.

Daly, A.J. and Zachary, S. (1978) Improved multiple choice models. In: Hensher, D.A. and Dalvi, M.Q., Editors, 1978. *Determinants of Travel Choice*, Saxon House, Sussex.

Davidson, W., R. Donnelly, P. Vovsha, J. Freedman, S. Ruegg, J. Hicks, J. Castiglione, and R. Picado (2007) Synthesis of first practices and operational research approaches in activity-based travel demand modeling. *Transportation Research Part A*, 41(5), pp. 464–488

de Palma, A., Motamedi, K., Picard, N., and Waddell, P. (2005) A model of residential location choice with endogenous housing prices and traffic for the Paris region. European Transport , 31, 67-82.

Deepak, S., A. Ansari, and S. Gupta. (2002). Investigating Consumer Price Sensitivities Across Categories. Working Paper, University of Iowa.

Demographia (2007) Accessed in November 2007, Available at http://www.demographia.com/

Dubin, J., and D. McFadden (1984) An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica,* 52, 345-62.

Edwards, Y.D. and Allenby, G.M. (2003). Multivariate analysis of multiple response data. *Journal of Marketing Research*, 40, 321-334.

Eliasson, J., and L. Mattsson (2000) A Model for Integrated Analysis of Household Location and Travel Choices, Transport Research 34A, pp. 375-394.

Eluru, N., A.R. Pinjari, J.Y. Guo, I. Sener, S. Srinivasan, R. Copperman, and C.R. Bhat. (2007) Population Updating System Structures and Models Embedded with the Comprehensive Econometric Microsimulator for Urban Systems (CEMUS). Technical paper, Department of Civil, Environmental and Architectural Engineering, The University of Texas at Austin, July 2007.

Ewing, R., and R. Cervero (2001) Travel and the Built Environment – Synthesis. *Transportation Research Record*: Journal of the Transportation Research Board, No. 1780, TRB, National Research Council, Washington, D.C., pp. 87-114.

Ewing, R., P. Haliyur, and W. Page (1994) Getting around a traditional city, a suburban planned unit development, and everything in between. *Transportation Research Record*: Journal of the Transportation Research Board, No. 1466, TRB, National Research Council, Washington, D.C., pp.53-62.

Ewing, R., R. Pendall, and D. Chen (2002) Measuring Sprawl and Its Impacts. Smart Growth America, Washington, D.C.

Ewing, R., W. Schroeer, and W. Greene (2005) School Location and Student Travel Analysis of Factors Affecting Mode Choice, Transportation Research Record 1895, pp. 55-63.

Fotheringham, A. S. (1983) Some Theoretical Aspects of Destination Choice and Their Relevance to Production-Constrained Gravity Models. Environment and Planning, 15A, 1121-1132.

Frank, L. D., and G. Pivo (1994) Impacts of mixed use and density on the utilization of three modes of travel: Single occupant vehicle, transit and walking. In Transportation Research Record: Journal of the Transportation Research Board, No. 1466, TRB, National Research Council, Washington, D.C., pp. 44-52.

Friedman, B., P. Gordon, and J. Peers (1994). Effect of neotraditional neighborhood design on travel characteristics. Transportation Research Record: Journal of the Transportation Research Board, No. 1466, TRB, National Research Council, Washington, D.C., 1994, pp.63-70.

Gliebe, J.P., and Koppelman, F.S. (2002) A model of joint activity participation between household members. Transportation, 29(1), 40-72.

Golob, T.F. and McNally, M.G. (1995) A model of household interactions in activity participation and the derived demand for travel. Presented at the European Institute of Retailing and Service Studies (EIRASS), Conference on Activity-Based Approaches, Eindhoven, The Netherlands, May.

Gossen, R. and Purvis, C.L. (2005) Changes in Women's Travel Time. Expenditures, 1990 – 2000. Paper presented at the Transportation Research Board Conference on Research on Women's Issues in Transportation, Conference Proceedings No. 35. http://onlinepubs.trb.org/onlinepubs/conf/CP35v2.pdf.

Goulias K. G. (2002) Multilevel analysis of daily time use and time allocation to activity types accounting for complex covariance structures using correlated random effects. Transportation, 29(1), 31-48.

Goulias, K.G., and Henson, K.M. (2006) On altruists and egoists in activity participation and travel: who are they and do they live together? Transportation, 33(5), 447-462.

Greenwald, M. J. and Boarnet, M. G. (2001) Built Environment as Determinant of Walking Behavior - Analyzing Nonwork Pedestrian Travel in Portland, Oregon. Transportation Research Record, 1780, 33-42.

Greenwald, M.J. (2003) The Road Less Traveled: New Urbanist Inducements to Travel Mode Substitution for Nonwork Trips. Journal of Planning Education and Research 23, pp. 39-57.

Greenwald, M.J., and M.G. Boarnet (2001) Built Environment as Determinant of Walking Behavior - Analyzing Nonwork Pedestrian Travel in Portland, Oregon. Transportation Research Record 1780, pp. 33-42.

Guevara C.A. and Ben-Akiva, M.E. (2006) Endogeneity in residential location choice models. Transportation Research Record, 1977, 60-66.

Guo, J.Y., Bhat, C.R., and Copperman, R.B. (2006) Effect of the Built Environment on Motorized and Non-Motorized Trip Making: Substitutive, Complementary, or Synergistic? Technical paper, Department of Civil and Environmental Engineering, University of Wisconsin–Madison.

Habib, K.M., and E.J. Miller (2006). Modeling activity generation: A utility based model for activity-agenda formation. Presented at the 11th International Association for Travel Behavior Research, Kyoto.

Hammond, D. (2005) Residential Location and Commute Mode Choice. Dissertation, submitted in fulfillment of the MSC in Transport and Planning, University of Wales, Cardiff, United Kingdom.

Handy, S. L. and Clifton, K. J. (2001) Local Shopping as a Strategy for Reducing Automobile Travel. Transportation, 28(4), 317-346.

Handy, S. (1996) Methodologies for Exploring the Link Between Urban Form and Travel Behavior. Transportation Research Part D, Vol. 1, No. 2, pp. 151-165.

Handy, S., X. Cao, and P.L. Mokhtarian (2005) Correlation or Causality between the Built Environment and Travel Behavior? Evidence from Northern California. Transportation Research 10D(6), pp. 427-444.

Handy, S., X. Cao, and P.L., Mokhtarian (2006) Self-Selection in the Relationship Between Built Environment and Walking? Evidence from Northern California. Journal of the American Planning Association, 72(1), pp. 55-74.

Handy, S., X. Cao, T.J. Buehler, and P. Mokhtarian. (2005) The Link Between the Built Environment and Travel Behavior: Correlation or Causality? Presented at 2005 Annual Transportation Research Board Meeting, on TRB CD-ROM.

Handy, S.L., and K.J. Clifton (2001) Local Shopping as a Strategy for Reducing Automobile Travel. Transportation 28(4), pp. 317-346.

Hanemann, W.M. (1978). A Methodological and Empirical Study of the Recreation Benefits From Water Quality Improvements. Ph.D. Dissertation, Harvard University.

Harvey, A.S., and Taylor, M.E. (2000) Activity settings and travel behaviour: A social contact perspective. Transportation, 27(1), 53-73.

Heckman, J. J. (1979) Sample Selection Bias as a Specification Error, *Econometrica*, 47(1), 153-161.

Herriges, J.A., Kling, C.L., and Phaneuf, D.J. (2004). What's the use? Welfare estimates from revealed preference models when weak complementarity does not hold. *Journal of Environmental Economics and Management*, 47, 55-70.

Hess, D. (2001) Effect of free parking on commuter mode choice - Evidence from travel diary data. Journal of the Transportation Research Board, No.1753, TRB, National Research Council, Washington, D.C., 2001, pp.35-42.

Hunt, J.D., and J.E. Abraham (2007) Influences on Bicycle Use. Transportation, 34(4), pp. 453-470.

Jones PM (1979) New approaches to understanding travel behaviour: the human activity. approach. In Hensher DA, Stopher PR (ed.) Behavioral travel modeling. Redwood Burn Ltd., London, pp. 55-80.

Jones, P. M., Koppelman, F.S., and Orfeuil, J.P. (1990) Activity analysis: state of the art and future directions, in Developments in Dynamic and Activity-Based Approaches to Travel Analysis, 34-55, Gower, Aldershot, England.

Kapur, A., Bhat, C.R. (2007) On Modeling Adults' Daily Time Use by Activity Purpose and Accompaniment Arrangement. Forthcoming, Transportation Research Record.

Khattak, A.J., and D. Rodriguez. (2005) Travel Behavior in Neo-traditional Neighborhood Developments: A Case Study in USA. Transportation Research Part A, Vol. 39, No. 6, pp481-500.

Kim, J., Allenby, G.M., and Rossi. P.E. (2002). Modeling consumer demand for variety. *Marketing Science*, 21, 229-250.

Kitamura, R. (1988) An evaluation of activity-based travel analysis. *Transportation* 15(1-2), pp. 9-34.

Kitamura, R., P.L. Mokhtarian, and L. Laidet (1997) A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area. Transportation 24, pp. 125-158.

Kockelman, K.M. (1997) Travel behavior as a function of accessibility, land use mixing and land use balance: Evidence from the San Francisco Bay Area. Transportation Research Record: Journal of the Transportation Research Board, No. 1607, TRB, National Research Council, Washington, D.C., pp. 116-25.

Kraan, M. (1996) Time to travel?; a model for the allocation of time and money. Unpublished Ph.D. dissertation, Department of Civil Engineering, University of Twente, The Netherlands.

Krizek, K. J. (2000) Pretest-Posttest Strategy for Researching Neighborhood-Scale Urban Form and Travel Behavior. Transportation Research Record 1722, pp. 48-55.

Krizek, K.J. (2003) Residential Relocation and Changes in Urban Travel - Does Neighborhood-Scale Urban Form Matter? Journal of The American Planning Association 69(3), pp. 265-281.

Kuhn, H.W. and Tucker, A.W. (1951). Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman (Ed.), University of California Press, Berkeley, CA, 481-492.

Kurani, K.S. and Kitamura, R. (1996) Recent developments and the prospects for modeling household activity schedules. Report prepared for the Los Alamos National Laboratory, Institute of Transportation Studies, University of California, Davis, CA.

Kurani, K.S. and Lee-Gosselin, M.E.H. (1996) Synthesis of past activity analysis applications. Presented at the Travel Model Improvement Program (TMIP) Conference on activity-based travel forecasting, New Orleans, June 2-5.

Lawrence, F., M.A. Bradley, S. Kavage, J. Chapman, and Keith Lawton Consulting (2007) Urban Form, Travel Time, and Cost Relationships with Tour Complexity and Mode Choice. Forthcoming, Transportation.

Lee, L.-F. (1983) Generalized Econometric Models with Selectivity. *Econometrica*, 51, 507-512.

Lee, B.A., and Guest, A.M. (1983) Determinants of Neighborhood Satisfaction: A Metropolitan-Level Analysis. The Sociological Quarterly 24, 287-303.

Lee, L. -F., and M. Pitt (1986). Microeconomic Demand Systems with Binding Non Negativity Constraints: The Dual Approach. Econometrica, 54, 1237-1242.

Lerman, S.R (1976) Location, Housing, Automobile Ownership and Mode to Work: A Joint Choice Model. Transportation Research Record 610, pp. 6-11.

Levinson, D.M. (1999) Space, money, life-stage, and the allocation of time. Transportation, 26(2), 141-171.

Lewbel, A. (2004) Simple Estimators For Hard Problems: Endogeneity in Discrete Choice Related Models. Working Papers in Economics No. 604, Department of Economics, Boston College.

Litman, T. (2002) The costs of automobile dependency and the benefits of balanced transportation, Victoria Transport Policy Institute, Canada.

Louviere, J., Train, K., Ben-Akiva, M., Bhat, C.R., Brownstone, D., Cameron, T.A., Carson, R.T., DeShazo, J.R., Fiebig, D., Greene, W., Hensher, D., and D. Waldman (2005) Recent Progress on Endogeneity in Choice Modeling. Marketing Letters, Special Issue: Sixth Invitational Choice Symposium, 16(3-4), 255-265.

Lu, X. and Pas, E.I. (1997) A structural equation model of the relationships among socio-demographics, activity participation and travel behavior. Presented at the 76th Annual Meeting of the Transportation Research Board, Washington, D.C.

Lu, X., and E.I. Pas (1999) Socio-demographics, activity participation and travel behavior. Transportation Research Part A, 33, 1–18.

Lund, H. (2003) Testing the Claims of New Urbanism. APA Journal, Vol. 69, No. 4, pp. 414-429.

Ma, Y. and P. B. Seetharaman. (2004). The Multivariate Logit Model for Multicategory Purchase Incidence Outcomes. Working Paper, Rice University.

Ma, Y., P. B. Seetharaman, and C. Narasimhan. (2005). Empirical Analysis of Competitive Pricing Strategies with Complementary Product Lines. Working Paper, Washington University in St. Louis.

Maat, K., and H.J.P. Timmermans (2007) Household Car Ownership in Relation to Residential and Work Location. Paper presented at the Transportation Research Board annual meeting, January 2007.

Manchanda, P., A. Ansari, and S. Gupta. (1999). The Shopping Basket: A Model for Multicategory Purchase Incidence Decisions. *Marketing Science* 18(2), 95–114.

Marans, R. W. (1979) The Determinants of Neighborhood Quality. Annual Housing Survey Studies. No. 3. Washington, D.C.: Office of Policy Development and Research, US. Department of Housing and Urban Development.

McFadden, D. (1978). Modelling the choice of residential location. In: Karlqvist, A., Lundqvist, L., Snickars, F. and Weibull, J.W., Editors. *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam.

Meloni, I., Guala, L., and A. Loddo (2004)Time allocation to discretionary in-home, out-of-home activities and to trips. Transportation, 31(1), 69-96.

Meloni, I., Spissu, E., and M. Bez (2007) A Model of the Dynamic Process of Time Allocation to Discretionary Activities. Transportation Science, 41(1), 15-28.

Merriman, D. (1994) Compensation for Commutes in the Land and Labor Markets: Some Evidence from the Tokyo Metropolitan Area. Journal of Regional Science 34(3): 297-324.

Miller E.J., J.R. Matthew, and J.A. Carrasco (2005) A Tour-based Model of Travel Mode Choice. Transportation 32, pp. 399-422.

MORPACE International, Inc. (2002) Bay Area Travel Survey Final Report, March 2002. ftp://ftp.abag.ca.gov/pub/mtc/planning/BATS/BATS2000/

Mokhtarian, P.L., and X. Cao (2007) Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B*, Volume 42, Issue 3, pp. 204-228.

Newman, P., and J. Kenworthy (1998) *Sustainability and Cities: Overcoming Automobile Dependency*, Island Press.

O'Toole, R. (1996) The Vanishing Automobile and Other Urban Myths. Thoreau Institute, Oregon.

Pas, E.I. (1996) Time and travel demand analysis and forecasting: theory, data collection and modeling. Presented at the Conference on Theoretical Foundations of Travel Choice Modeling, Stockholm, Sweden, August 7-11.

Pas, E.I. and A.S. Harvey (1997) Time use research and travel demand analysis and modeling. In P. Stopher and M. Lee-Gosselin (editors), Understanding Travel Behavior in an Era of Change, 315-338, Elsevier Science Ltd., Oxford.

Phaneuf, D.J. (1999). A Dual Approach to Modeling Corner Solutions in Recreation Demand. Journal of Environmental Economics and Management, 37, 85-105.

Phaneuf, D.J., Kling, C.L., and Herriges, J.A. (2000). Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand. *The Review of Economics and Statistics*, 82(1), 83-92.

Pinjari, A.R., Pendyala, R.M., Bhat, C.R., and Waddel, P.A. (2007) Modeling Residential Sorting Effects to Understand the Impact of the Built Environment on Commute Mode Choice. Forthcoming, Transportation.

Pinjari, A.R., R.M. Pendyala, C.R. Bhat, and P.A. Waddell (2007b) Modeling Residential Sorting Effects to Understand the Impact of the Built Environment on Commute Mode Choice. Forthcoming in Transportation.

Rajagopalan, B.S., and K.S., Srinivasan (2008) Integrating Household-Level Mode Choice and Modal Expenditure Decisions in a Developing Country: Multiple Discrete-Continuous Extreme Value Model. Proceedings of the 87th Transportation Research Board Annual Meeting, Washington, D.C. (January, 2008).

Rajamani, J., C.R. Bhat, S. Handy, G. Knaap, and Y. Song (2003) Assessing the Impact of Urban Form Measures in Nonwork Trip Mode Choice After Controlling for Demographic and Level-of-Service Effects, Transportation Research Record 1831, pp. 158-165.

Rich, J.H., and O.N. Nielsen (2001) A Microeconomic Model for Car Ownership, Residence and Work Location. Paper presented at the European Transport Conference (PTRC), Cambridge, UK.

Rietveld, P. (2001) Biking and Walking: The Position of Non-Motorized Transport Modes in Transport Systems. In K. J. Button and D.A. Hensher (Ed.), Handbook of Transport Systems and Traffic Control, pp. 299-319, Pergamon, London, United Kingdom.

Roads and Traffic Authority (RTA) (2007) Cycling in Sydney: Bicycle Ownership and Use. Prepared by the Transport Data Centre, New South Wales Department of Transport.

Rodriguez, D.A, H.M. Young, and R.J. Schneider (2005) An Easy to Compute Index for Identifying Built Environments that Support Walking. Presented at the 85th Annual Meeting of the Transportation Research Board, January, Washington, D.C.

Rodriguez, D.A., and J. Joo (2004) The relationship between non-motorized mode choice and the local physical environment, *Transportation Research D*, Vol 9, issue 2, pp. 151-173.

Russell, G. J. and W. A. Kamakura. (1997). Modeling Multiple Category Brand Preference with Household Basket Data. *Journal of Retailing* 73(1), 439–461.

Russell, G. J. and A. Peterson. (2000). Analysis of Cross Category Dependence in Market Basket Selection. *Journal of Retailing* 76(3), 367–392.

Salon, D. (2006) Cars and the City: An Investigation of Transportation and Residential Location Choices in New York City. Dissertation, Department of Agricultural and Resource Economics, University of California, Davis.

Sen. S. (2006). A Joint Multiple Discrete Continuous Extreme Value (MDCEV) Model and Multinomial Logit Model (MNL) for Examining Vehicle Type/Vintage, Make/Model and Usage Decisions of the Household. Ph.D. dissertation, Department of Civil, Environmental and Architectural Engineering, University of Texas at Austin.

Sen, S., and C.R. Bhat (2006). The Impact of Demographics, Built Environment Attributes, Vehicle Characteristics, and Gasoline Prices on Household Vehicle Holdings and Use. Technical paper, Department of Civil Engineering, The University of Texas at Austin.

Schwanen, T. and P.L. Mokhtarian (2003) Does Dissonance between Desired and Current Neighborhood Type Affect Individual Travel Behavior? An Empirical Assessment from the San Francisco Bay Area. Proceedings of the European Transport Conference (ETC), October 8-10, 2003, Strasbourg, France.

Schwanen, T., and P.L. Mokhtarian. (2004) The Extent and Determinants of Dissonance Between Actual and Preferred Residential Neighborhood Type. Environment And Planning B, Vol. 31, No. 5, pp. 759-784.

Schwanen, T. and P.L. Mokhtarian (2005a) What Affects Commute Mode Choice: Neighborhood Physical Structure or Preferences Toward Neighborhoods? Journal of Transport Geography 13 (1), pp. 83-99.

Schwanen, T., and P.L. Mokhtarian (2005b) What if You Live in the Wrong Neighborhood? The Impact of Residential Neighborhood Type Dissonance on Distance Traveled. Transportation Research 10D(2), pp. 127-151.

Sener, I.N. and Bhat, C.R. (2007) An Analysis of the Social Context of Children's Discretionary Activity Participation. Forthcoming, Transportation.

Shay E. and A. Khattak (2005) Auto ownership and use in neo-traditional and conventional neighborhoods, Presented at 2005 Annual Transportation Research Board Meeting, Washington D.C., January 2005.

Siegel, J. (1975) Intrametropolitan Migration: A Simultaneous Model of Employment and Residential Location of White and Black Households. Journal of Urban Economics 2: 29-47.

Simma, A. and K.W. Axhausen (2001) Structures of Commitment in Mode Use: A Comparison of Switzerland, Germany and Great Britain, Conference paper STRC 2001. First Swiss Transport Research Conference Monte Verità/Ascona, March 2001.

Simpson, W. (1980) A Simultaneous Model of Workplace and Residential Location Incorporating Job Search. Journal of Urban Economics 8, 330-349.

Song, Y. and G. J. Knapp (2004) Measuring Urban Form, Journal of the American Planning Association, 70, pp. 210-226.

Srinivasan, S. (2002) Quantifying Spatial Characteristics of Cities. Urban Studies, 39, pp. 2005-2028.

Srinivasan, S. (2004) Modeling household interactions in daily activity generation. Dissertation, Department of Civil, Environmental and Architectural Engineering, The University of Texas at Austin.

Srinivasan, S., and C.R. Bhat (2006). A Multiple Discrete-Continuous Model for Independent- and Joint- Discretionary-Activity Participation Decisions. *Transportation*, 2006 TRB Special Issue, Vol. 33, No. 5, pp. 497-515.

Srinivasan, T.C., and R.S. Winer (1994). Using neoclassical consumer-choice theory to produce a market map from purchase data. *Journal of Business and Economic Status*. 12, 1-9.

Terza, J., Wilson, P. (1990). Analyzing Frequencies of Several Types of Events: A Mixed Multinomial-Poisson Approach. *Review of Economics and Statistics*, 72, 108-115.

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

TDM Encyclopedia of Victoria Transport Policy Institute (2007) Accessed on November 30, 2007. Available at http://www.vtpi.org/tdm/

Train, K. (1980) A Structured Logit Model of Auto Ownership and Mode Choice. The Review of Economic Studies 47(2), pp. 357-370.

Transportation Research Board (TRB) (2005) Smart Growth and Transportation: Issues and Lessons Learned. Conference Proceedings 32, Washington, D.C.

Transportation Research Board and Institute of Medicine (TRB-IOM). (2005). *Does the Built Environment Influence Physical Activity? Examining the Evidence* – Special Report 282. Washington, DC. January.
Available at http://onlinepubs.trb.org/onlinepubs/sr/sr282.pdf.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26, 24-36.

von Haefen, R.H. (2003). Incorporating observed choice into the construction of welfare measures from random utility models. *Journal of Environmental Economics & Management*, 45(2), 145-165.

von Haefen, R.H. and Phaneuf, D.J. (2005). Kuhn-Tucker demand system approaches to nonmarket valuation. In *Applications of Simulation Methods in Environmental and Resource Economics*, R. Scarpa and A.A. Alberini (Eds.), Springer.

Vovsha P. and Bradley, M. (2005) Advanced Activity-Based Models in a Context of Planning Decisions. Transportation Research Record, 1981, 34-41.

Waddell, P. (1993a) Exogenous Workplace Choice in Residential Location Models: is the Assumption Valid in a Multinodal Metropolis? Geographical Analysis, 25(1), 65-82.

Waddell, P. (1993b) A Multinomial Logit Model of Race and Urban Structure. Urban Geography, 13(2), 127-141.

Waddell, P. (2001) Towards a Behavioral Integration of Land Use and Transportation Modeling. In D. Hensher, (Ed.), Travel Behavior Research: The Leading Edge, New York: Pergamon, pp. 65-95.

Waddell, P. (2006) Accessibility and Residential Location: The Interaction of Workplace, Housing Tenure, Residential Mobility and Location Choices. In Modelling Residential Location Choice, John Preston, Francesca Pagliara and David Simmonds eds. Ashgate.

Waddell, P., and Nourzad, F. (2002) Incorporating Non-motorized Mode and Neighborhood Accessibility in an Integrated Land Use and Transportation Model System. Transportation Research Record, 1805, 119-127.

Waddell, P., M. Outwater, C.R. Bhat, and L. Blain (2002) Design of an Integrated Land Use and Activity-Based Travel Model System for the Puget Sound Region, Transportation Research Record 1805, pp. 108-118.

Wegener, M., P. Waddell, and I. Salomon. (2002) Sustainable Lifestyles? Microsimulation of Household Formation, Housing Choice and Travel Behavior. In W. Black and P. Nijkamp, (Eds.), *Social Change and Sustainable Transport*. Bloomington, IN: Indiana University Press, pp. 125-131.

Weisbrod, G.E., S.R. Lerman, and M. Ben-Akiva (1980) Trade-offs in Residential Location Decisions: Transportation versus Other Factors. Transport Policy and Decision Making 1(1), pp. 13-26.

Wales, T.J. and Woodland, A.D. (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics*, 21(3), 263-85.

Walker, J. (2002) The Mixed Logit Model: Dispelling the Misconceptions of Identification. *Transportation Research Record* 1805: Journal of the Transportation Research Board, pp. 86-99.

Wang, Y. (2003). Estimation of Recreation Demand Models Using Dual Approaches. Ph.D. dissertation, Department of Agricultural & Resource Economics, North Carolina State University.

Wigan, M.R. (1984) Bicycle Ownership, Use and Exposure: Participation and Activity Patterns in Melbourne, Australia. Transportation Research Part A, 18(5-6), pp. 379-398.
Wooldridge, J. (2002) Econometric Analysis of Cross Section and Panel Data. MIT Press.

Williams, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A,* Vol 9, pp. 285–344.

Yamamoto, T., and Kitamura, R. (1999) An analysis of time allocation to in-home and out-of-home discretionary activities across working days and non- working days. Transportation, 29(2), 231-250.

Ye, X. and Pendyala R. M. (2005) A Model of Daily Time Use Allocation Using Fractional Logit Methodology. In H.S. Mahmassani (ed.) Transportation and Traffic Theory: Flow, Dynamics, and Human Interaction. Pergamon, Elsevier Science Ltd., Oxford, UK, pp. 507-524.

Ye, X., R.M. Pendyala, and G. Gottardi (2007) An Exploration of the Relationship between Mode Choice and Complexity of Trip Chaining Patterns. Transportation Research 41B(1), pp. 96-113.

Zhang, M. (2004) The role of land use in travel mode choice: Evidence from Boston and Hong Kong. Journal of the American Planning Association, Vol. 70, No. 3, pp. 344-61.

Zhang, M. (2006) Travel Choice with No Alternative: Can Land Use Reduce Automobile Dependence? Journal of Planning Education and Research 25(3), pp. 311-326.

Zhou, B., and K. Kockelman (2008) Self-Selection in Home Choice: Use of Treatment Effects in Evaluating the Relationship between the Built Environment and Travel Behavior. Proceedings of the 87th Annual Meeting of the Transportation Research Board, Washington D.C., January 2008.

Vita

Abdul Rawoof Pinjari was born in Kurnool town of Andhra Pradesh state in India. He went to Jawahar Navodaya Vidyalaya, Kurnool for his schooling and secondary education. In 1998, he joined the Indian Institute of Technology (IIT), Madras, from where he received the degree of Bachelor of Technology (B.Tech.) in Civil Engineering in July 2002. Mr. Pinjari then enrolled at the University of South Florida (USF), Tampa, in the Fall of 2002 for graduate study in Civil Engineering with a specialization in Transportation Systems. He received the degree of Master of Science in Civil Engineering (M.S.C.E.) from USF in May 2004. Subsequently, in August 2004, he began his doctoral study in Transportation Systems at the University of Texas at Austin.

Permanent address: House number 104, Sri Lakshmi Builders
Opposite to Raj Theatre
Kurnool - 518001
Andhra Pradesh, India.

This dissertation was typed by the author.