Michael Norman Bagley
September 1999
Civil and Environmental Engineering

Incorporating Residential Choice into Travel Behavior-Land Use Interaction Research: A Conceptual Model with Methodologies for Investigating Causal Relationships

Abstract

This dissertation investigates the factors that influence an individual's residential choice. The role that residential choice plays in other individual decisions is also investigated, with an emphasis placed on understanding the importance of land use configuration on individual travel demand. To achieve this, a conceptual model of residential choice/preference was developed that was a comprehensive reflection of those relationships supported by the literature and by informed judgement. The complexity of this model is seen in the many interdependent relationships that involve residential choice. For example, in choosing a place to live, a household may evaluate a dwelling unit and/or neighborhood according to how it fits along several interrelated dimensions, such as: housing type, neighborhood type, distance to work, distance to shopping and other household-related activities, type of mode to work and car ownership.

The measurement of neighborhood type as a continuous variable through factor analysis was another important part of this work. A two-factor disaggregate solution representing traditional and suburban neighborhood dimensions was used in this study. This approach allowed for a single area to possess attributes of both types of neighborhoods, and allowed individuals within the same area to face different neighborhood characteristics – a flexibility amply justified by the empirical results. Further, it has the statistical advantage of producing continuous measures of

endogenous variables, a trait that is desirable in both regression and structural equation models.

The data analyzed in this study came from 852 individuals from five neighborhoods in the San Francisco Bay Area. Information such as trip records, life style preferences, and attitudes towards urban transportation, housing and the environment, were incorporated with household demographic and socio-economic data to perform multivariate statistical analyses of an individual's residential choice. Specifically, three different sets of models were estimated: 1) a binary model of residential choice (adjusted $?^2$ = 0.52), where residential choice alternatives included suburb (= 1) and traditional (= 0), 2) single-equation regression models of elements in the conceptual model (with adjusted R^2 values ranging from 0.39 for residential choice = traditional to 0.02 for travel demand = daily walk/bike miles rate), and 3) structural equation models (with good model fit indices such as the relative fit index = 0.913)

A particularly noteworthy finding supported by all of the models referred to above is that attitude and lifestyle variables play the greatest role in explaining residential choice and travel demand. Results suggest that the association commonly observed between neighborhood type and travel patterns is not one of direct causality, but due to correlations of each of those variables with others. In particular, is believed that when attitudinal, lifestyle, and sociodemographic variables are accounted for, neighborhood type has very little influence on travel behavior.

Dissertation Committee Chair

CHAPTER 1

INTRODUCTION

Spatial interaction modeling is of fundamental importance to transportation and land use researchers and planners. Studies in this area stemmed from interest in and concern regarding rapid urban growth and development during the 1960s (e.g., Kain, 1962b; Lierop, 1986). In the hope of understanding the growth and geographic arrangement of human settlements, researchers developed theoretical approaches to urban modeling oriented toward describing the spatial pattern of urban growth (e.g., Alonso, 1964).

As the urban spatial structure is fundamentally influenced by the location decisions of households, one of the most important spatial interaction modeling components is residential choice. Lierop (1986) categorized residential neighborhood location research into three areas: 1) residential mobility, 2) residential choice (both housing and neighborhood choice), and 3) relocation. Before continuing into the work on residential choice, it is useful to briefly introduce the concepts of residential mobility and relocation. Residential relocation research involves the study of an individual's decisions to move from *and* locate to a residence, while residential mobility research only concerns the decision to move (Porell, 1982). Neither residential mobility nor relocation are a focus of this study since we only have specific data about the current residence of each respondent. The lack of data on each person's past home locations and motivations for moving to the current location prevent a proper residential relocation study. However, the rich data we have on the present neighborhood and individual attitudes permits a useful study of residential choice, the focus of this dissertation. An important residential mobility study (Verster, 1985) is reviewed in the next chapter, with respect to the insight it offers into residential choice behavior.

Residential choice research is complex because household location decisions are highly interdependent. In choosing a place to live, a household may evaluate a dwelling unit and/or neighborhood on several interrelated dimensions, such as: housing type, neighborhood type, distance to work, distance to shopping and other household-related activities, and availability of various modes of transportation. The interconnectedness of these dimensions also shows the value of residential choice studies to improve our understanding of peoples' travel behavior.

The spatial choice of residential location faced by an individual or household depends on both spatial and nonspatial factors. In terms of spatial factors, the location of a person's residence relative to his or her activities of greatest interest (and/or frequency) in a geographic location at a given time is very important. Possibly equally important to the decision are the nonspatial factors, such as the person's socioeconomic characteristics, attitudinal orientation, lifestyle, and familiarity with the alternatives involved.

Though residential location will be at the center of the study's hypotheses and modeling efforts, investigating the core relationships between travel behavior and land use configuration of the residential neighborhood will be an important part of this research. Understanding these relationships will improve the theory and empirical development of residential choice models.

The organization of this dissertation is as follows. The remainder of Chapter 1 contains a short description of research on the impact of neighborhood design on travel behavior, a definition of residential location, and the objectives of the dissertation. Chapter 2 is a selected review of relevant past research. An original conceptual model of residential choice, with supporting literature, is discussed in the next chapter along with tables of potential model variables. The fourth chapter of the dissertation includes a description of this study's data set and background, along with a discussion of variable development for residential choice modeling. Analysis of a binary logit model of residential choice (traditional neighborhoods versus suburban neighborhoods) follows, and leads to Chapter 6 which contains formulation of continuous measures of residential choice that can be modeled using regression techniques. Analysis and discussion of single-equation regression models form the next chapter, a chapter which provides the foundation for the development of the more complex multi-equation structural equation models (SEMs) presented in Chapter 8. Chapter 8 also contains discussion of the concept of causality in relation to the empirical estimation of a modification of the conceptual model presented in Chapter 3. The final chapter contains a short summary of the main findings and implications of the research, as well as a discussion of potential future extensions of this dissertation.

1.1 Land Use and Travel Behavior

The investigation of interactions between travel behavior and land use is one key stream of research motivating the present study. In simple terms, work in this area examines how urban form affects peoples' travel patterns. A specific part of this research is the study of neighborhood design, where researchers are investigating travel variability among individuals in different neighborhoods (e.g., Cervero and Radisch, 1996; Friedman *et al.*, 1994).

The passage of the Intermodal Surface Transportation Efficiency Act (ISTEA) in 1991 created a new transportation policy focus for the 1990s, one that, among other things, emphasized planning for non-motorized modes of travel. Consequently, urban planners and transportation professionals have an even greater interest than before in creating land development policies that facilitate more non-motorized trips. This motivation has supported the implementation of pedestrian- and transit-oriented developments (PODs, TODs), where it is expected that people are more likely to walk and bike to a destination than is the case for an automobile-dominated neighborhood. In fact, this potential decrease in auto trips in POD and TOD neighborhood designs has some researchers examining whether land use strategies can significantly mitigate traffic congestion (see, for example, Atash, 1993).

In conclusion (as will be discussed further later), research has demonstrated that land use and travel decisions are related to neighborhood design (e.g., Friedman *et al.*, 1994). Consequently, as neighborhood design has the potential to influence a households' decision on residential location, the relationship between them is very connected to land use and travel behavior factors (Handy, 1997). For example, an individual who enjoys walking and/or using public transit is likely to choose a location where she can have that option. The careful consideration of the complex interrelationships mentioned above are important to model development and interpretation.

1.2 Defining Residential Location for Model Development

Residential location can be categorized in two primary ways: as a type of dwelling unit, such as a single-family home or high-density apartment, and as a type of neighborhood, such as a quiet rural community or a lively city-center area. Researchers have also studied specific neighborhoods and/or dwelling units. For example, Lindstrom (1997) selected two neighborhoods that were known to be very different in their socio-economic diversity and residential density to test the impact of race and urban form on choice of residential neighborhood.

Residential choice researchers have generally selected either dwelling unit type or neighborhood type as a dependent variable in model development, with the other component included among the explanatory variables. This is natural for two reasons. First, the dependent variable chosen will be based on the goal of the study, which may be oriented toward understanding only one of those two choices. Second, precisely determining the causality of residential choice is difficult due to the interdependence between the neighborhood and dwelling unit variables. It seems reasonable that some individuals or households will focus on the type of dwelling unit as the main determinant of residential choice; that is, some people will decide that they need a single-family dwelling unit with a very large yard, and then they will seek a neighborhood that meets this requirement (dwelling unit -----> neighborhood choice). On the other hand, though hypothesized to be a less-likely scenario, neighborhood choice may influence a household's dwelling unit decision. For example, a quiet and safe neighborhood with parks nearby may be the primary focus of a residential choice, and once a neighborhood(s) is found that satisfies this need the type of dwelling unit chosen may be affected by alternatives available in the chosen neighborhood (neighborhood choice ----> dwelling unit). Another example is a household's desire to locate in a particular neighborhood that is very prestigious, and then choosing a dwelling unit available in that neighborhood (e.g., a high-rise condominium). In other cases, the choice of neighborhood and dwelling unit may be made more or less simultaneously, with the choice set consisting of various neighborhood/dwelling unit combinations.

1.3 Developing a Measure for Neighborhood Type

The residential choice models in this dissertation use neighborhood type as the dependent variable. There are many different ways to define a neighborhood. However, for this dissertation we will focus on a neighborhood characteristic that has been studied in connection with travel, and that is how "(neo)traditional" it is (see, for example, Friedman *et al.*, 1994). The level of "traditionalness" a neighborhood has may be determined by how well it embodies traditional qualities such as high-density development, mixed uses, and a grid-based street structure.

It can be seen from the foregoing discussion that defining a neighborhood dependent variable for residential choice models is a complex task. Models with dependent variables based solely on an individual trait (such as housing size) are likely not to capture well the variation in an individual's residential choice. This is due to the fact that a person or household chooses a combination of traits when selecting a residence. As each of the traits of this study's neighborhoods are measured separately, it is necessary to develop a residential choice dependent variable that represents a variety of traits simultaneously. In Chapter 6 a measure of neighborhood type that was developed through factor analysis is described and is the basis for a majority of the residential choice empirical work found in this dissertation.

1.4 Dissertation Objectives

There are three interrelated objectives for this dissertation. The first mission is to present a realistic conceptual model of the relationships among residential choice, job location, travel behavior, and other major components involved with spatial interaction modeling at a disaggregate level. The next aims are to provide a better understanding, through empirical models, of what motivates an individual to choose a certain type of residential location, and of the relationship between residential location and travel behavior. In particular, we want to identify the role that attitudinal and lifestyle variables play in people's residential location choice and travel behavior.

Achieving the above-mentioned objectives will result in some useful contributions to the land use and travel research area. First, the development of a realistic conceptual model of the relationships among residential choice, job location, travel behavior, and other major components involved with spatial interaction modeling at a disaggregate level can lead to the improvement of residential choice models. By better understanding the underlying factors motivating an individual or household to select a particular type of residential location, transportation and urban planners can better understand the market for various neighborhood types (neotraditional, rural, etc.). In doing so, researchers may be better able to predict adoption of those types of neighborhoods, and in turn, better estimate the impact of changing residential land use patterns on travel behavior and vehicle emissions.

A large majority of the research to date has mainly consisted of pointing out correlations between travel patterns and characteristics of urban form. An important next step taken in this study is the development of insight into the behavior underlying these correlations (i.e., the causal relationships). Knowledge of the causal relationships is important for improved land use/travel policies to be created (Handy, 1997). For example, is the association found in previous studies (e.g., Cervero, 1996b and Friedman *et al*., 1994), of denser and more mixed land uses with fewer vehicle-trips and lower distances traveled, a directly causal one, or is the influence of land use on travel more indirect or even a spurious artifact of correlation with other variables such as attitudes? The answer could have a large impact on the above mentioned land/use travel policies.

The incorporation of factor analysis into empirical land use/transportation analysis, a contribution of this study, may help answer the above question. Attitudes, lifestyles, and neighborhood type are examples of variables that cannot be well defined by a single measure, and are usefully modeled as latent variables or factors derived from the combination of multiple observed indicators. Indeed, given that a person's perceptions, attitudes and preferences have a significant impact on all of her or his decisions, the use of attitudinal variables in the estimation of models can be valuable. In particular, few residential choice studies have incorporated attitudinal variables (see for example, Prevedouros, 1992), with the present study advancing the state of the art in that regard. In conclusion, the conceptual model and the statistical models (especially the structural equations modeling in Chapter 8) developed in this dissertation may contribute to a better understanding of individual residential choice, and the potential causal impact of urban form on travel demand.

1.5 Chapter 1 Summary

Chapter 1 provides a foundation for the development of this dissertation's analysis of residential choice. The present study was placed in the context of spatial interaction modeling and research on land use/travel behavior interactions. The selection of residential location modeling in general, and neighborhood type as the key dependent variable in particular, was explained on the basis of the available data. A discussion of the major objectives of the study and their importance to transportation and urban planning research concluded Chapter 1.

CHAPTER 2

RESIDENTIAL CHOICE LITERATURE

2.1 Introduction

Modeling an individual's or household's residential choice is a complex task due to the many interrelated variables that are theorized to be part of it. Much of the residential location research to date has primarily focused on two such variables, travel and housing characteristics. This focus is likely the result of the classic economic residential location models by Alonso (1964) and Muth (1969), models that represent residential location as a consumer allocation problem. In short, the models are based on the assumption that households have a fixed annual income which they use to buy land (a residence), travel (based mainly on commute distance), and other commodities. This constrained utility maximization process is generally described in terms of the relationship between rent or housing prices and distance from the central business district (CBD). It is assumed that households balance residential consumption (i.e., housing prices that decrease with increased distance from the CBD) with commuting distance (i.e., travel costs that increase with increased distance from the CBD). Lerman (1975) and Shin (1985) drew on this classic framework by placing great emphasis on travel and housing explanatory variables (such as mode to work, commute travel time, and type of dwelling unit) in their disaggregate logit models of residential location.

Prevedouros (1992) extended the types of variables studied in residential choice research by including personality characteristics in the modeling of location decisions. Interestingly, no one study was found that rigorously accounted for the many different variables that have been hypothesized to affect residential choice. In particular, attitudes and lifestyle dimensions were generally not included in the reviewed residential choice studies.

To achieve the goals of depth and breadth in the review of the residential choice literature for this dissertation, two different approaches were taken. Studies that highlighted concepts and methodologies that are most directly pertinent to this dissertation are discussed first. Further, the reviews of these papers contain more reflective evaluation and discussion (depth). The second approach provides additional breadth, by briefly summarizing numerous other papers.

2.2 Extended Review of Selected Key Studies

This section highlights some of the landmark studies in this area of research, as well as some lesser-known, but germane work. Each of the studies described below is different in empirical methodology and/or theoretical assumptions. Advantages and disadvantages of these studies may be used to improve future work on residential choice modeling. By including additional important variables (such as attitudinal variables), and more realistic structures and assumptions, the models developed in this dissertation benefit from this earlier work while extending it.

2.2.1 An Econometric Model of Urban Residential and Travel Behavior

Kain (1962a), in his flagship study of urban transportation and land-use planning, uses an econometric framework to model the residential and tripmaking behavior of workers in

Detroit. He uses survey data from 40,000 households to estimate the relationships among residential space consumption, auto ownership, choice of transportation mode, and length of journey-to-work. Employing least-squares multiple-regression techniques, Kain estimates a system of 9 equations that represent the relationships among the above 4 listed variables.

Kain assumes a sequential or *causal* ordering for his nine equations. He assumes that a worker: 1) first decides what amount of housing s/he wants, 2) then decides whether to buy an auto, 3) then determines to drive or to use transit, and 4) that the first three decisions determine the mean length of his or her commute. This structural equation approach to investigating the relationships among residential choice and travel produced reasonable results, such as: the percentage of people living in two-family units increased when commute distance increased, and zones with lower housing costs and larger household sizes tended to have a higher proportion of single-family units.

Though this study showed how an economic framework could shed light on the interdependencies between travel and housing consumption, it did not provide insight into individuals' travel and housing choice behavior due to its use of aggregate data in the model estimations.

2.2.2 Location and Land Use

In his book Location and Land Use (1964), Alonso established the foundations of a mathematical model of urban land use. Alonso assumes a single central business district (CBD) that is the location of all employment, and that households reside on a infinite plain surrounding it. Further, the travel time per unit of distance is the same in every direction on the plain.

Each household consumes an amount of housing, goods, and travel time to work that maximizes its utility function. As households must determine how much of their budgets to allocate to the consumption of each item, this problem can be viewed as a constrained maximization problem. An important conclusion from Alonso's bid rent theory is that housing price decreases as the distance to the CBD increases. This consumer behavior theory of utility maximization is seen in a majority of the related studies following Alonso's.

2.2.3 A Disaggregate Behavioral Model of Urban Mobility Decisions

One of the first efforts at modeling residential choice at a disaggregate level was undertaken by Louviere (1979). He developed a theory on the individual preference for single-family dwelling units by looking at residential "bundles" as potential choice alternatives for 185 Wyoming residents living near the University of Wyoming, in the city of Laramie (sample size for experiments 1, 2, and 3 are 35, 50, and 100 respectively). Subjects rated a number of hypothetical residential bundles, with an example of one such bundle being a single family home with monthly payments of $400, 20 blocks from work, 34 blocks from the central business district, 9 blocks from neighborhood shopping, 15 blocks from schools, 3 bedrooms, 2 baths, a garage, conventional construction, and new paint and fencing. Louviere interviewed realtors and home developers to determine the eleven factors (the neighborhood and dwelling unit characteristics having the greatest influence on individual home selections) that are seen above in the bundle. These eleven factors were used as explanatory variables in regression and

ANOVA models to determine what most influenced individuals' residential choice. Louviere's models showed that his study sample was most concerned with housing cost, quality of neighborhood (a categorical variable defined by Louviere in his stated preference survey; 1 = below average quality, 2 = average quality, and 3 = above average quality), and square footage in the decision to purchase a single-family home.

Louviere conducted a near identical experiment to the one described above (experiment 3), with the main difference being the use of 13 factors (as opposed to 11) determined by a different set of realtors and developers. Both experiments provided the following two conclusions: 1) personal characteristics such as gender and age greatly affected the model coefficients for the variables describing preference, and 2) individuals are not likely to share the same utility function.

2.2.4 Residential Mobility of Job Changers

Verster (1985), in an effort to include an influence often left out of studies of spatial and travel interaction, considers the effect that job location has on moving behavior and residential choice. The analysis is based on survey responses from about 4000 people from the Randstad region of the western Netherlands, where each person was defined as the head of the household and categorized by one of the following four groups: 1) moved to a new residence and took new job, 2) moved to a new residence and kept old job, 3) stayed at old residence and took new job, and 4) neither moved nor changed jobs. Verster notes that as "the places where individuals settle for living and working are the spatial framework for their other daily activities"

(p. 195), any changes in residence or employment location for an individual will create large impacts on his/her lives, and consequently, the choice to change both job and residence will likely not be made simultaneously. That observation is less likely to hold for larger countries such as the United States, in which a job change often necessitates a location change. However, the degree to which this observation holds for the United States (and other countries) may be changing with the rise in two-career households imposing more geographic constraints than before.

Verster makes some important points about causality in modeling. First, it is noted that even though the sample was stratified into four different types of mobility groups, it is possible that any of these groups of decision makers will not have a "homogeneous preference structure". Next, Verster provides the following statistics about the potential direction of causality between residence and work changes: "Of the respondents whose latest residential move was associated with a change of work, 93 percent when asked about the causality answered that the house move was a result of the job change (in only two percent of the cases the causality was the other way about)" (pg. 196). Taken alone this can be a powerful reason for creating a model structure that has job location as a cause of residential choice - at least for the "head of the household". However, Verster points out that the literature on residential mobility and migration "gives conflicting results for some variables" (pg. 198). For example, coefficient signs and magnitudes for the variables household size and household income have been found to be inconsistent across various studies, potentially giving credence to the idea of different populations having different preference structures. Further, while the concept of "head of household" has fallen into disuse with the increase of women entering professional careers, it is still quite possible that in a two-worker household, a residential change may be prompted by a job change for one worker, while the converse is true for the other worker.

A multinomial logit model structure was used to estimate the residential-location choice of people who changed jobs and moved. To create a relevant and tractable set of residential alternatives to be modeled, each worker's choice set was assumed to contain the 20 municipalities nearest to his or her employment zone (where the study area comprised 25 employment zones and 160 municipalities). Verster drew two main conclusions from the results: 1) travel cost variables (such as commuting distance) were significant in models of residential mobility (probability of moving) and residential choice (destination), and 2) the magnitude of the impact of changes in travel costs on an individual's behavior was positively associated with the travel costs to which one had been accustomed. The second conclusion is interesting in that it supports the hypothesis that experience influences attitude, and illustrates the value of having longitudinal data to investigate such effects.

2.2.5 Causal Analysis of Trip-Chaining Behavior

Kitamura *et al.* (1990) studied a large sample of commuters from an urban region in Japan (Osaka) in an effort to understand their trip chaining behavior. Survey data from 7,611 commuters that made exactly one non-work stop (such as a trip to a store or restaurant) on work days were studied to determine the relationship between commuting distance and non-work stop location, and the impact of travel time on activity duration. Two specific goals of the project were to identify the effect of time-space constraints on the spatial distribution of non-work stops and to investigate the causal mechanism underlying commuters' trip chaining behavior. This last objective is particularly relevant to this dissertation's task of exploring the causal mechanisms in an individual's residential location choice behavior.

Spatial distributions of home, work, and activity (non-work stop) locations were analyzed to achieve the first goal. Primary findings from this analysis included: 1) non-work stop locations were most frequent in places near home and work, with the greatest concentration of trip stops being near work; 2) commuting distance was the principal determinant influencing the choice and duration of non-work stops by commuters; and 3) additional time spent traveling to reach a non-work activity (stop) was positively correlated with the time spent participating in the non-work activity (stop).

A path analysis was conducted with log-linear models based on contingency tables to reach the second aim of finding a hierarchical relationship among the following trip-chaining factors (all in categorical data form): commuting distance, added travel distance (the extra distance traveled to get to a non-work stop), travel mode to reach non-work stop, time spent at non-work stop, and total travel time of trip-chaining event. Three types of log-linear causal models were estimated. First, 24 five-tier linear hierarchical models were developed with the assumption that commuting distance was the first factor (the "primordial" factor) in the hierarchy. This was the simplest model design in that it didn't allow for simultaneous interaction among the factors, and thus, it assumed that one factor was chosen and then another until all five factor choices were made. The second and third model types allowed for factors to be in the same level (or tier), where two factors could influence each other but not have any hierarchical relationship. Specifically, model type two was a four-tier linear hierarchical model where the first or last choice was composed of two interacting factors (three independent factors in a chain preceded or followed by two interacting factors for a total of 4 tiers), and model type three was a three-tier hierarchical model where the first or last two tiers were composed of two sets of two interacting factors and the other tier was an independent single factor. Results from the best models (selected as "best" based on interpretability and chi-square values) in the first group were used to design the more sophisticated model type two and three structures. The major conclusions drawn from these models were that several causal structures fit the observations equally well (i.e., gave similar chi-square statistics), and that trip-chaining decision processes can vary across individuals, making it necessary for travel behavior researchers to design models that accommodate heterogeneity in individual decision structures.

This study has useful implications for residential choice research. By empirically demonstrating that there is more than one decision structure underlying trip chaining behavior, it suggests the possibility that there will also be more than one causal structure pertaining to residential choice behavior. The authors' conclusions that travel behavior analysis should be set up to handle "heterogeneity in decision structures" indicates the need for future studies to include the development of multiple conceptual and analytical frameworks that can accommodate the potentially varying causal structures of individuals' residential choice behavior.

2.2.6 Modeling the Choice of Residential Location

McFadden's (1978) study of residential location was pioneering in that he incorporated economic choice behavior into an empirically feasible discrete choice model of housing location. He first describes choice with a multinomial logit model, having individual dwellings as the alternatives. McFadden then assumes that dwellings with the same observed attributes will be perceived as the same, leading to a multinomial logit model with dwelling types as the choice alternatives (thus reducing the number of alternatives greatly), the preference among which is influenced by other attributes. By relaxing the assumption of independence of the error terms (representing unobserved variables in the utility functions for each alternative), he permits a structure of similarities among alternatives. Further, McFadden derives the family of generalized extreme value models using stochastic utility maximization. He goes on to show how the nested model is a special case of the generalized extreme value model. In essence, this paper opened the door for more applications of residential choice modeling with logit models (see for example, Quigley, 1985; Shin, 1985; and Cho, 1997).

2.2.7 A Two-Stage Housing Choice Forecasting Model

Tu and Goldfinch (1996) offer a recent example of a housing choice study. Their modeling methodology was defined by two independent stages: estimating the likelihood that a household would choose a particular housing sub-market bundle (i.e., neighborhood and dwelling type), and estimating the likelihood that a household would choose a bundle of housing components such as central heating and gardens (i.e., dwelling unit choice). In the first stage, the households were segmented into specific socioeconomic classes, including single young-person households and households with dependent children. Tu and Goldfinch assumed that all households within a particular socioeconomic group had the same housing choice behavior if they had the same income. This assumption would be more credible if the attitudes and lifestyles of households within the same socioeconomic group were the same (which is of course not the case). Consequently, their study would be improved with the inclusion of explanatory variables relating to household attitudes and interests.

On the other hand, Tu and Goldfinch argued that households in different socioeconomic groups have different decision-making processes. Thus, in some instances a simultaneous choice of neighborhood and house might be the best representation of the decision making process, whereas in other instances those two decisions might be made sequentially. Tu and Goldfinch assumed that households were rational, meaning that they would gain market information (i.e., learn about the available neighborhoods and dwelling units) and choose dwellings that would maximize their residential utility. These ideas and logic provided a foundation on which their two-stage housing choice model was built.

Tu and Goldfinch utilized major contributions from two of the authors mentioned earlier in this chapter. First, similar to Louviere, Tu and Goldfinch viewed the household residential choice decision making structure as a process whereby households select the neighborhood and dwelling unit "bundle" that maximizes their consumption utility function. Both neighborhood bundle components (such as school quality, transport connections, and safety) and dwelling unit bundle components (such as housing size, quality, and age) were used as explanatory variables in their housing sub-market models. Next, using McFadden's residential-choice framework, the authors chose a multinomial logit (MNL) model specification to determine the probability that an alternative (whether a neighborhood or dwelling unit) will be chosen by a household. Tu and Goldfinch addressed the issue of independence from irrelevant alternatives (IIA), and stated that the error terms are not independent because, for the dwelling unit model, the "independent variables include neighborhood components shared by dwellings in the same neighborhood" (pg. 525), and similarly, neighborhoods in the neighborhood model may share unobserved characteristics common to a larger region containing each neighborhood. They still chose to use the multinomial logit form and argued (without substantiation) that the "IIA axiom may not be violated if a buyer makes the choice after obtaining full market information" (pg. 525). Further, they suggested that the good model fits (each rho-squared value was above 0.70) obtained using MNL on their data implied that any bias due to IIA was insignificant. However, this author is not aware of any theoretical work linking model goodness of fit to lack of violation of IIA (other than the general guideline that improving the model specification may reduce or remove the IIA violation). Thus, it may be possible to obtain an apparently strong goodness-of-fit on a biased model.

Data collected from the 1972 Lothian Region (Scotland) Household Housing Survey were used to estimate the two independent sets of multinomial logit models. Specifically, models of sub-market and dwelling unit choice were estimated on three different socioeconomic groups: single young-person households (N=125), young-couple households (N=154), and households with dependent children (N=329). The dependent variable for the sub-market models had 63 potential alternatives, based on 7 neighborhoods and 9 housing types (such as "flat with less than or equal to 2 bedrooms" and "terraced house with 3 or 4 bedrooms") within each neighborhood. Independent variables for the sub-market models included the neighborhood average dwelling-unit size, the neighborhood average dwelling-unit age, distance to work (for household head), and school quality. Findings from the first-stage model included: 1) distance to work was a significant variable for all the socioeconomic groups, but least significant with young-couple households (possibly due to dual-earner households' difficulty with optimizing commute length), 2) young-person households preferred to live near shopping areas, and 3) households with dependent children strongly preferred neighborhoods with good schools.

Whereas the first-stage models were calibrated on revealed preferences, the second-stage models were stated preference (actually a transformation of ranked importances to presumed preferences). The dependent variable for the second stage (i.e., dwelling-unit choice) had eight alternatives, based on the absence or presence of three characteristics: large kitchen (1 if yes, 0 if not), central heating (1 if yes, 0 if not), and private garden (1 if yes, 0 if not). The independent variables were three binary variables representing presence of the three characteristics listed above (e.g., large kitchen), and one income variable. The income variable was the only nonsignificant variable in the model, suggesting that a household's budget constraint did not impact the choice of characteristics like kitchen size in the final housing choice. Lastly, the authors concluded that researchers using cross-sectional data should do modeling based on subgroups.

2.2.8 Personality Characteristics and Residence Location Decisions

Though a relatively recent study, Prevedouros (1992) gives the best example (to the author's knowledge) of the use of attitudinal variables in models of residence location decisions. Using the Minnesota Multiphasic Personality Inventory (MMPI) test to measure the personality of 1300 mail-back survey respondents in Chicago, Prevedouros investigated the relationships among personality characteristics (sociability, materialism, and suburbanism) and location selection, automobile ownership, and travel characteristics. Factor analysis was used to obtain the personality factors, which were then used in a cluster analysis to obtain personality types. It was found that "the suburbanism dimension of a respondent's personality has an association with the type of suburb where a household decides to relocate" (pg. 385): specifically, a respondent with a high, positive factor score was more likely to choose a low-density suburb than a high-density suburb. Further, it was concluded that "certain personality characteristics correlate well with residence location, automobile ownership, and daily travel characteristics" (pg. 391). In summary, Prevedouros effectively used attitudinal variables to gain a deeper understanding of an individual's transport and residence location behavior.

2.3 Brief Review of Other Relevant Studies

This section presents the second approach to the travel behavior/land use and spatial interaction literature review, a brief discussion of a more comprehensive set of studies. Research from many different disciplines, including economics, geography, sociology, transportation engineering, and urban planning was explored for findings related to residential preference/choice. A table format is used to allow a large amount of information to be presented in a space-efficient manner.

Specifically, Table 1 contains summaries of references that were helpful in identifying potential dependent and independent variables needed to operationalize the conceptual model. It is important to note that many of the studies involving traditional neighborhood design focused on its relationship with travel behavior, and not on residential choice. However, the studies are still useful in that they present many findings on the relationship of socio-demographic and travel-related variables to traditional neighborhood (land use) development. Many of these studies are rigorous and complex, often having multiple research objectives and modeling methodologies that could not be fully described in a few paragraphs (e.g., Lerman, 1975). Consequently, the synopses below contain descriptions of only select areas of each study, the areas that were most pertinent to this dissertation.

Before summarizing each paper, an **"Overview**" description is given that provides the statistical basis and research focus of the paper as it relates to the present study. Included in the overview description will be study type (quantitative or qualitative, and empirical, theoretical, or conceptual), statistical method (such as regression), and study focus (such as residential preference/choice or travel demand). This will allow a quick inspection of the varying range of methods and goals contained in this diverse set of literature.

Table 1: A Selective Summary of References Supporting the Conceptual Model

REFERENCE	STUDY DESCRIPTION
Aldana *et al*. (1973)	**Overview:** a quantitative, empirical study of travel demand using factor analysis, analysis of covariance, and discriminant analysis. In this study, models of travel demand for Boston, Massachusetts residents were developed that take "into account the simultaneous and interdependent character of decisions about travel, location, and automobile ownership." Data from more than 38,000 households, 117,000 persons, and 300,000 trips were used in estimating the models. Factor analysis was used on these data to obtain different market segments for separate model estimation, such as a model estimated with data solely on young bachelors and a model estimated on blue-collar workers. Twelve demographic groups were defined along three characteristics: 1) residential location (central city and suburbia), 2) automobile

ownership (none, one, and two or more), and 3) use of public transportation (consistent transit user or non-transit user). Models predicting number of individual trips were estimated by standard analysis of covariance techniques using data from each group. Exogenous variables in the models included household income, binary variables for working outside of residential zone and spouse working, and walking time to transit. A 2-group discriminant technique was used to develop a mobility model of transportation choice (number of cars to own and frequency of transit usage) conditional on location. In the first model, the probability that a household uses transit given residential location and automobile ownership was estimated. In the second model, the probability that a household owns 0, 1, and 2+ cars given residential location was estimated. The computation of joint probabilities of using or not using transit and owning 0, 1, or 2+ cars was possible using the results from the two separate models. Using all of the results together the authors were able to calculate the "probabilities of using or not using transit conditional on location, but unconditional on the choice of automobile ownership level"[pg.8]. Conclusions of the study included: 1) life-cycle stage and socio-demographic variables were "extremely" useful for segmenting samples for use in disaggregate modeling, and 2) dual-earner households were more likely to use transit.

Ben-Akiva and Bowman (1998)

Overview: a quantitative, empirical study of residential location using nested logit models.

A discrete choice model system that incorporated activity and travel schedules into a household's residential location choice was the focus of this paper. An activity-based travel demand model system previously developed by the authors was summarized. A residential accessibility measure was described, with accessibility being defined as "the expected value of the individual's maximum utility among the activity schedules available, given a residential location" (pg. 1133). The inclusion of this accessibility measure was a key step in the estimation of their integrated residential choice model system. In short, the system was "linked as a sequentially estimated nested logit system, with conditional models supplying expected maximum utility, or logsum variables, to the higher level models." The highest level model was the residential location model, which incorporated results (such as measures of accessibility derived from the activity-based travel demand model system) from the lower-level models (such as a daily activity model). Data used in the empirical work came from a 1991 24-hour household travel diary survey from the Boston metropolitan area (N=1259) and descriptive statistics (such as crime rate and residential density) from 787 geographic zones that comprised Boston's metropolitan area. The dependent variable for the residential choice model comprised the 787 geographic zone alternatives. To simplify the modeling process, the authors estimated model parameters with "a sample of eight alternatives, drawn by stratified importance sampling, for each household." Explanatory variables used in the model included: 1) accessibility measures (used in the activity-based model system) such as "expected utility of the daily activity schedule, given the daily activity pattern and work location, households with one adult, a worker", 2) residential density, and 3) annual income. Findings include: accessibility to non-work activities had a strong influence on households' residential choices, and residential density was an important factor in households' residential choices. This study was unique in the attention that was given to non-work related activities. The activity-based model system provided accessibility measures that were based largely on non-work trips.

Boarnet and Sarmiento (1998)

Overview: a quantitative, empirical study of travel demand using an ordered probit model.

This study focused on the link between land use at the neighborhood level and individual non-work travel demand. The authors stated that "almost all recent empirical work on land use and travel behavior has lacked a clear behavioral framework" (pg. 1155). Data from a 1993 two-day travel diary for 769 southern California residents were used to estimate an ordered probit model. The dependent variable for the model was the number of non-work automobile trips made by an individual during the two-day survey period. Model explanatory variables consisted of sociodemographic variables (such as gender, race, and household income) and land-use variables (such as population density, percentage of grid-like street pattern, and service employment divided by land area). Model findings included: 1) age (negative coefficient) and female dummy variable (positive coefficient) were significant explanatory variables at the 5 percent level, and 2) no land-use variable was statistically significant. New land-use variables that "control for co-variance between residential location choice and unobserved aspects of trip generation behavior" were developed using an instrumental variable technique (i.e., substituting variables such as "proportion of housing built before 1960" that were correlated with land use, but unrelated to trip generation behavior). The new land-use variable for service employment (previously found to be insignificant) was significant in a similar model of non-work automobile trips. The authors noted the importance of "both controlling for residential location choice and using different levels of geographic detail when studying the link between land use and travel behavior" (pg. 1166).

Boehm and Ihlanfeldt (1991)

Overview: a quantitative, empirical study of residential preference/choice using probit models.

The primary goals of this study were to identify the neighborhood attributes that were most important to households and how neighborhood preferences varied among different types of households. The data used came from household records of the 1985 American Housing Survey (a sample that included over 40,000 households). Sixteen household types were defined by race (black, white), income (high, low), dwelling unit (multifamily, single family), and location (central city, suburbs). An extensive set of explanatory variables (such as quality of streets and availability of open space) hypothesized to affect satisfaction with neighborhood were developed and incorporated into 16 probit models, one model for each household type. The dependent variable modeled was individual neighborhood satisfaction (where each individual response appeared in the market segment model to which his or her household belongs), a value between 0 (lowest satisfaction) and 10 (highest satisfaction) that was interpreted as an "ordinal utility index." Empirical results include: 1) open space was significant for all of the high-income, white households (but not for any of the other households), suggesting that this group prefers neighborhoods with lower population densities, and 2) variables representing crime rates, noise levels, and building appearance were significant in all of the models across the 16 household types.

Carnahan *et al*. (1974)

Overview: a qualitative, conceptual review and analysis of population density and behavior.

A detailed review of the literature on the determinants and consequences of population density was presented. Population density was described as a "composite of several different measures of land use", including: population per room, dwelling size, number of dwellings per structure, number of structures per residential area, and percentage of area used for residential land. Two very different neighborhoods can have the same population density due to the many different components defining this measure. It was noted that changes in transportation (such as a new transit line or more car usage) had impacted the formation of densities in urban areas. In general, higher auto ownership levels were associated with lower population density areas. However, a contradicting trend was also indicated, where new concentrations of density were occurring in suburban communities due to changes in workplace locations and the development of new popular structures such as high-density condominiums. In terms of density's impact on residential behavior, the authors point out the potential attractions of both high and low density: 1) high-density areas were more likely to contain cultural and artistic opportunities important to some people, and 2) low-density areas had more open space such as parks that may appeal to other people. An important hypothesis of the authors was that little of the variance in the behavior of people (such as residential and mode choice) could be "attributed to density independent of other social structural variables" (pg. 502) (such as attitudinal variables).

Cervero (1996a)

Overview: a quantitative, empirical study of travel demand using linear regression.

The impact of neo-traditional neighborhood design on commuting was the focus of this study. A regression model of transit modal share was estimated using 1990 US Census tract data from four San Francisco Bay Area counties. Individual commute data from people living in both transit- and auto-oriented neighborhoods (N= 898) was combined with the census tract data. Explanatory variables used included density (households/acre), income (natural logarithm of household income), and neighborhood type (transit oriented or automobile oriented). Major findings included: 1) increases in neighborhood density corresponded to increases in transit mode share, and 2) income was negatively associated with transit mode share.

Cho (1997)

Overview: a quantitative, empirical study of the joint choice of tenure and dwelling type using multinomial logit models.

This study contained a housing choice model where it was assumed that households chose type of tenure (rent or own) and type of dwelling unit (single-detached dwelling or multiple-unit dwelling) simultaneously. It was assumed that households followed the principle of consumer behavior theory (i.e., a household examines all of the available alternatives and chooses the one that maximizes its housing utility under a budget constraint). Data collected in 1993 on the choices of tenure and housing type for 785 households residing in the city of Chongju, Korea were used to estimate multinomial logit models of the joint choice of tenure and dwelling unit type. The dependent variable consisted of the following four alternatives: owner-occupied detached dwelling

(OD), owner-occupied multiple dwelling (OM), rented detached dwelling (RD), and rented multiple dwelling (RM). Independent variables included occupational type, housing price divided by household income (a proxy for budget constraint), age and education of household head, and children (binary variable equal to 1 if school-age children were present, 0 if not). In addition to the full model, separate models were estimated on two subsamples of the data, respondents who lived in a "high-quality neighborhood" (rating of 8 points or higher; N=547) and respondents who lived in a "low-quality neighborhood" (rating of 7 points or lower; N=238). The quality of the neighborhood was determined by an environmental quality index, where respondents rated their neighborhood on three areas (5 possible points per area for an upper bound of 15 points): satisfaction with public transport, environmental pollution, and public services. Conclusions included: 1) the inclusion of neighborhood attributes into residential choice modeling was valuable, 2) older respondents were more likely to choose owner-occupied detached residences, and 3) the presence of children, though significant in the nonsegmented model, was not significantly associated with housing choice in the models segmented by neighborhood quality.

Crane (1996)

Overview: a qualitative, conceptual study of the impacts of land use on travel demand.

This paper contains a behavioral model of travel demand (in terms of frequency and length of trips) as a function of neighborhood design (in terms of accessibility). A literature review on the transportation benefits of neotraditional design, as well as a discussion of weaknesses found in these studies, is given to motivate the new behavioral model developed. Specifically, a framework for measuring the changes in individual travel behavior due to improvements in community access from changes in land use is presented and discussed in terms of neotraditional design and travel demand. An important conclusion was that the transportation benefits of neotraditional and transit-oriented designs may be oversold, because the reduction in miles traveled due to shorter trips may be outweighed by an increase in trip frequency.

Dussault (1997)

Overview: a qualitative, empirical study of residential choice using personal interviews.

Real-estate agents in the Sacramento, California region were interviewed on the topic of "best places to live". Neighborhood characteristics (such as quality of local schools, crime rate, and distance to shopping) and dwelling unit characteristics (such as design of home and back yard size) that the agents saw as most important to new home-buyers were reviewed. The anecdotal evidence presented suggested that variables such as school quality and safety were key factors in a household's residential choice.

Frank and Pivo (1994)

Overview: a quantitative, empirical study of travel demand using linear regression and correlation analysis.

Empirical models testing the impacts of land-use mix, population density and employment density on travel behavior were developed from a diverse set of data sources, including: Puget Sound Transportation Panel, U.S. Census Bureau, and Washington State Department of Economic Security (sample analyzed included 1,680 households and 28,955 trips). Dependent variables included percent of trips by single-occupant vehicle and transit. Census tract data were used to develop urban form variables. Findings include: 1) higher population densities were positively correlated with the percentages of individual person trips that were transit and walk based, and 2) mixing of land uses was negatively correlated with the percentage of individual person trips that was single-occupant-vehicle based.

Friedman *et al.* (1994)

Overview: a quantitative, empirical study of travel demand using comparative trip frequencies.
The authors use data from a 1980 regional travel survey of San Francisco Bay Area households to investigate the impact of neotraditional design on travel (N= 672: 450 suburban households and 222 traditional households). Specifically, they define particular communities as being "traditional" (based on a set of criteria, including "had an interconnecting street grid and residential neighborhoods in close proximity to nonresidential land uses") and others as being suburban, and then compare the trip rates between the two types of communities. Statistical analysis only included frequencies of types of trips (such as work-based transit trips and home-based-nonwork-auto trips). The findings indicated that community design had a significant impact on travel behavior. For example, respondents from traditional communities took a greater number of walk and transit trips than respondents from suburban communities.

Galster and Hesser (1981)

Overview: a quantitative, empirical study of residential preference/choice using two-stage least-squares structural equation modeling.
A model of residential satisfaction was defined and estimated in this study. A path analysis approach was used with the following three components: 1) objective independent variables based on neighborhood (such as residential density), respondent (such as marital status) and dwelling (such as number of bathrooms), 2) subjective intervening variables such as "rundown" (respondent's perception of neighborhood deterioration) and "in common" (respondent's measured feeling of community and belonging in a neighborhood), and 3) subjective dependent variables defined by neighborhood satisfaction, housing-quality satisfaction, housing-quantity satisfaction, and housing-type satisfaction. Many hypotheses were formed among the three components, where the direction of causality goes from the objective independent variables to the subjective intervening variables, and then to the subjective dependent variables. An example hypothesis was: residential density (objective variable) will influence a respondent's perception of how much she or he will have in common with other residents (intervening variable), and that in turn, will influence a respondent's neighborhood satisfaction (dependent variable). The data used in the model was taken from a household survey of 767 residences in Wooster, Ohio, in which respondents' perceptions of neighborhood and dwelling unit characteristics were measured along with their socio-demographic characteristics. Similar path models were estimated based on segmentations of the sample (such as a model estimated solely on married couples). All model coefficients were calculated using two-stage least-squares structural equation techniques. Model results indicated that "certain types of people (younger, married, female heads, blacks, those with many children) consistently evidence less satisfaction for any given

residential context due to different needs" (pg. 752), suggesting that different household types will have different residential choice decision structures.

Gilbert and Foerster (1977)	**Overview:** a quantitative, empirical study of mode choice using attitudinal variables and regression analysis. A review of literature on attitudinal variables and transportation created the foundation for this study's investigation of the importance of attitudinal variables in models of mode choice. It was noted that the rationale behind using subjective data (e.g., attitudes) was that it can improve the "correspondence between the individual decision processes and the theoretical relationships used in transport planning" (pg. 322). Attitudes were defined as individuals' ratings on attributes "for which the underlying physical or economic attributes are not obvious." Data collected from an attitudinal survey of Chapel Hill (North Carolina) residents were used to estimate regression models of mode choice (bus or car). Only respondents who had a choice between the two modes were analyzed (N = 144) so that the attitudinal data on both modes would not be biased by situational constraints. Explanatory variables used in the models included both objective measures (such as travel time and travel costs) and subjective measures (such as attitudes towards cars and buses). Models containing attitudinal and objective variables had significantly more explanatory power than models with objective variables alone. The major conclusion of the paper was that attitudes and perceptions can be important explanatory variables in empirical models.
Hamilton and Roell (1982)	**Overview:** a quantitative, empirical study of urban commutes. The focus of this paper was the examination of the impact a city's housing supply and job locations had on its citizens' average commute length. A review of commuting theoretical models assuming centralized employment and decentralized employment was given. Estimation of the average commute that would take place "if all households were to choose jobs and homes so as to minimize the sum of commuting plus land rent" was done for several U.S. cities, and then compared to actual city average commute lengths. It was found that actual commute lengths were much greater than the theoretically optimal commute lengths, indicating that substantial gains were possible through residential relocation. Many reasons were noted for the "wasteful" commuting, including relocation costs exceeding the value of commute reduction savings, and dual-earner households not being able to optimize commutes (as they were not always on the same "ray from the CBD"). These findings pointed to the need for a workplace location model to be an integrated part of any commute model system.
Heikkila *et al.* (1989)	**Overview:** a quantitative, empirical study of residential location land values using a regression model. The main hypothesis of this study was that residential land values are impacted by a residence's accessibility to multiple "subcenters" (i.e., employment centers and recreation centers). One example given was the multiple-worker household, where a residential location "offering accessibility to many employment nodes" would have increased value as household workers would have a greater chance to find work at a reasonable distance from home. Property-specific data on 10,928 households from Los Angeles was combined with census-tract data from the 1980 Population Census to estimate a

regression model of residential land value. The dependent variable was housing price divided by lot size, and independent variables included age of home, number of bathrooms, distance to central business district, distance to Glendale, and distance to ocean. The focus on accessibility variables in this study was unique. The variable distance to ocean was viewed as a proxy for "air quality as well as a measure of accessibility to the dominant recreational resource of the region" (pg. 225). Model results included: 1) neighborhood income (median household income of residents) had a significant and positive coefficient, and 2) distance to central business district was the least significant of all the accessibility variables. It was concluded that accessibility to non-work activities may play an increasingly important role in future residential land value studies.

Hensher and Taylor (1983)

Overview: a quantitative, empirical study of residential relocation using multinomial logit models.

This study looks at students' residential relocation decisions. Multinomial logit models of residential relocation choice (dependent variable alternatives: moved once, moved twice, and did not move) were based on data from a sample of 62 students (61.3% had moved). Significant variables in the model included travel time to college before the move (where higher travel times meant a greater probability of relocating) and car ownership (where having a car increased the probability of relocating).

Horowitz (1995)

Overview: a quantitative, empirical study of the joint choice of residential location and commute mode using a multinomial logit model.

A joint residential choice and travel mode choice model was theoretically developed and empirically estimated in this study. Three data sources were used: 1) the 1968 Washington, D.C., area transportation survey of 30,000 households (providing data on travel behavior and socioeconomic characteristics), 2) the 1970 U.S. census (providing information on neighborhood attributes of census tracts of the households surveyed, such as average housing costs and residential densities), and 3) estimates of tract-to-tract travel times and costs by car and by bus. A multinomial logit specification was used with the dependent variable defined by census tract ("a geographical unit that has an area of about 0.25 square mile and that may contain several hundred to over 1,000 dwelling units", pg. 209) and commute travel mode (bus or car). The total number of census tract/mode combinations was not specified. However, it is likely that the model was estimated on a subset of alternatives containing the chosen combination and a randomly selected set of possible additional combinations per person. Explanatory variables included in the model were of three types: transportation (such as commute time), household (such as natural log of income), and location (such as residential density of tract and school quality in tract). It was found that improvements in transit had large effects on mode choice, but not on residential location choice.

Hunt *et al.* (1994)

Overview: a quantitative, empirical study of residential preference using logit models.

The authors developed stated response surveys of residential location preference by defining 48 different houses described by monthly cost, number of bedrooms, travel time to work and shopping, and proximity to rail. Residents (N=390) of Calgary, Canada, who participated in the 1992 survey were shown a subset of 4 of these houses (with the subsets randomly selected for each respondent and collectively including all 48 alternatives across the sample) and ranked them from best to worst. Further, socioeconomic and household characteristic data were collected on the respondents. Logit models of residential choice (where the dependent variable consisted of the rankings of the 4 housing scenarios given by each respondent) were estimated using the exploded logit technique. This technique was chosen because it can "predict the full ranking of the alternatives in an observation - in contrast to the more limited prediction of the single, most-preferred alternative in standard logit analysis" (pg. 82). The validity of the stated preference method was discussed. Significant explanatory variables included cost divided by household income and travel time to shopping.

Joseph *et al.* (1989)

Overview: a quantitative, empirical study of residential preference using conjoint measurement (i.e., measures of utility).
Understanding the preferences of home buyers for rural residences was the focus of this paper. Conjoint measurement was used to analyze data from 22 potential home buyers in Ontario, Canada. As people may develop preferences for a residence based on the joint effect of two or more variables (such as house size and price), conjoint measurement (a method "concerned with the joint effect of two or more independent variables on the ordering of a dependent variable", pg. 48) was chosen to measure residential preference. Six attributes for rural residential preference were defined: lot size (small to large), house size (3 bedrooms to 5+ bedrooms), location (isolated, scattered, or village), price of home ($75,000 to $150,000+), school bus (on route or off route), and roads (paved or unpaved). Respondents evaluated hypothetical choice alternatives defined by varying levels of the above six attributes. Based on these evaluations, combined utilities for the attributes were calculated via the conjoint analysis method, allowing an overall measure of an individual's utility for a particular residential alternative to be determined. Further, an examination of trade-offs between attributes was performed. Two findings from the study were: 1) trade-offs between servicing and other residential attributes was common (e.g., "respondents would rather live on a small private lot with full services [such as paved roads and bus service] at a lower price, than pay $125,000 or more to live on a larger, equally private lot with less services", pg. 60), and 2) rural home buyers value privacy the most, pointing to a "behavioral push for low-density residential development in the countryside" (pg. 61).

Kitamura *et al.* (1997)

Overview: a quantitative, empirical study of travel demand using regression models and factor analysis.
The impact of land use and attitudinal orientations on travel behavior were investigated in this study. Respondents (N= 1,380) from five neighborhoods in the San Francisco Bay Area, California region, were surveyed on many aspects of urban life. The neighborhoods varied on several dimensions, including residential density, transit access, and sidewalk and bike trail availability. Data on attitudes toward topics like transit and the environment were obtained and factor analyzed to obtain explanatory variables for use in regression models of travel behavior. Further, land use variables such as population density and

level of mixed land use were significant independent variables in models of travel behavior, for which dependent variables included: person trips (number), transit trips (number and fraction), and non-motorized trips (number and fraction). Important findings included: 1) land use and attitudinal variables contributed significantly to model explanatory power, and 2) "attitudes are certainly more strongly, and perhaps more directly, associated with travel than are land use characteristics" (pg. 156).

Lansing and Marans (1969)

Overview: a quantitative, empirical study of residential preference using frequencies and correlation analysis.

In an effort to understand better the attributes that were important in individuals' evaluations of their neighborhoods (defined as the "immediate vicinity of their homes"), the authors of this study obtained survey information from home interviews (N=1042) of residents in the Detroit, Michigan region and compared it to information obtained from a professional planner's evaluation. Specifically, each household was asked to evaluate the immediate area in which they lived on 2 measures of environmental quality: 1) overall evaluation of neighborhood (like it very much to dislike it), and 2) attractiveness of neighborhood (very attractive to unattractive). The planner rated 99 neighborhoods (corresponding to the neighborhoods of about one third of the respondents) on the degree of openness (open, neutral, or enclosed), the degree of pleasantness (pleasant, neutral, or unpleasant), and degree of interest (interesting, neutral, or dull). A correlation analysis indicated that there was a positive relation between the households' ratings on attractiveness and the planner's rating on openness, and a negative relation between the households' ratings on attractiveness and the planner's rating on interest. Households found to "like their neighborhood very much" reported that appearance ("well kept up"), noise level ("quiet"), and low density ("families not too close together") were important features of their neighborhood evaluation.

Lerman (1975)

Overview: a quantitative, empirical study of the joint choice of residential location, housing type, auto ownership and mode type using multinomial logit models.

A behavioral representation of the household mobility decision was the framework for this comprehensive study. A three-stage hierarchy of choice was assumed, where a household first had an employment location, and based on this location, a bundle of mobility decisions were made (residential choice, housing, auto ownership, and commute mode choice) that impacted short term travel choices (such as frequency of trips and time of day). Data from the 1968 Washington, D.C. Home Interview Survey were used to estimate multinomial logit models of choice. The model specifications were based on the assumption that households were faced with a joint choice structure, in which each combination of location (nine different census tracts), housing type (single family, walk-up apartment, and high rise), ownership level (zero, one, and two+ autos), and mode to work (car and transit) represented a possible alternative for the dependent variable. Explanatory variables used in model development included household size, residential density, total two-way commute time, and per pupil school expenditure. Further, models segmented by number of household workers (one versus two or more) were developed to test the hypothesis that the decision making process for these two types of households would be very different. Coefficient estimates for all of the models estimated had the expected signs (positive or negative) and reasonable magnitudes, providing support for the model specifications.

Lindstrom (1997)

Overview: a qualitative, empirical study of residential choice using frequencies and responses from an open-ended interview.
Investigating the importance households placed on social relations and solidarity was the focus of this work. The residential decisions of 50 randomly selected households in two affluent but dissimilar neighborhoods on Chicago's North Shore (Evanston and Wilmette) were analyzed via an open-ended home interview and questionnaire. Evanston was described as a neighborhood with racial diversity and an "urban ambience", while Wilmette was described as a low-density suburb occupied primarily by white professionals. Both neighborhoods were considered to be part of the prestigious Chicago North Shore, indicating a high average income and education of its respondents. There were no significant differences between the respondents of each neighborhood in terms of life cycle, age, and number of children. The similar characteristics of the respondents allowed the author to focus on the "social and cultural factors in the demand for specific housing styles, lot sizes, density ratios, and neighborhood ambience" (pg. 22). It was found that respondents who chose homes in Evanston placed great value on diversity, with one respondent quoted: "We wanted Evanston. It was the type of community we had been accustomed to - semiurban, culturally and racially mixed." Respondents who chose homes in Wilmette noted that good schools and a small town atmosphere were very important. One respondent said: "We like the look of Wilmette, the age of the homes, size of trees, the brick streets. There is a sense of community here and of families" (pg. 33). Though the impact of factors such as commute distance and transit accessibility on residential decisions was not investigated, the large influence that a sense of community had on household residential choices was made clear.

Los and Nguyen (1983)

Overview: a qualitative, theoretical model of the joint choice of residential location/type and mode.
A joint model of residential choice (defined by housing type and location zone) and mode choice (defined by automobile or bus) was formulated and solved as a convex programming problem. The authors used the bid rent concept of Alonso to determine residential household preferences (given that the level of household utility varies with housing type and location zone) and assumed that job location was determined prior to the choice of a residence. Other sophisticated mathematical techniques were employed to establish a model algorithm that may be used in applications such as estimating the impact of transportation investment on residential location. It was noted that a substantial amount of data would be required to calibrate their model, something that may prevent its use by practitioners in the near future.

Louviere (1979)

Overview: a quantitative, empirical study of residential preference using analysis of variance and regression models.
Two experiments (similar in design and purpose with different sample sizes, 35 and 50 respectively) were conducted in which potential home buyers from

Laramie, Wyoming were asked to evaluate descriptions of residences in terms of how well each residence met their needs if "they were in the market for a home." The evaluation scores (1= not well to 20 = perfect match) for each residence were then used as the dependent variable in one-way ANOVAs on each of eleven factors related to residential preference (such as distance to work and landscaping). Mean evaluation scores differed significantly (at the 0.05 level) across the factor levels of all 11 of the factors, with landscaping and garage factors accounting for the most variation in mean scores. A regression analysis was also performed with the eleven factors as explanatory variables and the dependent variable being the residential evaluation scores. All of the accessibility explanatory variables in the regression model (such as distance to work, and distance to neighborhood shopping) had negative coefficients, indicating that the utility of a residential location decreased as the distance from it to other places of interest increased. [See also Section 2.2.3]

Louviere and Timmermans (1990)

Overview: a quantitative, empirical study of residential preference using linear regression.

The hierarchical information integration (HII) approach to studying complex decision making was applied to residential preference in this paper. Briefly defined, HII is a theory of how individuals "process information in a hierarchical manner in decision-making situations that involve many attributes" [pg. 130]. Specifically, individuals are hypothesized to categorize the characteristics that influence residential preference into subsets that represent decision constructs. To operationalize this modeling approach, 315 people in the Roermond region of the Netherlands who had recently moved were surveyed with respect to their residential preference decision. In the design of the HII experiments, four constructs influencing residential preference were defined: house (such as size of backyard and number of rooms), residential environment (such as amount of greenery and number of children's playgrounds), relative location (such as distance to work and distance to neighborhood shopping), and social and economic (such as presence of friends and family near new residence). Respondents rated scenarios from each of these four constructs (e.g., there were 16 house attribute descriptions varying by number of rooms, cost, design, and size, and 8 social and economic descriptions varying by the number of friends and family in the area, workplace location, and previous residence) with values ranging from 0 (the worst) to 10 (excellent). The ratings were assumed to be cardinal measurements, and consequently, regression analyses were used to statistically describe a subject's evaluations (i.e., the dependent variable was the value placed on the particular construct scenario). Explanatory variables for each of the four regression models were attributes that defined the four constructs of residential location preference, such as number of rooms (for the house regression model) and amount of greenery (for the residential environment regression model). Findings included: 1) privacy and view were the most significant variables in the residential environment model, 2) distance to work was the most significant variable for the relative location model, and 3) house and residential environment constructs had the greatest influence on residential preference.

Lu (1998)

Overview: a quantitative, empirical study of migration decisionmaking using logit models.

A conceptual model of migration decisionmaking based on behavioral theories from social psychology was developed and empirically estimated in this paper. It was hypothesized that "perceptions of, and attitudes toward, housing and neighborhood" were key to modeling migration decisionmaking. Data from the 1985, 1987, and 1989 national core and supplement files of the American Housing Survey on the mobility decisions, sociodemographic characteristics, and housing characteristics of more than 135,000 people were used to estimate logit models of mobility behavior. Mobility models were estimated (sample sizes ranging from 33,198 to 35,726) on five mutually exclusive household types such as single person and married couple without children. The dependent variable was equal to one if the household had moved within two years (i.e., 1985-1987), and equal to zero if no move occurred. Explanatory variables (defined for the conditions prior to the move) included neighborhood satisfaction, household type (such as single-person household), education, and log of income. Conclusions included: 1) attitudinal variables (such as neighborhood satisfaction) were important factors in a household's migration decision, and 2) sociodemographic variables such as age and income had "significant direct effects on migration over and above their indirect effects channeled through attitudinal variables" (pg. 1492).

Madden (1981) **Overview:** a quantitative, empirical model of commute length using regression analysis.

In this paper, gender differences in commute length were investigated by "measuring the effects of household characteristics and job characteristics on the work-trip behavior of men and women in different categories of households." A theoretical equation for commute length was determined mathematically through the manipulation of a model of one-earner household utility (where household utility was defined as a function of housing consumed, other goods consumed, and leisure time). An important assumption of the utility model was that commute length was impacted by both household residential choice and workplace choice. Regression models, segmented by socio-demographic and life-cycle variables (such as presence of children, number of workers, and marital status), were estimated using data from the 1976 Panel Survey of Income Dynamics (data collected from over 5,000 families at the University of Michigan, and filtered to contain people who were considered to have "urban commuting decisions"; sample sizes ranged from 124 to 166 cases). The dependent variable for the models was the natural logarithm of commute length, and explanatory variables included logarithm of weekly work hours, distance of home from city center, logarithm of spouse's wages, and logarithm of number of children. Model results indicated that wages and distance from city center were positively associated with commute length, while population size was negatively associated. A major conclusion of the paper was that women have shorter commute lengths due to two main reasons (pg. 193): 1) "women's lower wage rates and shorter work hours reduce the earnings return to their commute", and 2) "household responsibilities increase the cost of longer commutes."

Menchik (1972) **Overview:** a quantitative, empirical residential preference/choice study using correlation analysis.

An investigation of preference and choice of residential environments for individuals was the basis of this study. Home interviews of residents (N = 457) in five areas of Pennsylvania (ranging in description from rural to dense city) were conducted to obtain data on the characteristics important to them in their decision to choose a residence. Based on the following open-ended interview question, "When you were trying to decide on a place to live what were all of the characteristics that were important to you?", four preference dimensions were identified. Each dimension comprised many attributes, such as ruralness for dimension 1 (natural environmental beauty), population density for dimension 2 (man-made environment), access to work and shopping for dimension 3 (accessibility), and lot and house size for dimension 4 (housing characteristics). Four preference variables were developed by "dividing the number of responses falling into each category (dimension) by the total number of responses made by that respondent." The four preference variables were: 1) preference for a natural environment, 2) preference for uncongested surroundings, 3) preference for high accessibility, and 4) preference for a large house lot. Choice variables matching the categories of the preference variables were also developed (using individual data from respondent surveys), such as "choice of accessibility to work" being defined by the negative logarithm of commute time given by respondent. It was noted that a weakness of the study was the quality of developed choice variables due to data limitations. Empirical analysis and discussion was primarily based on product-moment correlations between the choice and preference variables. Second-order correlation coefficients indicated that individuals' preferences of residence generally matched their choices.

Onaka (1983)

Overview: a quantitative, empirical residential mobility study using logistic regression models.

This study contained a model of residential mobility, with an emphasis on the relationship between the socioeconomic characteristics of a household and housing preference. Before and after data from 1667 rental households (randomly sampled within the 48 contiguous states in the USA, from the 1966 Survey of Economic Opportunity conducted by the US Bureau of the Census) were used in logistic regression models, where the dependent variable was mobility behavior (i.e., moved within the year or not) and the independent variables were household characteristics (such as household size and age of head of household) and dwelling unit characteristics (such as number of rooms and repair needs). The explanatory variables were based on data obtained in the before survey, since the before conditions are the ones on which a decision to move is based. An expected finding was the interaction effect of household size and number of rooms, where the "marginal utility of an additional room is greater for the larger household than for the small" (pg. 764). Further, there was a positive relationship between housing size and the probability of moving.

Prevedouros (1992)

Overview: a quantitative, empirical residential choice and travel demand study using factor analysis, analysis of variance and regression models.

The data used in the empirical analysis in this paper was from a mailback household survey of 1,300 Chicago suburb residents who had moved between 1987 and 1989. An exploratory analysis of variance (ANOVA), with the dependent variable being location decision ["reside in a low-density suburb", "reside in a high-density suburb", "relocated into a low-density suburb" (presumably from a high-density suburb), and "relocated to a high-density suburb" (presumably from a low-density suburb)] and five explanatory variables including income and number of household workers, was conducted in this study. Only 17% of the variance was explained in this ANOVA, indicating a need to include other factors like personality to improve understanding of the location decision. A factor analysis on personality attributes was performed, and three personality factors were obtained: 1) sociability, 2) materialism, and 3) suburbanism. Various travel-related regression models were estimated using the personality factors as explanatory variables. For example, the dependent variable, non-work trip distance by auto (miles), was regressed on the sociability factor, yielding an R-squared of 0.24. One conclusion of the paper was that "some degree of association does exist between the personality of the respondents and the type of suburb chosen for the residence location of their household" (pg. 386). [See also Section 2.2.8]

Quigley (1985)

Overview: a quantitative, empirical residential choice and public services study using a nested logit model structure.

Consumer choice models of dwelling unit, neighborhood, and public services were presented in this study. Household choice was estimated via a nested choice model in three stages: 1) the choice of dwelling unit given neighborhood and town, 2) the choice of neighborhood given town, and 3) the choice of town, characterized as bundles of services and amenities (such as public expenditures). The models were developed based upon data from 584 "recent mover rental households in the Pittsburgh metropolitan area" that were randomly taken from a Southwestern Pennsylvania Regional Planning Commission home interview survey conducted in 1967 of nearly 25,000 households. The dependent variable for dwelling unit choice had 20 alternatives, differentiated by two measures of residential density (i.e., either a single attached dwelling or a duplex), two measures of quality of dwellings (i.e., structural condition and age), and cost (i.e., monthly income of respondent minus rental payments). Independent variables for this stage included number of baths and bedrooms per person. The dependent variable for neighborhood had 4 alternatives, namely 4 census tracts with data on characteristics such as median rent of dwellings and fraction of households that are white. Independent variables for this stage included respondent's monthly commute in hours by auto and by transit, and median rent. There were two alternatives for the town dependent variable (not a specific town, but a larger census tract region equivalent to two census tracts/two neighborhoods), with census tract/town data including the amount of expenditures spent on school and public services. Independent variables for this model included proportion of black students in schools and transit travel time. Important empirical findings included: measures of accessibility of a neighborhood were significant (e.g., auto commute time was significant in the choice of a neighborhood), and the independence of irrelevant alternatives assumption among dwelling unit, neighborhood, and town alternatives would have been inappropriate (i.e., a nested choice model was necessary).

Rutherford *et al.* (1996)

Overview: a quantitative, empirical travel demand study using cross-tabulations.

The travel pattern differences between people who live in mixed-use (neotraditional) neighborhoods and people who live in neighborhoods "with more homogeneous land use patterns" were explored in this paper. Data from a two-day travel diary and demographic survey of 900 households in three greater Seattle, Washington neighborhoods (characterized by two or more land uses) were combined with a similar rich data set collected two years earlier by the Puget Sound Regional Council (PSRC) to allow travel patterns to be compared between different neighborhoods. Conclusions included: 1) that people living in neotraditional neighborhoods (such as Queen Anne and Wallingford) traveled fewer person-miles than people living in nonmixed-use neighborhoods (such as the suburban Kirkland), and 2) "mixed land uses have the same effect (reduced travel) on weekend trips as [on] weekday trips" (pg. 55).

Ryan and McNally (1995)

Overview: a qualitative, conceptual review of the impact of neighborhood design on travel demand.
This study provided a historical context for the neotraditional design movement. Design aspects of such neighborhoods were presented along with its political and social background. A discussion of the transportation impacts of neotraditional design was given. No empirical models were developed, but a solid review of the present understanding of neotraditional design was presented. It is noted that the challenge for a neotraditional designer is to create a community that is "held together by the human element" (i.e., to include characteristics such as walkable streets and civic centers), but that is also capable of handling households' increasing demands for "convenient travel". Also discussed was the fact that research on neotraditional neighborhood design impacts is still "in its infancy", and that many questions about the viability of this design still remain.

Salomon and Ben-Akiva (1983)

Overview: a quantitative, empirical travel demand study using cluster analysis and logit models.
The importance of lifestyle variables was a major focus of this study. It was suggested that the concept of lifestyle may contribute to differentiating between population groups on such choices as residential location and travel mode. Cluster analysis was used to identify lifestyle groups, including "family oriented economically active" and "young family oriented childbearing household". Logit models based on data from each group (individually and pooled) were estimated where the dependent variable had 20 alternatives based on different travel destinations (such as the central business district and a respondent's residential zone) and mode choices (auto, transit, and walk). Model findings showed that weights of "certain attributes of the utility functions" (i.e., the explanatory variables such as "in-vehicle travel time for auto" and "out-of-pocket travel costs, all modes") significantly varied (by sign and magnitude) across lifestyle groups, indicating that lifestyle impacts travel demand. For example, a lifestyle-group characterized by "lower-income, older" people had the lowest probability of choosing an auto for a trip.

Shin (1985)

Overview: a quantitative, empirical study of the sequential choice of residential location, housing type, auto ownership, and mode to work using a nested logit model structure.
A sequential household decision-making structure (as opposed to an independent or joint structure) was investigated in this study for the choices of

residential location, housing type, auto ownership, and mode to work. As the number of alternatives and attributes in residential choice processes is vast, a nested logit model structure was employed that simplified the complex multi-stage choices into a "tree-like decision making process." Data from the 1981 San Francisco Bay Area Home Interview Survey was used in the model estimations. An important characteristic of the sequential structure in a nested logit model is that "the lower level of the decision [in this case the lowest level is mode to work] is iteratively incorporated, through the composite form of utility, into the higher level(s) of decision throughout each stage of the sequential model structure." In the first stage, the mode to work model dependent variable was defined by three alternatives (drive-alone, shared-ride, and transit) and had explanatory variables such as in-vehicle travel time and number of autos per worker. The dependent variable in the second stage model of auto ownership had three alternatives (own zero, one, and two or more cars) and explanatory variables such as household size and automobile cost divided by annual income. The next stage's dependent variable was identified by three housing types (single-unit, townhouse-unit, and multi-unit) and included explanatory variables such as "marital status for a single-unit dummy variable" and "household head age for townhouse-unit dummy variable". The last stage of the sequential logit model contained a dependent variable for residential location choice defined by "eleven hypothetical centroids of cities in Santa Clara County (California)" and explanatory variables such as crime rate and residential density. The sequential model structure was used to develop models for both single- and dual-income households. Findings from the models indicated that single- and dual-worker households had different decision processes in choosing residential locations.

Simpson (1987)

Overview: a quantitative, empirical study of the joint choice of workplace location, residential location, and urban commuting distance using two-stage least-squares techniques.

This study presented an empirical and conceptual analysis of urban spatial structure, emphasizing the roles of workplace location and residential location. It was noted that "a model of workplace location decisions is needed to interact with the Alonso model", as the Alonso model greatly underpredicted commute distance. Two causal equations were developed, in which the dependent variable of one was an explanatory variable in the other. The first equation described a model explaining the distance from the city center to the workplace, with explanatory variables that included distance from the city center to the place of residence, job skill level, and local employment conditions. The second equation described a model predicting the distance from the city center to the place of residence, with explanatory variables that included distance from the city center to the workplace, number of workers in the household, and household income. It was hypothesized that the residential location and workplace location decisions were simultaneously determined, and consequently, two-stage least squares estimation was employed. Model estimation was completed with data from the 1979 Metro Toronto Travel Survey (N=3508). Segmented models were also estimated on the following subsamples: heads of households, nonheads, owners, and renters. One finding was that "the responsiveness of workplace location (CCJOB) to residential location (CCHOME) in the workplace location equation is greater for nonheads and owners than for heads and renters" (pg. 124).

Skaburskis (1997)

Overview: a quantitative, empirical study of gender differences in housing demand using logit and regression models.

The impact of gender on household formation, tenure choice, housing consumption, and location preference was examined in this paper. The increase in women's income generation and women's expanded role in the workforce were the main trends evaluated in terms of housing demand. A key belief expressed by the author was that "changes in women's opportunities and outlooks may be among the most important new determinants of urban form" (pg. 275). Data from the 1991 Census Public Use Micro Data Files (over 90,000 households from Toronto and Vancouver, Canada) were used to estimate housing demand models for each major household type, including single-parent households and couple-households with dependents. Sample sizes varied in each model, but were never below 30,000 cases. Four types of housing consumption models were developed using the following dependent variables: home ownership (yes or no), condominium ownership (yes or no), amount of money spent on rental housing (rent/income), and home value level (mortgage payment/income). Explanatory variables used included personal income, ethnicity, household size, and age. Findings included: 1) single women and single men had a similar probability of owning a home, 2) women single parents were less likely to be home owners, and 3) women with low income tended to spend the highest proportion of income on rent. In addition, logit models of location preference were estimated using the following dependent variables: central city (yes or no), and suburban house (yes or no). Independent variables used included children (present or not), income, and other language (equal to 1 if respondent's home language was not English or French). Findings included: 1) women tended to prefer a central location more than men with similar characteristics, and 2) higher-income couples tended to prefer to locate in the central city (contrary to what was expected). It was hypothesized that the demand for suburban homes by nonfamily households will decrease as the income difference between men and women decreases. Also, it was hypothesized that reduction in family formation (e.g., less married couples and fewer kids) may reduce demand for suburban locations in the future.

Steiner (1994)

Overview: a qualitative, conceptual study of residential density and travel patterns.

The author analyzed several sets of literature (such as travel patterns and density, travel patterns and socioeconomic characteristics, location theory and residential choice) to give increased insight into the interactions among residential choice, land-use characteristics of an area, and transportation choices of households. Many important concepts and variables from these literature streams were presented. In addition, a proposal for future research into areas like life cycle characteristics and other location factors that may be important in a household's residential choice was presented. Two weaknesses of past residential choice studies pointed out by the author were: 1) a failure to look at the trade-offs made in dual-earner households, and 2) a failure "to address the importance of non-work locations (e.g., schools and personal services) in the decision about where to move" (pg. 41).

Stopher and Ergun

Overview: a quantitative, empirical study of travel demand using

(1982) analysis of variance, factor analysis, and logit models.

Determining the demand for recreational and cultural activities by urban residents was the focus of this study. The data were obtained from mail-out surveys of residents (N = 638) from two suburbs (statistically different in socioeconomic traits) of Chicago. Information collected included perceptions of recreation activities, frequency of participation in recreation activities, and socioeconomic characteristics. Logit models of choice were estimated with the dependent variable consisting of the frequency with which a respondent participated in various recreation activities ("with the logit model predicting the probability that a given activity will be chosen on one occasion" [pg. 27] given that an activity occurs) and explanatory variables that included distance from trip origin to activity site and escapism (a factor score representing "the ability to get away from day-to-day demands and pressures"). Location (suburb 1 and suburb 2) and various other variables (such as income and education) were explored for potential model segmentation through an analysis of variance procedure (ANOVA). The results from the ANOVA, a two-way analysis of variance on choice (of recreation activity in terms of frequency of participation) versus the "various socioeconomic and situational variables available for segmentation", indicated that income, location, and attractiveness (respondent ratings of attractiveness of activity) were significant main effects (i.e., could be potentially effective segmentation variables). Segmentation was investigated so that different homogeneous subgroups of the respondent population could be identified, and consequently, models would allow each group to "exhibit different weights" for explanatory variables describing a choice alternative. It was concluded that "there are significant variations in decision making for urban recreation activities between people living in different locations of an urban area" (pg. 33). This pointed to the potential problems of transferability of results from these models to other geographic regions. Further, joint segmentation on location and other socio-demographic variables showed improved statistical performance of the choice models estimated. [See also Section 3.4.1]

Tardiff (1977) **Overview:** a quantitative, empirical study of causality between travel demand and attitudes using direct and two-stage least squares models.

Investigating the directions of causality between attitudes and behavior for travel (e.g., attitudes toward travel being determinants of travel behavior, or vice versa) was the focus of this study. Data from a 1973 study of transportation attitudes and behavior of 211 residents of West Los Angeles, California were used to create modal and attitudinal models, separately and simultaneously. The dependent variable for the modal model was a dummy variable equal to one if the bus mode was chosen and equal to zero otherwise. The dependent variable for the attitudinal model was a measure of an individual's perceptions of the importance of eight modal attributes of a car versus a bus (a quasi-continuous variable ranging from -32 for a pro-car person to +32 for a pro-transit person). First, models describing each direction of causality were estimated separately using direct least squares. Only five different variables were included in the models: one attitudinal variable ("comparative satisfaction with the bus and car") and four non-attitudinal variables, including distance to bus stop and auto availability. Using the outputs of this first set of models (estimates of the original endogenous variables) as instrumental variables for the endogenous right-hand-side variables in the simultaneous model allowed consistent coefficient estimates to be obtained. Results indicated that "behavior causes attitudinal response" (pg. 401). However, the author cautions that the value of the models "lie[s] not in their ability to definitively select an appropriate causal mechanism, but in the fact that they suggest that more attention should be paid to the uses and implications of attitudinal variables in transportation models" (pg. 403). [See Chapter 8 for more discussion on causality]

Timmermans *et al*. (1992)

Overview: a quantitative, empirical study of residential and employment preference/choice using regression and multinomial logit models.

A decompositional choice model was used in this study of residential- and employment-choice behavior of dual earner households. It was argued that residential preferences cannot be determined by inspecting residential choices, as "real-world choices do not necessarily reflect individual preferences" (pg. 517). To overcome this, the authors developed hypothetical residential scenarios (defined by different characteristics of possible residential environments and job situations) that were rated by study participants (N = 187 dual-earner couples from the Netherlands). Variables defining residential environments included type of dwelling (e.g., apartment or house), cost (rent or mortgage) and size of area (e.g., 20,000 inhabitants in uptown residential area or 250,000 inhabitants in a city center). Variables defining job situations included distance to work, monthly income, and flexibility of work schedule. The two partners in each household were first asked to separately rate each of 32 residential scenarios (where each scenario comprised a bundle of particular job and residential environment characteristics) on its job profile, its residential profile, and overall (on a scale from 1 = bad to 10 = great). The sample average overall rating for each of the 32 scenarios was regressed against the average job and residential ratings. Both variables were highly significant in explaining the overall rating, but the residential environment coefficient was the largest (0.58 versus 0.51). Next, the two partners in each household were presented with 32 new pairs of residential scenarios, where this time each scenario was characterized in terms of four variables: each partner's hypothetical rating on each of the two characteristics, job and residential environment. Each couple jointly chose "one alternative from each pair that would reflect their choice process in the real world." A multinomial-logit model was estimated "to link preference ratings to subsequent joint choice behavior." Empirical findings included: 1) residential environment attributes were important in forming preferences, but job considerations had a greater influence in the joint decision-making process, and 2) the choice behavior of households with children showed that the residential environment was a bigger factor for them than job issues.

To *et al*. (1983)

Overview: a quantitative, empirical study of residential location using an alternative specification of the Alonso-Muth model.

In an effort to understand why higher-income households were moving to suburbs in North American cities, residential choice models based on the Alonso-Muth model (i.e., a model which measures a household's utility as a function of a composite good: quantity of housing and travel) were developed. Data from a 1972 survey of nearly 1500 households in Montreal, Canada were used in the model estimations (where coefficients were estimated with ordinary least squares). Two dependent variables were tested, both being measures of housing quantity. Principal-components analysis was used to generate a linear combination of housing attributes (such as number of rooms, number of toilets, and heating control) as the first measure of housing "quantity", whereas the second measure was simply the number of rooms in the dwelling unit. Independent variables for the models included household income and distance to work. It was concluded that the "number of rooms" variable was an appropriate measure of the quantity of housing. Model results confirmed the authors' expectations, that "the quantity of housing consumed increased as income and distance increased" (pg. 346).

Tu and Goldfinch (1996)

Overview: a quantitative, empirical study of residential choice using multinomial logit models.

Data collected from the 1972 Lothian Region (Scotland) Household Housing Survey were used to estimate multinomial logit models of residential choice. The modeling was structured as a two stage process in which households first chose a housing sub-market (a neighborhood and housing type) and then a dwelling unit (defined by characteristics such as kitchen size and private garden availability). Models at each stage were estimated on three different socioeconomic groups: single young-person households (N=125), young-couple households (N=154), and households with dependent children (N=329). The dependent variable for the sub-market models had 63 potential alternatives, based on 7 neighborhoods and 9 housing types such as "flat with less than or equal to 2 bedrooms" and "terraced house with 3 or 4 bedrooms". Independent variables included average dwelling-unit size, average dwelling-unit age, distance to work (for household head), and school quality. The dependent variable for the second stage had eight binary alternatives, based on the absence or presence of three characteristics: large kitchen, central heating, and private garden. The independent variables were three binary variables representing presence of the three characteristics listed above (e.g., large kitchen). Findings from the first-stage model included: 1) young-person households preferred to live near shopping areas, and 2) households with dependent children strongly preferred neighborhoods with good schools. [See also Section 2.2.7]

Waddell (1993)

Overview: a quantitative, empirical study of residential and workplace choice using a nested logit model structure.

Testing the assumption that workplace choice is exogenous in determining residential location choices of households was a major objective of this study. It was hypothesized that the "degree to which residence location is driven by workplace location, or the converse, may vary with the degree to which workplace locations are dispersed in a multinodal city, as well as by individuals' household relationship, tenure, ethnicity, and socioeconomic status." Data on the 650 census tracts in the Dallas-Fort Worth SMSA and data from the 1980 Census sample of housing units (filtered to allow only one-worker households) within the Dallas-Fort Worth area were used to estimate joint choice and nested logit models (N = 16,000). The dependent variable in the joint logit specification (a multinomial logit structure) is the "joint probability that a worker will choose a particular combination of residence, workplace, and housing tenure". Each worker had eight alternatives in his/her choice set, based on two possible workplaces (i.e., two different tracts in which her/his job could be located) , two tenure types (rent or own), and two possible residences (i.e, two different tracts in which he/she could reside). A two-level nested logit structure was tested, with the dependent variable for the first nest (top level) being "the probability that a worker will choose a particular workplace", and the dependent variable for the second nest being defined as the "joint probability that a worker will choose a particular combination of residence and tenure, given the prior choice of workplace." The same independent variables (such as travel time to work, population and employment densities of a residence tract, and income of worker) were incorporated into each of the three model specifications. Important hypotheses in residential choice research were supported by the empirical results, including: 1) households "clustered" in residential areas by socioeconomic status, stage of life cycle, and race and ethnicity, and 2) residence and workplace are better modeled as jointly determined (as opposed to the major urban residential-location models like Alonso's that assumed workplace to be exogenous in determining residential location).

Webber (1983)

Overview: a qualitative, conceptual study of the role of life cycle in a household's intraurban mobility.

The significance of life cycle changes as a facilitator of residential mobility was investigated in this paper. A review of theoretical and empirical studies of household migration was given that sets the foundation for the author's conceptual development of a life cycle matrix that can be incorporated into urban models of migration. The number, age, gender, and relations between household members are the primary characteristics that define "the stage of a household in the life cycle". Though no empirical work was presented, various hypotheses based on the life cycle theoretical model were presented, including: "the propensity of households to migrate depends on their stage in the life cycle", and "the consumption (of housing, goods) of households depends on their stage in the life cycle."

Weisbrod *et al*. (1980)

Overview: a quantitative, empirical study of residential choice using a multinomial logit model structure.

In an effort to understand the tradeoffs between transportation and other factors like job location and neighborhood safety in households' residential location decisions, data from 791 households in the Minneapolis/St. Paul, Minnesota metropolitan area (in 1970) were used to estimate three multinomial logit models, representing: 1) moving and tenure type joint choice, 2) residential location, housing type and auto ownership joint choice (given moving and tenure type choices), and 3) mode to work travel choices (given location choice). The three models were integrated through a recursive structure, in which "the estimation of each step depends on the expected utility of subsequent choices." The dependent variable for model one (i.e., the moving and tenure type joint choice model) had 2373 alternatives, defined by different locations (702 location zones, where data were available for attributes such as median rent, median annual household income, and crime rate) that a respondent can move to (or stay at) and the tenure decision (rent or own). The explanatory variables included household income, number of children, and work-trip access for both the primary and secondary worker. The next model in the sequence had 14,814 alternatives in the dependent variable, differentiated by the residential location (702 possible zones), housing type (such as 4-room

	unit or six-room unit), and auto ownership choice (such as own 1 auto or own 2+ autos). Significant explanatory variables in this model included "teacher/pupil ratio", "housing value/size per income", and "proximity to industrial land". The mode choice model's dependent variable was defined by three alternatives: transit, auto-drive alone, and auto-shared ride. Its independent variables included: in-vehicle travel time and travel cost. A major conclusion was that sociodemographic factors played a much greater role in peoples' choices of residential location than did transportation or other public services.
White (1977)	**Overview:** a qualitative, theoretical model of residential location choice and commuting by gender. This study looked at the economic theory behind commuting distances for men and women. A household utility model was presented and then maximized (subject to budget and time constraints) to allow for certain hypotheses to be tested. For example, manipulation of the model indicated that a "wife's supply of labor increases with the speed of commuting, since the time cost of working declines." A household's valuation of space, time, and accessibility were discussed in terms of their contributions to household utility. An important conclusion of the paper dealt with the complexity of a dual-worker household, where it was noted that "two-worker households can choose for purely rational reasons to locate in cities so that women workers commute shorter distances than men" (pg. 50).
Young (1984)	**Overview:** a quantitative, empirical study of residential choice using an Elimination-by-Aspects model structure. A survey of attitudes and behavior of a sample (N=716) of new residents from 3 suburban areas of Melbourne, Australia provided data that were used to estimate Elimination-by-Aspects (EBA) models of residential location choice. The three areas were purposefully chosen to be very different in physical and social characteristics, but all were "located on the same transport corridor". All respondents were familiar with each of the three study areas, and were interviewed to obtain "measures of their perception of the suitability of each area as a possible residential location". Attributes such as "closeness to friends", "pedestrian safety", and "traffic congestion" were three of twenty measures obtained (all on a 100-point semantic scale). The dependent variable for the models was location choice, having the three different suburban areas as its alternatives. The explanatory variables incorporated into the models were the same 20 attributes (such as "affordability of dwelling unit" and "closeness to parks") that were evaluated for suitability across the three areas by the respondents. Using maximum-likelihood procedures, parameters for the non-compensatory EBA models were estimated (i.e., tolerance values for each of the attributes in terms of their importance to residential location were given). "Schools" (school quality) was found to have the lowest tolerance value, indicating that schools had the greatest impact on the final residential choice in the EBA models.

2.4 Chapter 2 Summary

The first part of Chapter 2 presented a detailed analysis of findings and conclusions from eight papers that contained concepts and methodologies important to this dissertation. The first two papers reviewed were landmark studies, containing the fundamental concepts that form the basis for more advanced residential choice modeling. The paper by Verster presented a way of looking at causality for spatial and travel interaction, with commuting (travel) costs influencing home and work location. Kitamura *et al*.'s causal analysis of trip chaining addressed the concept of causality in transportation and serves as an example to follow in terms of assessing causality in residential choice. The last papers by McFadden and Prevedouros include two major advances to spatial modeling, looking at neighborhood attributes in an economic utility framework and including attitudinal variables to explain residential choice, respectively. The strengths and weaknesses of all of the methodologies and hypotheses seen in these papers were guide posts in the development of the empirical models found in this dissertation.

The second part of Chapter 2, the brief reviews of a more comprehensive set of literature, started with a discussion of the main dimensions seen in the residential preference/choice literature. A short description of the classic economic theory of location choice was given along with example studies. The bulk of the section, however, was composed of a tabular summary (see Table 1) of numerous papers pertinent to this dissertation.

Table 1 is fundamental to the development of the conceptual and empirical models of this dissertation. First, in the next chapter, measures hypothesized to be potential explanatory variables in residential preference/choice model development are categorized by type (such as travel demand variables and neighborhood characteristics), and supported by references to author(s) from Table 1. These potential explanatory variables are considered for inclusion in the conceptual and empirical models built here. Second, the findings from the work discussed in Table 1 provide a foundation for the hypothetical examples that are used to illuminate the conceptual model formulated in Chapter 3.

CHAPTER 3

CONCEPTUAL MODEL OF RESIDENTIAL CHOICE

3.1 Introduction

Findings from the literature review suggest that residential choice modeling requires analysis of many interdependent relationships. Determining which relationships to include in a study of residential location is not simple. One relationship involved with residential location that has been extensively studied is the correlation between travel demand and residential neighborhood (see e.g., Cervero and Radisch, 1996). On the other hand, another set of relationships that this author hypothesizes to be important in a study of residential location, the influence of attitudes and lifestyles on residential choice, has seldom been analyzed. In this chapter a conceptual model of residential choice is presented and discussed. The model is a comprehensive reflection of the relationships supported by the literature and by informed judgement. While there may be variables pertaining to residential choice that are not mentioned in this dissertation, it is believed that the most important variables are included here.

The model is shown in Figure 1 (pg. 59), with the arrows representing hypothesized relationships between the rectangles (representing sets of variables defining a particular category such as neighborhood characteristics) they connect. For example, the hypothesis that a person's residential choice will directly influence his or her travel demand is illustrated by an arrow pointing from the *Residential Preference/Choice* category to the *Travel Demand* category. Numerous categories are interdependent, and consequently, many of the connecting arrows have heads at both ends. Each arrow (head) is numbered, and referred to by that number in the discussion of the corresponding hypothesis (so that later reference to that particular hypothesis is simplified by having the corresponding number placed in the section heading for that hypothesis). For example, arrow number 1 represents the hypothesis that an individual's or household's socio-demographic and life cycle characteristics influence residential preference/choice, while arrow number 2 denotes the reverse relationship, that residential preference/choice has an impact on that individual's or household's socio-demographic and life cycle characteristics (see Sections 3.2 and 3.3).

It is clear that the complexity of the residential choice process makes isolating causality very difficult, and thus, the main goal of the model is to demonstrate the many interconnecting relationships that are important to understanding an individual's or household's residential choice behavior. Ideally, empirical findings based on the conceptual model may help to uncover the relative magnitudes of the influence of each component, and be used to verify or modify the structure of the relationships in Figure 1.

Hypotheses and related examples for each of the conceptual model relationships (such as residential location's influence on travel demand) are presented below. Each rectangle in Figure 1 is actually a generic concept embodying many potential specific measures. For example, *Neighborhood Characteristics* includes attributes such as crime rate, residential density and school quality, all of which are potential explanatory factors in residential choice models. It is important to note that each hypothesis can be viewed in terms of both individual and household residential choice behavior. A mix of individual and household examples is provided in the following sections to illustrate this.

In addition to examples highlighting the conceptual model, selected variables that may influence residential choice, identified from a literature review of travel behavior/land use and spatial interaction studies, are listed in tables following the discussion of various conceptual model components. All of the variables listed are believed by the author to have an impact on an individual's preference for and choice of a residence. Many of them are taken directly from previous models of residential choice, and variables that are not, such as explanatory variables used in empirical models of trip frequency, are nevertheless considered relevant to that choice. These variables guided the selection of both dependent and independent variables in this dissertation's model development. The variables comprising the different dimensions (rectangles) of the conceptual model may be modeled as dependent variables in one hypothesis and as explanatory variables in another hypothesis. For example, household size (one of the many potential variables in the *Socio-Demographics & Life Cycle* category) can be viewed as an explanatory variable in a model of travel demand (e.g., one hypothesis could be that larger households tend to make more vehicle trips), and can be viewed as a dependent variable in a model of socio-demographics and life cycle (e.g., one hypothesis could be that an individual who lives on a farm is more likely to choose a large household size).

Before presenting examples of the conceptual-model hypotheses, it is important to explain further the most important dimension in the model, the residential preference/choice dimension. Figure 1 shows that in the conceptual model developed for this dissertation, residential preference and residential choice have been placed together (i.e., they are in the same rectangle). It is acknowledged that there can be a difference between preference and choice, as a household may have to choose a less-preferred residence due to constraints (e.g., Timmermans *et al*., 1992 and Hunt *et al*., 1994). However, distinguishing between the two is a challenge, requiring a rigorous research design to capture households' decision-making structures (ideally while in the process of choosing a residential location). This type of design allows a researcher to analyze the attributes considered in a household's development of residential preference and then make a comparison to actual residential choice. If there is a difference between the two, then identification of constraints (for example, a better understanding of the household bargaining process) can be attempted. The difficulty of "infer[ring] preferences from overt choice behavior" has moved many researchers to study preference directly (Louviere and Timmermans, 1990, pg. 127); an inspection of Table 1 lends support to this point.

The difficulty of distinguishing between residential preference and residential choice (as noted above) influenced the decision to keep them together in this dissertation's conceptual model. However, some researchers do separate the dimensions, by designing experiments specifically to account for the differences between preference and choice (see, for example, Kroes and Sheldon, 1988). An example of an older study which attempted to distinguish between preference and choice of a residence is Menchik (1972, pg. 145), where the author concluded "that preferences, thus defined, do express themselves to some extent through market choice."

A review of the studies just mentioned shows that researchers can conceptually distinguish preference and choice, but in practice it is seldom done. Many of the same explanatory variables could be used in models of residential preference and residential choice, as preference and choice will be correlated. Further, it has been found that constraints, which are supposed to distinguish choice from preference, can in fact influence preference (see e.g., Mokhtarian and Salomon, 1997). The key dependent variable used in this dissertation's empirical analysis is a measure of residential choice (see Chapter 6), but constraints (e.g., income) are incorporated, as well as variables expected to influence preference (e.g., attitudes and lifestyles). Thus, preference is indirectly modeled.

Based on the above reasoning, preference and choice are considered together in the conceptual model. However, examples given to illustrate various hypotheses will include both preference and choice contexts. Each hypothesis is discussed in turn, in the order of the arrow numbers shown in Figure 1. In some sections, extra discussion will be given to clarify conceptual issues before examples are given. When useful, specific conclusions and references from other studies will be provided.

3.2 Socio-Demographics & Life Cycle -------1-> Residential Preference/Choice

The socio-demographic and life cycle characteristics of an individual have an important role in the development of his or her residential preference/choice. Before giving an example of this relationship, it is important to distinguish life cycle from lifestyle. Simply put, life cycle represents a stage in a person's (or household's) life, such as the child rearing years or retirement period (Nijkamp *et al*., 1993). Lifestyle, on the other hand, may be described based on a person's activity patterns (Salomon and Ben-Akiva, 1983). An individual who is involved in many social activities may be considered to have an outgoing or active lifestyle. Different people at any given life cycle may represent a variety of different lifestyles.

With the above distinctions in mind, two examples of how socio-demographics and life cycle can affect a person's residential preference (1 and 2) and one example of its impact on residential choice are given (3), respectively: 1) high-income households may be more likely to prefer high prestige neighborhoods (i.e., may weight prestige more heavily in their utility function), 2) a parent of young children may desire a neighborhood that is pedestrian friendly and containing a nearby park, and 3) a low-income college student may choose to share a small apartment with others due to monetary constraints instead of living in his or her first preference. It is useful to consider example 2 further, as it provides insight into the subtlety of the difference between preference and choice. In this example, household composition could either be a constraint to choice (i.e., the household "must" live in a neighborhood that is better-suited for children though they may prefer a different neighborhood that is not well-suited for children), or a facilitator to preference (i.e., in an evaluation of two similar neighborhoods, they will prefer the one that is better-suited for children).

Researchers have found that socio-demographic and life cycle characteristics are very important in residential preference/choice. Aldana *et al*. (1973, pgs. 2,4) describe the household as the "basic decision unit" for residential location choice, and note that "social class" and "stage of the life cycle" are key identifiers of households. Webber (1983) presents examples of both socio-demographic and life cycle impacts on residential relocation preference/choice: 1) housing consumption (e.g., the size of the house or apartment) changes with changes in household income levels, 2) the change in the number of children in a household impacts space needs, creating new housing preferences, and 3) retired people are more likely to be able to choose their true preference as constraining factors such as commute distance are not obstacles in their choice of residence. Cho (1997) finds that housing choice is influenced by both socio-demographic and life cycle variables, including education, occupation, and children. Skaburskis (1997) shows that gender is a significant factor in housing demand. Lastly, Nijkamp *et al*. (1993) believe that life cycle is more significant than socio-demographic variables in explaining residential relocation decisions. Taken together, it is evident that socio-demographic and life cycle variables play an important role in an individual's or household's residential preference/choice.

Selected socio-demographic/life cycle variables that may influence the preference and/or choice of residential location are listed in Table 2. This variable table and the others in this chapter are not exhaustive, but rather contain only a selection of the numerous possibilities. As indicated earlier, Table 1 provides descriptions of all the studies referenced in Table 2 and the similar tables that follow.

3.3 Residential Preference/Choice -------2-> Socio-Demographics & Life Cycle

The socio-demographic and life cycle characteristics of a household or individual may be influenced by residential preference/choice. For example, a household may postpone having children due to living in a small dwelling unit. An individual who prefers to live in a high-density area where parking is scarce and transit highly accessible may decide to own fewer or no vehicles. The reverse direction of causality, represented by hypothesis 1, is much more common. In fact, no references investigating this specific hypothesis were found. However, for completeness, this hypothesis and others similarly neglected in the literature are included in the conceptual model (Figure 1).

3.4 Attitudes & Lifestyle -------3-> Residential Preference/Choice

An individual's attitudes and lifestyle can have a powerful effect on his or her residential preference/choice. A household preferring to live near the city center due to an affinity for social and/or cultural activities is such a case. Similarly, a household that has many activity demands (i.e., an active lifestyle) and scarce travel time may prefer a residence that has many desired destinations located near it (to facilitate trip chaining). A person who is very sensitive to traffic congestion is likely to prefer a less congested neighborhood on the urban fringe (but relatively close to work).

Researchers have indicated that lifestyle and attitudes are important in the residential location decision making process. Ben-Akiva and Bowman (1998, pg. 1143) incorporated individual activity and travel schedules (i.e., the trips a person makes) in a residential location choice model, and note that a person's "daily pattern reflects a longer-term lifestyle decision, one that should be integrated at the long-term level with other mobility and lifestyle choice models." Lu (1998) included attitudinal variables representing an individual's satisfaction with her/his neighborhood and dwelling unit in

Table 2: Socio-Demographic & Life Cycle Variables Hypothesized to Influence Residential Preference/Choice

VARIABLE	SELECTED REFERENCE(S)
age of head of household, worker	Webber (1983), Nijkamp (1993)
annual pre-tax income of household	Waddell (1993), Hunt *et al.* (1994)
blue-collar worker (job status)	Aldana *et al*. (1973), Cho (1997)
education level	Stopher and Ergun (1982), Cho (1997)
gender	White (1977), Skaburskis (1997)
housing ownership	Waddell (1993)
housing price divided by housing income	Cho (1997)
income of worker (after tax) per month	Timmermans *et al.* (1992)
marital status	Shin (1985)
number of cars available for use in household	Hunt *et al.* (1994)
number of children	Louviere (1979), Waddell (1993)
number of drivers	Weisbrod *et al*. (1980)
number of licensed drivers in household	Lerman (1975), Hunt *et al.* (1994)
number of people in household	Onaka (1983), Hunt *et al.* (1994)
remaining income (natural log of money left after taxes, home and transportation costs)	Horowitz (1995)
years lived in dwelling unit	Louviere (1979)

models of residential mobility. Prevedouros (1992, pg. 391) concluded that personality characteristics (i.e., developed attitudinal variables like social introversion) “correlate well with residence location selection”.

Thus, some variables related to individual attitudes and lifestyles were found in the travel behavior/land use and spatial interaction literature. Many such variables were found to influence significantly preference and/or choice of residential location. Table 3 contains a listing of selected variables from these studies. A quick inspection of the list shows that some variables, such as “distance to entertainment”, “participation in activities”, and “leisure time” may appear to be endogenous to residential choice, and thus, it is important to discuss the issue of endogeneity here.

It is clear that the distance (from home) to any activity will depend on a person’s residential location (hence the endogeneity), but the more important question is whether or not the actual distance to a location will influence a person’s residential location decision. The same logic applies in varying degrees to many of the rest of the variables (such as participation in activities). The variables listed in Table 3 were used differently, depending on the author and study (see Table 1), but can all be viewed as measures of attitudes and lifestyles that influence residential choice. For example, Young (1984) chose to study the influence of distance to activities on a person’s residential choice. However, instead of measuring the respondents’ actual distances to activities (which would have generated an endogeneity bias), Young asked respondents to rate on a scale of 0 to 100 (100 denoting very important) the importance of living in a residence close to entertainment, close to relatives, and close to parks. These

importance ratings can

Table 3: Attitudes & Lifestyle Variables Hypothesized to Influence Residential Preference/Choice

VARIABLE	SELECTED REFERENCE(S)
affinity for material possessions	Prevedouros (1992)
desire to be near people (yes, no)	Menchik (1972)
family oriented	Salomon and Ben-Akiva (1983)
friendly neighbors nearby	Dussault (1997)
inactive	Salomon and Ben-Akiva (1983)
introversion	Stopher and Ergun (1982), Prevedouros (1992)
leisure time	White (1977)
living close to parks (respondent importance rating)	Young (1984)
living close to entertainment (respondent importance rating)	Young (1984)
living close to golf (home buyer request)	Dussault (1997)
living close to people of same age, social level (respondent importance rating)	Young (1984)
participation in activities (like swimming, cultural events and nature)	Stopher and Ergun (1982)
perception of bus mode of travel	Tardiff (1977)
personal achievement	Stopher and Ergun (1982)
proximity to friends	Hensher and Taylor (1983), Young (1984)
proximity to relatives	Young (1984)

legitimately be viewed as proxies for attitudinal and lifestyle variables in a model of residential choice, and because they are not measures of actual distance, endogeneity is not a concern.

3.5 Residential Preference/Choice --------4-> Attitudes & Lifestyle

"When in Rome, do as the Romans do." Where a person lives can have a direct impact on her or his lifestyle and attitudes (and hence, behavior). For example, a household living in a high-density urban area with increasing crime rates may reorient its leisure activities to spend less time on local outside activities such as walking the dog. Likewise, an individual that lives in an area famous for its lakes and wildlife could start to place a greater value on the environment. The direction of causality represented by this hypothesis, though less common than the direction discussed in Section 3.4, has been investigated by researchers. The study by Stopher and Ergun (1982) is one which examines this hypothesis. A description of this study is given next.

3.5.1 Case Study: Impact of Residential Choice on Lifestyle

Stopher and Ergun (1982) investigated the effect that residential location has on an individual's or household's lifestyle. Using data collected from surveys of residents in two suburbs of Chicago (N = 638), the researchers developed two logit models with activity type as the dependent variable (i.e., the frequency of participation in various activities defined the dependent variable, with the logit model "predicting the probability that a given activity will be chosen on one occasion" [pg. 27]), and explanatory variables

that included personality traits (such as extroversion) and activity measures (such as distance, availability and attractiveness of a list of specific types of activities). The data were segmented by location, with one model estimated on survey information from Des Plaines, Illinois, and one estimated on data from Evanston, Illinois. A major conclusion drawn from the study was that location (i.e., residential choice) played a significant role in explaining the variation in peoples' choices for type and number of urban recreation activities (i.e., lifestyle). For example, it was found that activities that allowed a person to get close to nature (and away from urban life) were significantly more important for respondents in Des Plaines than for respondents in Evanston, while activities that helped a person get away from "day-to-day demands and pressures" were significantly more important for respondents in Evanston.

The authors considered the potential location bias that could result from varying demographics between the two locations by including socioeconomic measures (such as age and education) as segmentation variables. Empirical analysis indicated that segmentation by location gave much better model results than segmentation by socioeconomic variables.

Four of the explanatory variables in the models -- personal achievement, extroversion, ability to get close to nature and escapism -- were derived from a factor analysis of the respondents' perceptions of various activities. Each of these variables was statistically significant in one or more of the choice models estimated in the study, adding important explanatory power over that obtained solely from typical socio-demographic independent variables. The implication is that by identifying and measuring (using factor scores) the key dimensions underlying many types of activities people do, researchers can gain insight into the lifestyle attitudes and behaviors of respondents.

3.6 Socio-Demographics & Life Cycle -------5-> Attitudes & Lifestyle

The socio-demographic characteristics of a household and/or the life cycle stage it is in are hypothesized to impact its attitudes and lifestyle. Two examples are: 1) a couple that has just finished putting their children through college may start a lifestyle of leisure and be less concerned about spending money on pleasure items, and 2) a single mother may be less likely to participate in "night-life" activities (due to family responsibilities/constraints) than would a young, single woman. Stopher and Ergun (1982) found that socio-demographic characteristics (such as age) were useful segmentation variables in their models of activity participation (lifestyle).

It is important to note that a person's socio-demographic characteristics and life cycle are not the only determinants of his/her attitudes and lifestyle. In fact, it is likely that attitudes will differ among people with the same socio-demographic characteristics, indicating that other complex factors are involved.

3.7 Attitudes & Lifestyle -------6-> Socio-Demographics & Life Cycle

A household's attitudes and lifestyle can have an impact on its socio-demographic and life cycle characteristics. For example, an ambitious individual desiring to participate in a career that requires many years of education may postpone entering the workforce and/or getting married. A household that is very environmentally proactive may choose to own fewer than one automobile per licensed driver.

Salomon and Ben-Akiva (1983) found a particular lifestyle group that could be considered “family-focused”, in that these households chose to establish a family with children at the expense of their total household income. In short, one member of these households chose to “participate in lower paying jobs in the labor market” (thus reducing the total household income), probably to balance work and family responsibilities.

3.8 Socio-Demographics & Life Cycle -------7-> Travel Demand

The socio-demographic characteristics and/or life cycle stage a household (or individual) is in are hypothesized to impact its travel demand. For example, an elderly person may take fewer nighttime trips (possibly due to poor vision and/or few evening social activity demands). On the opposite end, a household with several children involved in activities like soccer is likely to make a large number of trips each week. A young, athletic individual may use non-vehicular modes of travel, such as biking and walking, more often than other types of individuals.

Travel demand models containing socio-demographic and life cycle characteristics as explanatory variables are common in the literature. Cervero and Radisch (1996) found that income (“annual salary of respondent”) had a negative coefficient for a binary (1 = transit, 0 = auto) logit model of mode choice (i.e., a higher income would reduce the probability a respondent chooses transit). Madden (1981) concluded that women have shorter commute trips (see also White, 1977). Boarnet and Sarmiento (1998) found that age was negatively associated with number of non-work trips and that the number of children under 16 in a household was positively associated with number of non-work trips.

3.9 Travel Demand -------8-> Socio-Demographics & Life Cycle

An individual's (or household's) demand for travel will influence his or her socio-demographic and life cycle characteristics. One example is a person whose circumstance has recently changed, such as a new requirement to travel to distant locations that are not well-served by transit, resulting in the first-time purchase of a car (or even just the acquisition of a driver's license). A person who takes a lot of trips (e.g., to take care of a sick relative or for business-related needs) may postpone marriage or starting a family until her or his travel demands decrease. Literature investigating this hypothesis was not found, but it is important to note the less obvious relationships as they too may be affecting a household's or individual's decisions.

3.10 Travel Demand -------9-> Attitudes & Lifestyle

A person's travel demand influences her or his attitudes and lifestyle. For instance, a person that frequently uses reliable transit may develop more negative attitudes toward driving (especially in congestion) and strengthen her or his positive perceptions of transit. An individual who works at home (i.e., has no commute travel demand) may be likely to pursue more activities outside of the home, like walking to the park, to avoid "cabin fever". A person who is required to travel a lot (due to work and other obligations) may be more likely to pursue relaxing, at-home activities with his or her free time.

Similar to Section 3.9 (i.e., that travel demand influences socio-demographics and life cycle), the opposite orientation for this hypothesis (i.e., that attitudes and lifestyle impact travel demand) is expected to be more significant. However, research based on the theory that people try to minimize inconsistencies between attitudes and behavior (cognitive dissonance reduction) has indicated that behavior (e.g., travel demand) can impact attitudes. For example, Tardiff (1977), using cross-sectional data, included an explanatory variable on transit usage (equal to one if the respondent used a bus, zero otherwise) in a model developed to explain attitudes toward transportation modes (car and bus). The transit usage variable was significant (t = 5.64) and positive, indicating that respondents who used the bus had higher satisfaction for bus travel (i.e., that behavior impacts attitudes). The reverse relationship was also tested, with the final conclusion being that "relationships between transportation attitudes and behavior [are] more complex than previously hypothesized" (pg. 397). As attitudes are hypothesized to develop over time, longitudinal data would be more suitable for investigating this type of hypothesis.

3.11 Attitudes & Lifestyle -------10-> Travel Demand

An individual's attitudes and lifestyle will impact his or her demand for travel. For example, a person who is uncomfortable around strangers is more likely to drive than use transit to get to places. A household with an active lifestyle is more likely to make a large number of trips than an inactive household.

Many researchers have explored the relationship that attitudes and/or lifestyle have on travel demand. Gilbert and Foerster (1977) found that attitudinal variables (such as a variable based on “I hate to be tied to fixed schedules for traveling”) added significant explanatory power to models of mode choice (specifically, whether or not transit was used). Bentler and Speckart (1981) investigated the causal relationship between attitudes and different types of behavior, concluding that the direction of causality is not unidirectional (i.e., there is a complex interaction and one is not predominantly causing the other). Lifestyle’s role in travel demand has also been explored. For example, Prevedouros (1992) found that individuals who were socially extroverted traveled more than individuals who were socially introverted (in terms of distance by auto for non-work trips).

3.12 Dwelling-Unit Characteristics -------11-> Residential Preference/Choice

A person’s residential preference/choice is heavily influenced by the values he or she places on dwelling-unit characteristics. For instance, a dwelling unit that is structurally sound and clean is more likely to be preferred by a household than a dwelling unit that is poorly kept. A large lot with elegant landscaping may be essential characteristics in a particular household’s residential choice bundle.

Louviere (1979, pg. 374) notes that individuals form their overall preference or utility for a residence based on the sum of “marginal utility values of the attributes of the residential bundle”. Many researchers have recognized the importance of including dwelling-unit characteristics when designing residential bundles to be evaluated by respondents. In a study of residential preference/choice, Menchik (1972) found that dwelling-unit characteristics were important factors in an individual’s residential choice. Respondents in this study noted the following dwelling-unit characteristics as most important: 1) house-design characteristics such as room layout and heating, 2) house quantity, such as the size of rooms and the number of bathrooms, and 3) lot size. Hunt *et al*. (1994) identified cost (rent or mortgage) and size (number of bedrooms) as two of the most important dwelling-unit characteristics in the formation of residential location preference.

Selected dwelling-unit characteristic variables that may influence the preference and/or choice of residential location are listed in Table 4.

3.13 Neighborhood Characteristics -------12-> Residential Preference/Choice

Neighborhood characteristics influence a household's residential preference/choice. Many examples of this hypothesis can be posed, including: 1) a low-density, pedestrian-friendly neighborhood is more likely to be chosen as the place of residence for a young family than is a high-density, congested neighborhood, and 2) the level of quiet and safety a neighborhood has may be more important to an elderly person than to a younger person.

Similar to dwelling-unit characteristics, many researchers have acknowledged the importance of including neighborhood characteristics when designing residential bundles to be evaluated by respondents. In a study of residential preference, Boehm and Ihlanfeldt (1991) found that open space and access to good shopping were valued

Table 4: Dwelling-Unit Characteristic Variables Hypothesized to Influence Residential Preference/Choice

VARIABLE	SELECTED

	REFERENCE(S)
age of home	Quigley (1985), Timmermans *et al.* (1992)
availability of parking	Hoinville (1972)
building period (before 1975, after 1975)	Timmermans *et al.* (1992)
building type (bungalow, multifamily, walk-up apartment, house, etc.)	Lansing and Marans (1969), Lerman (1975), Quigley (1985)
classic architecture	Dussault (1997), Lindstrom (1997)
cost (to own)	Louviere (1979)
expected financial gain from reselling unit	Young (1984)
front footage of lot	Galster and Hesser (1981)
garage	Louviere (1979)
heating fuel type used in unit	Shin (1985)
housing quality	Lu (1998)
landscaping (yes, no)	Menchik (1972) Louviere (1979)
layout of rooms (design)	Menchik (1972)
lot size	Kain and Quigley (1970)
monthly payments (rent or mortgage)	Quigley (1985), Hunt *et al.* (1994)
new paint and fencing	Louviere (1979)
number of bedrooms	Onaka (1983) Timmermans *et al.* (1992)
number of bathrooms	Heikkila *et al.* (1989), Hunt *et al.* (1994)

Table 4: Dwelling-Unit Characteristic Variables Hypothesized to Influence

Residential Preference/Choice - (Continued)

VARIABLE	SELECTED REFERENCE(S)
proximity to traffic	Menchik (1972)
square footage	Onaka (1983) Louviere (1979)
tenure - rent or own	Weisbrod *et al.* (1980)
type of construction (local builder or pre-fabricated)	Louviere (1979)

neighborhood characteristics. Horowitz (1995) included the neighborhood characteristic of school quality as an explanatory variable in a model of the joint choice of residential location and mode to work.

Selected neighborhood characteristic variables that may influence the preference and/or choice of residential location are listed in Table 5.

3.14/15 Neighborhood Characteristics <-13-------14-> Dwelling-Unit Characteristics

The hypothesis that there is an interrelationship between neighborhood characteristics and dwelling-unit characteristics is not commonly discussed, though each is often separately noted for its influence on a household's residential preference/choice. There is typically a strong correlation between them, in which case the distinction is not as important. For example, a large, expensive Victorian home would generally be found in a neighborhood with similar characteristics (i.e., wealth, upscale neighbors, etc.), and consequently, the neighborhood and dwelling unit would not be distinguished. On the other hand, there are situations that are very distinct such as a low-income multi-unit structure in a wealthy, low-density neighborhood. For this scenario a household that otherwise may be interested in a low-density neighborhood may be very opposed to living in the multi-unit structure it contains.

Examples of stereotypical relationships between dwelling-unit characteristics and neighborhood characteristics are given next, but it is important to note that there are always exceptions to these generalizations (for example, Carnahan *et al.* (1974) note that some low-density suburban neighborhoods have areas of high density development).

Table 5: Neighborhood Characteristic Variables Hypothesized to Influence Residential Preference/Choice

VARIABLE	SELECTED REFERENCE(S)
accessibility	Ryan and McNally (1995), Ben-Akiva and Bowman (1998)
age of buildings	Boehm and Ihlanfeldt (1991)
air quality of the region	Young (1984)
attractiveness	Lansing and Marans (1969)
children's playground present	Louviere and Timmermans (1990)
classic architecture	Dussault (1997)
commute distance	Louviere (1979)
connected grid street patterns	Friedman *et al*. (1994)
crime rate	Weisbrod *et al*. (1980), Dussault (1997)
degree of interest (interesting or dull)	Lansing and Marans (1969)
distance to friends (family)	Hensher and Taylor (1983)
distance to recreational activities	Louviere and Timmermans (1990)
distance to shopping (both major and local)	Louviere (1979)
distance to schools	Louviere (1979)

elderly population (percent elderly)	Weisbrod *et al*. (1980)
employment density	Waddell (1993), Frank and Pivo (1994)
fraction of nonwhite households in tract	Aldana *et al*. (1973), Horowitz (1995)
fraction of husband-wife family households	Quigley (1985)
greenery (amount of grass, trees, etc.)	Louviere and Timmermans (1990)

Table 5: Neighborhood Characteristic Variables Hypothesized to Influence Residential Preference/Choice - (Continued)

VARIABLE	SELECTED REFERENCE(S)
housing opportunities (availability of housing)	Weisbrod *et al*. (1980)
integrated civic and commercial centers	Ryan and McNally (1995)
mean age of housing units in neighborhood	Waddell (1993)
median monthly rent	Quigley (1985)
median property values	Shin (1985)
mixed land uses (high, low)	Rutherford *et al.* (1996)
neighborhood appearance	Boehm and Ihlanfeldt (1991), Horowitz (1995), Dussault (1997)
neighborhood prestige	Hunt *et al.* (1994)
neighborhood type (city, suburb, traditional, etc.)	Horowitz (1995),

	Cervero (1996a)
noise	Boehm and Ihlanfeldt (1991)
nearby parks	Dussault (1997)
open space	Boehm and Ihlanfeldt (1991), Ryan and McNally (1995)
parking availability	Boehm and Ihlanfeldt (1991)
pedestrian safety	Young (1984)
percent owner-occupied	Quigley (1985)
presence of "anti-residential" land uses	Menchik (1972)
privacy (distance between neighbors)	Joseph *et al.* (1989)
property tax/household	Weisbrod *et al*. (1980)
public transit (availability, quality)	Boehm and Ihlanfeldt (1991)
quality of natural environment	Menchik (1972)

Table 5: Neighborhood Characteristic Variables Hypothesized to Influence Residential Preference/Choice - (Continued)

VARIABLE	**SELECTED REFERENCE(S)**
residential density (households/acre)	Houghton (1971), Carnahan *et al*. (1974), Horowitz (1995)

road conditions (surface quality, maintenance)	Boehm and Ihlanfeldt (1991)
school quality (such as $/pupil funding)	Horowitz (1995), Dussault (1997)
sense of community (coherent unit)	Ryan and McNally (1995)
shopping opportunities	Boehm and Ihlanfeldt (1991)
size of municipality and location of dwelling (e.g., 250,000 inhabitants located in the city center)	Timmermans *et al.* (1992)
tidiness of area	Young (1984)
traffic levels	Louviere and Timmermans (1990)
transit-oriented land pattern	Ryan and McNally (1995)
type of construction (local builder or pre-fabricated)	Louviere (1979)

First, characteristics such as high percentages of multi-unit buildings, high transit accessibility, and mixed-land uses are more likely to be associated with a traditional neighborhood (Ryan and McNally, 1995), while open space and low-density structures (i.e., mainly single-family units) generally indicate a suburban-type neighborhood. As an example in the opposite direction (i.e., arrow 14), a neighborhood in the tornado-prone Midwestern U.S. is more likely to have a basement (as a storm shelter) compared to a neighborhood in Florida, where high water tables often make basements impractical.

3.16 Attitudes & Lifestyle -------15-> Job Location

A person's job location may be influenced by her or his attitudes and lifestyle. For example, an individual may choose a job in a rural location because she or he prefers a rural lifestyle. Joseph *et al*. (1989) find that respondents are willing to make great sacrifices to fulfill lifestyle needs such as privacy, and job location could be central to such a tradeoff for some. Also, a person may take a job in a high-density, urban location so that he or she can have access to many social activities after work.

3.17 Socio-Demographics & Life Cycle -------16-> Job Location

Socio-demographic and life cycle characteristics can have an impact on job location. For instance, a person juggling many household and child-care duties is more likely to choose a job that is located close to home or daycare to be able to reduce travel time. This balancing of commitments to work and family has been studied extensively (see e.g., Dasgupta, 1996), and many studies have found that women's commute trips have indeed been shorter than men's for these and other possible reasons (see e.g., Madden, 1981). In addition to family factors, some workers may choose a job location because of their income. For example, more than a quarter of the below-poverty-level households are without a vehicle, and consequently, they are more likely to choose a job that is accessible by transit (Harbaugh and Smith, 1998; see also Blumenberg and Ong, 1997).

3.18 Job Location -------17-> Residential Preference/Choice

A common hypothesis of researchers is that a household's (or individual's) job location impacts its residential preference/choice. In other words, job location is assumed to be exogenous in a model of residential location choice (see e.g., Alonso, 1964; Aldana *et al*., 1973; Horowitz, 1995). For example, a young individual takes his or her first job out of school (more concerned about getting a job than the location of the job) and then chooses a residence that is reasonably accessible to the job. Kain (1961), in a landmark study of the impact of journey-to-work on residential choice, noted that commute travel is a large part of a household's travel-time budget, and hence, a household will choose a residence that is located at a distance from work that is acceptable in terms of a limited travel-time budget. Traditional models of residential choice (see e.g., Alonso, 1964; Muth, 1969) have assumed job location to be at the city center, where individuals choose a residence by maximizing a utility function through a tradeoff on land and transportation costs (Dubin, 1985). It is this foundation that has led to the common model specification including job location as a predictor of residential preference/choice.

Selected job location-related variables that may influence preference and/or choice of residential location are given in Table 6. Some of the variables can be considered sociodemographic or travel-demand related, but are included here because of their association specifically with job location.

3.19 Residential Preference/Choice -------18-> Job Location

Though much less common, the reverse hypothesis that residential preference/choice influences job location is also important to test. An example of this is a person who decides that she wants to live in Washington state near her family, then finds a job in that location. The presence of more than one worker in a household adds further complexity to the relationship between residential choice and job location, as the household may be choosing a residential location based on the job location of one member, with another member then choosing a job location based on the residential location (see, e.g., Madden, 1981).

Research pertaining to this hypothesis has been completed in the past decade. For example, Waddell (1993) rejected the assumption that workplace location be considered exogenous in models of residential location, and instead, suggested a joint specification of workplace and residential choice. Verster (1985) noted a small portion of respondents who indicated that their job location was determined after their residential location choice - all heads of households (see Section 2.2.4).

Lastly, a new trend is highly-valued information workers having the ability and option to work anywhere (remotely through computers and telecommunications), demonstrating a "higher degree of geographic mobility" (Giuliano, 1989, pg. 148). These mobile professionals live where they choose, and then maintain their business based from an office in or near their homes.

3.20/21 Job Location <-19--------20-> Travel Demand

Though typically viewed as a one-way relationship with an individual's job location influencing his or her travel demand (arrow 20, see e.g., Kain, 1961), it is increasingly acknowledged that the reverse relationship is also reasonable (arrow 19). For example, a person who expects to commute by auto to work may prefer a job location (and residence, see Section 3.22) that is accessible by uncongested roadways. Another example is an individual who is required to take a large number of non-work trips (such as trips to the bank, grocery store, school, and daycare) and chooses a job location that is closer to home to save time. Madden (1981) notes a similar reasoning for why women have often chosen to work closer to home. Indeed, this hypothesis is most likely applicable to working members of a household who are limited in their household-bargaining abilities.

Job location impacting travel demand is the more common relationship direction seen in the literature (see, e.g., Simpson, 1987). At the most basic level, commute length is a major determinant of total distance traveled on a given day. Job location may affect the modal distribution of travel demand as well: a central-business-district job may influence an individual to take transit to work, even without a predisposition to do. Alternatively, if a person's job is located in a part of town having a high crime rate, he or

Table 6: Job Location-Related Variables Hypothesized to Influence Residential Preference/Choice

VARIABLE	SELECTED REFERENCE(S)
auto travel time to work	Quigley (1985), Ben-Akiva and Bowman (1998)
distance from home to workplace	To *et al.* (1983), Young (1984)
in-vehicle travel time from home to work	Hunt *et al.* (1994)
flexibility of work schedule	Timmermans *et al.* (1992)
length of (work) contract (less than a year, more than a year)	Timmermans *et al.* (1992)
number of workers in household (e.g., dual earner households)	White (1977), Ben-Akiva and Bowman (1998)
occupation	Gordon (1990), Cho (1997)
occupational status (seven point scale, by Hollingshead)	Tardiff (1977)
skill level of worker (low, high)	Waddell (1993)
transit travel time to work	Quigley (1985)
travel time to central business district	Waddell (1993)
work outside of residential zone (yes, no)	Aldana *et al*. (1973)
years worked at current employer	Madden (1981)

she may choose to commute by auto due to safety concerns. Similarly, an individual who has a very long commute may take fewer non-work trips than an individual with a short commute.

3.22 Travel Supply -------21-> Residential Preference/Choice

Travel supply impacts a household's (or individual's) residential preference/choice. It is important to note that travel supply is a special type of neighborhood characteristic. It is singled out to highlight the importance this particular type of neighborhood characteristic is hypothesized to have for residential preference/choice. For example, a household may choose to locate near a bus line (see, e.g., Harbaugh and Smith, 1998), or be influenced by the presence of bike ways in a neighborhood. Other travel supply variables (such as auto availability) are a special type of demographic characteristic (see Section 3.2).

Many researchers have investigated the importance of travel supply to location, land use, and urban form. Steptoe and Thornton (1986) studied changes in land use and economic activity of low-income minority communities due to construction of a new interstate highway. Forkenbrock and Foster (1990) investigated the reductions in transportation costs and increases in economic activity due to corridor highway investment. In terms of residential location, improved travel supply can greatly increase the desirability (value) of a home. Joseph *et al*. (1989) found that the presence of paved roads was a significant factor in the preference formation of households considering buying a rural residence.

Selected travel supply variables that may influence preference and/or choice of residential location are listed in Table 7.

3.23 Residential Preference/Choice -------22-> Travel Demand

A household's residential preference/choice directly impacts its travel demand. For example, a household that resides in an area where there are many outdoor activities (i.e., prefers a residence with access to outdoor activities) available is likely to make more non-work trips than a household that prefers to reside in an area with little recreational opportunity. Likewise, an individual that desires privacy and lives in a home far away from the city center is likely to travel more (in terms of distance) than an individual who lives in a residence close to many destinations, such as a movie theater and grocery store.

This hypothesis has been tested by many researchers. Stopher and Ergun (1982) concluded that residential location was a significant factor in predicting individual travel demand for urban recreational and cultural activities. Likewise, Ewing *et al*. (1994) found that the type of communities (traditional, suburban, etc.) individuals lived in had a large impact on their demand for travel in terms of trip lengths and modes used. For example, it was found that households in suburban areas traveled by car significantly more (about 70% more vehicle-hours) than households in traditional areas. Friedman *et al*. (1994) found significant differences in the travel demand of residents in suburban communities compared to residents in traditional communities, including: 1) more walk and transit trips taken in traditional neighborhoods, 2) more auto-driver trips in suburban neighborhoods, and 3) more total trips in suburban neighborhoods. The characteristic

Table 7: Travel Supply Variables Hypothesized to Influence Residential Preference/Choice.

VARIABLE	SELECTED REFERENCE(S)
distance to entertainment	Young (1984)
distance to public transport	Timmermans *et al.* (1992)
distance to shopping	Louviere (1979)
in-vehicle travel time to shopping center	Hunt *et al.* (1994)
proximity to light rail transit (within walking distance, not within walking distance)	Hunt *et al.* (1994)
traffic congestion	Young (1984)
school bus on route	Joseph *et al.* (1989)
public transport available	Young (1984)
travel time to central business district	Waddell (1993)
transit accessibility to shopping centers	Horowitz (1995)

differences between the two types of neighborhoods were described (similar to descriptions found in comparable studies), essentially noting that traditional neighborhoods were more dense, had greater accessibility to neighborhood locations and activities, and had better transit service than suburban neighborhoods.

Friedman *et al*.'s first two findings have been replicated in several similar studies. For example, Cervero and Radisch (1996) found that residents in a neotraditional town were more likely to use transit and walk than residents in a suburban town. Frank and Pivo (1994) looked at the impact of neighborhood characteristics such as employment density, population density, and mixed land use on individual travel demand. They found that employment density had the highest positive correlation between percent use of transit and percent walking, while mixed land uses had a negative correlation with percent use of single-occupant vehicles. Lastly, Kitamura *et al*. (1997) investigated the impact that neighborhood characteristics had on an individual's demand for trips (types and mode used), and found that parking availability and distance to transit were positively correlated with demand for auto trips.

3.24 Travel Demand -------23-> Residential Preference/Choice

A great body of literature has been generated on the relationship between travel demand and residential preference/choice. Transportation planners value knowledge in this area because it can help them to predict the travel impacts of new development (arrow 22), while developers value this knowledge because it can help them understand the market for different types of development (arrow 23). Of the two possible directions of causality, the hypothesis that a household's travel demand will affect its residential preference/choice is less often investigated, though still potentially important. For example, a retired household that does not travel much may seek a residence in an uncongested, quiet, low-density area. A person who likes to make frequent, short shopping trips may choose a residence that is accessible to many stores.

3.25 Travel Supply -------24-> Travel Demand

This last hypothesis of the conceptual model, that travel supply influences travel demand, is well established. For example, a person who has access to bike lanes in his or her neighborhood is more likely to bike than a person who does not have access to bike lanes. Horowitz (1995), in a study of residential choice and mode to work, found that improvements in transit had a large impact (strongly increased the odds) on it being the selected mode. Likewise, Friedman *et al*. (1994) found that suburban neighborhoods that had wide streets, low accessibility (e.g., unconnected streets), and lots of parking were associated with much higher household automobile trip rates.

3.26 Conclusions and Next Steps

The numerous interrelationships involved with residential preference/choice present a challenge to modelers of the subject. Many of the hypotheses discussed above (such as the impact of residential preference/choice on travel demand) have been successfully modeled before (see for example, Cervero, 1996a), but some have not been found by the author to be empirically investigated (such as the influence of residential preference/choice on attitudes and lifestyle) . It is not surprising that every hypothesis in Figure 1 has not been routinely modeled (especially simultaneously). First, the estimation of all of the parameters that define the conceptual model would require large quantities of detailed data, an expensive and highly time-consuming task. Second, some of the hypotheses are weaker than other related ones and/or very closely related to other hypotheses, making it difficult to find significant parameters. Consequently, it is anticipated that some hypothesized relationships will not be empirically justified, and based on this expectation, another empirically-based conceptual model will be presented later.

The hypotheses in the conceptual model denote a direction of causality. This is purposefully done, with a major goal of the dissertation being to understand the causal linkages that are represented by the statistical relationships in Figure 1. However, causality is a difficult conclusion to justify in an airtight manner, especially with cross-sectional data (see Chapter 8). Hence, the intent of the empirical analysis done in this dissertation is **not** to definitively "prove" causal relationships, but rather to assess the relative strengths of the relationships hypothesized among the components of the model.

Special emphasis will be given to the *Attitudes & Lifestyle* component of the conceptual model. The rich data available to this study on individual attitudes and lifestyles permits a rigorous empirical analysis of the interdependent relationships among individuals' attitudes and lifestyles and areas such as travel demand and residential preference/choice. This work will be a contribution to the travel behavior/land use and spatial interaction research literature, as the power of attitudinal and lifestyle variables in this context has seen only limited exploration to date.

The residential preference/choice component of the conceptual model will be measured by neighborhood choice variables (see Chapter 6). Any number of alternatives could have been defined as the measure of residential preference/choice (e.g., census tract and evaluation score of multi-unit dwelling), but a neighborhood measure was chosen due to its importance in the above literature review (e.g., Frank and Pivo, 1994). In addition, attitudes and lifestyles are likely to be more strongly associated with a residential location defined in terms of a neighborhood. For example, an outgoing individual can be easily associated with a lively neighborhood, but probably not with a certain type of dwelling unit or a specific census tract. Thus, it is suggested that the neighborhood measures used here offer more robust and transferable insights than measures based on specific geographic locations such as census tracts.

Three types of modeling structures will be employed in this dissertation. First, a logit structure will be used to test the importance of attitudinal and lifestyle variables in a binary model of residential choice (suburb versus traditional neighborhood). Second, a regression model structure will be implemented on a continuous measure of residential choice. Last, a structural equation model form will be implemented to estimate the strengths of relationships found in Figure 1.

3.27 Chapter 3 Summary

Chapter 3 describes the residential preference/choice conceptual model of the dissertation. Numerous examples are given to support the hypotheses represented by Figure 1. In addition to reviews of pertinent literature relating to hypotheses contained in the conceptual model, numerous variables representing different components of the conceptual model that are hypothesized to impact residential preference/choice were listed. In essence, this chapter presented the many interconnected causal hypotheses that are at the heart of the empirical modeling work presented later in the dissertation.

CHAPTER 4

EMPIRICAL CONTEXT AND DESCRIPTION OF DATA

4.1 Introduction

The data set used for this dissertation was developed for a land use-travel behavior study sponsored by the California Air Resources Board in 1992. The main objective of the researchers originally collecting these data was to examine the impacts of neighborhood type (i.e., land use) and individual attitudes on travel behavior. Thus, variability of neighborhood type was important in their development of regression-based travel-behavior models. The value of such variability was a motivating factor in the decision to use these data in this dissertation.

In an effort to obtain this variability, five Bay Area neighborhoods were chosen to represent extreme values in terms of key factors describing land-use type: public transit accessibility, land use mix, residential density, and employment mix. This would lead to greater variability in the model input variables, which is desirable in statistical modeling. However, to control for the effect that income can have on travel, neighborhoods with similar (medium) income ranges were purposefully chosen. This decision was important in that it helps distinguish the effect of income from the effect of land use on travel behavior. For example, members of a low-income family may use buses frequently due to not having enough personal vehicles available to them, not because their neighborhood has good public transit (though they may choose to live there because of good public transit). Fortunately, the variation of income *within* neighborhoods is high, which will also permit an explicit examination of income's effect on travel behavior and residential choice.

After the neighborhoods were chosen, great emphasis was placed on obtaining detailed information about each individual neighborhood and its residents. First, a sizeable amount of micro-scale data was collected on the land use, roadway network, and public transit from site surveys of each neighborhood (Concord, Pleasant Hill, North San Francisco, South San Francisco, and San Jose). This type of information would be critical to the identification of neighborhood types for this study's residential choice models. Second, to analyze the relationships between neighborhood type and individual travel demand, demographic, socio-economic, attitudinal, and travel related data were collected through mail-out surveys of residents in these same neighborhoods.

Though the data set is quite rich, with an abundance of information that is valuable to residential choice modeling (such as attitudinal data relating to housing, transportation, the environment, and policy, *and* preferences for various types of living areas), it was not collected specifically to capture household residential choice behavior. For example, data on people's residential choice decision processes, such as how important job location was in the selection of a specific residence, were not collected in this study (and would have been helpful in addressing the goals of this dissertation). Further, individual and joint evaluations of residential attributes helpful in understanding preference formation were not obtained. These factors led to the decision to keep residential preference and residential choice together in one category in the conceptual model (see Figure 1 and Chapter 3).

However, as the original study was aimed at understanding the influence of land use and attitudes on travel behavior (each being an important factor in the study of residential choice), the extension to residential choice is natural. In summary, the data that are available are very relevant, and include variables that are not found in most studies of residential choice. Specifically, data on attitudes toward travel issues (such as public transit and the environment), and on activities (a measure of lifestyle) represent valuable additions to residential choice modeling.

4.2 Generalizability of the Data

Households in the five neighborhoods were sent three surveys: 1) a trip diary, 2) a household survey, and 3) an individual survey. About 18% of those initially contacted agreed to participate in the study, and only 60% of these people participated in all three of the surveys. The low response rate led to an "under-representation of individuals younger than 35 years old, individuals without college education, and households with annual incomes of less than $20,000" (Kitamura *et al.*, 1997, pg. 157).

The underrepresentation of this segment of the population is not expected to undermine the usefulness of this study because the hypotheses to be tested and conclusions to be drawn involve modeling relationships among variables, not projecting sample distributions to the population as a whole. Further, the perception and attitudinal data collected from participants are also likely to represent non-respondents given that the measured perceptions and attitudes encompass a wide range of potential views (Kitamura *et al.*, 1997).

Not all of the survey data collected was used in the estimation of models for this dissertation. More than half of the households participating in the study (526 out of 963 HH) filled out more than one individual survey, for a total of 1331 individual surveys. The household surveys were, by instruction, completed by "one adult member of the household", and though this person was directed to consult with other household members for responses and opinions, it is likely that individuals within the household would provide different answers to the household survey questions (such as "In what type of area would you prefer to live?"). Though it may be econometrically feasible to incorporate intra-household variability and correlation in model estimation, the time and effort required to do so is beyond the intended scope of this dissertation. Consequently, the data sets used in this dissertation are based on purposefully selected subsets of the original surveys from 963 households. This selection process is described next.

First, all households that failed to complete at least one of each of the three different survey instruments (111 HHs) were excluded from this study's sample. This initial filtering step was performed to reduce the amount of missing data for important variables based on each different survey instrument. The next task was to choose the individual survey data for one member within each of the remaining 852 households. As noted before, the impetus for this step was to prevent intra-household correlation from being an issue in the later models. The selection of specific respondents from households was complicated by the fact that the household survey could not at this point be linked to a particular member of the household (i.e., person-specific identification numbers found on each of the other survey instruments were not found on the household surveys). Fortunately, a majority of the variables to be developed for the residential choice conceptual model (such as the attitudinal and lifestyle variables) are based on data that are available for and identifiable to a single specific person in the household. Further, it is suspected that in many instances all three surveys were filled out by the same person in the household.

To select which household member would be kept in the analysis, respondents from households with more than one survey participant were assigned a “sub” identification number (ranging from 1 to 5) by the initial data collectors, and all respondents who had a sub value equal to one were initially selected for this study sample. Although it could not be determined with certainty, it was believed the “sub 1" individual was most likely to be the one completing the household survey, and hence that this selection rule would give the greatest amount of congruence between the three data sources. This rule led to the inclusion of all responses from one-person households as well as from households where only one person participated in completing surveys. In other cases, where the respondent’s individual survey had a lot of missing data (true for less than 1% of sub 1 cases), a different participant within the household was selected for sample inclusion.

Though it was considered most important to have congruent data (i.e., responses from the three survey instruments representing one person), sample diversity was also highly desired. Consequently, 21 respondents who increased the variability of the sample on data needed for the conceptual model (such as people without a driver’s license or who worked part time) were granted an exception to the sub 1 rule (42 were identified, but only every second one was chosen). The selection of underrepresented respondents combined with the missing data rule led to the inclusion of a total of 59 cases having a sub number of 2 or higher (only 7% of the

852 cases chosen). Importantly, a comparison of descriptive statistics on key variables (such as education, gender, and income) between the complete data file (every individual survey included) and the selected data file (after filtering) showed that there were no significant differences.

4.3 Respondent Samples

The modeling analysis and policy discussion for this dissertation is primarily based on four different (but overlapping) samples with respective sizes, $N_1 = 852$, $N_2 = 492$, $N_3 = 615$, and $N_4 = 515$. This section provides an explanation for the varying sample sizes along with summary statistics and discussion of each sample.

The first objective completed was the development of dependent and explanatory variables on which transportation and residential choice models could be based. Attitudinal variables such as "pro-transit" and lifestyle variables such as "adventurer" were created from the full data set of 852 respondents. The data reduction technique factor analysis was used to develop these variables. This analysis was completed on the full data set (i.e., $N_1 = 852$), to maximize the use of the information available, before other considerations led to estimating models on various subsets of the data. Tables 8 and 9 highlight some of the major characteristics of the full sample.

A brief inspection of Tables 8 and 9 reveals that an average respondent from the sample is about 50 years old, has a college degree, and has a household income that ranges from \$35,000 to \$50,000. In addition, more than 15% of the respondents were

Table 8: Sociodemographic Characteristics of the Full Sample[1] (N_1 = 852)

Variable	5 Neighborhoods (N_1=852): NSF(N=155) SSF(N=168) CON(N=165) PH(N=192) SJ (N=172)
Occupation[2]: number, percent of non-missing cases (number of missing cases)	
Manager/administration	124, 15.3% (40)
Professional/technical	285, 35.1% (40)
Administrative support	94, 11.6% (40)
Retired[3]	131, 15.4%
Household composition: mean, standard deviation (number of missing cases)	
Household size	2.25, 1.09
No. people 16 or over	1.85, 0.90
No. people under 16	0.35, 0.74 (15)
No. full-time workers	0.98, 0.70 (1)
No. workers (part- and full-time)	1.18, 0.71 (1)
Personal characteristics: mean, standard deviation (number of missing cases)	
Age	50.23, 14.80 (221)
Education[4]	4.01, 1.29 (10)
Female (=1, Male = 0)	0.52, 0.50 (4)
Household income[4]	6.38, 1.38 (16)
Years lived in Bay Area	29.22, 18.76 (5)

[1] The values given in this table are based on raw data (i.e., no imputed means).
[2] Not all job categories are presented, and thus, percentages do not sum to 100%.
[3] The number of missing cases for occupation and employment status (including "retired) differed since they were obtained from two different questions.
[4] Education and household income were entered as categorical data (e.g., 5 = annual income ranging from $20,001 to $35,000), but here averaged as if they were continuous. Respondents were on average well-educated (a value of 4 represents completion of 4-year degree) with moderate income levels (a value of 6 represents annual income varying from $35,001 to $50,000).

Table 9: Travel and Residential Characteristics of the Full Sample[1] (N_1 = 852)

Variable	5 Neighborhoods (N_1=852): NSF (N=155) SSF (N=168) CON (N=165) PH (N=192) SJ (N=172)
General travel information: mean, standard deviation (number of missing cases)	
No. of vehicles	1.89, 1.03
No. of vehicles / driver	1.04, 0.50 (1)
Commute distance (1-way, miles)	12.20, 12.23 (286)
Daily person trips	4.23, 2.31 (103)
Daily vehicle-miles traveled	28.33, 29.11 (107)
Daily transit-miles traveled	4.02, 12.51 (108)
Daily walk/bike-miles traveled	0.51, 2.55 (108)
Residential characteristics: mean, standard deviation (number of missing cases)	
Home size (square feet)	1537, 637 (67)
No. of bedrooms	2.73, 0.97 (7)
Home value category[2] (for the 625 homeowners)	4.44, 1.17 (1)
Monthly rent category[2] (for the 217 renters)	3.47, 1.07
Most important reasons for choosing current neighborhood: number, percent[3]	
Housing cost	463, 54.3%
Close to shops and services	210, 24.6%
Close to work	210, 24.6%
Good school	107, 12.6%

[1] The values given in this table are based on raw data (i.e., no imputed means).

[2] Home value and monthly rent were collected as ordinal categorical variables. Reference points for each category include: 4 (home value ranging from $180,001 to $250,000), 5 (home value ranging from $250,001 to $375,000), 3 (monthly rent ranging from $501 to $700), and 4 (monthly rent ranging from $701 to $1,000).
[3] There were no missing data on for the "most important reasons" variables. Responses sum to more than 100% since more than one reason could be offered.

retired. Further, one may note that missing data is more of an issue for the travel-related variables like commute distance and daily person trips.

The first set of residential choice models was based on a subsample (N_2 = 492) of the 852 potential data records. A binary logit model structure was chosen due to its simplicity, where the dependent variable was type of neighborhood (= 1 if suburban, = 0 if traditional). Site characteristics (such as street pattern, bike paths, and distance to shops) were used to classify the five neighborhoods in this study as traditional or suburban. In reality, however, each neighborhood had some characteristics of both types. Pleasant Hill and South San Francisco appeared to be especially mixed in nature, so to begin analyzing neighborhoods that were more pure examples of each type, all cases from the Pleasant Hill and South San Francisco neighborhoods were discarded from the binary choice analysis.

Tables 10 and 11 highlight some of the major characteristics of the respondents from these three neighborhoods. A cursory investigation of these tables reveals that there are indeed differences among respondents across neighborhoods. The average individual from North San Francisco is younger and more educated than the average individual from Concord or San Jose. The average commute distance is longest for respondents from Concord while living next to a good school was more important to people in San Jose than to those in the other two neighborhoods.

The models developed in Chapters 7 and 8 required yet another modification of the full sample, resulting in $N_3 = 615$. Specifically, any respondent who was unemployed was removed from the analysis. In total, 237 people from the full sample of 852 were

Table 10: Sociodemographic Characteristics of the Binary Logit Model Sample[1] (N_2 = 492)

Variable	NSF (N=155) Traditional	CON (N=165) Suburban	SJ (N=172) Suburban
Occupation[2]: number, percent of non-missing cases (number of missing cases)			
Manager/administration	23, 16% (8)	19, 12% (4)	26, 16% (10)
Professional/technical	47, 32% (8)	46, 29% (4)	59, 36% (10)
Administrative support	22, 15% (8)	16, 10% (4)	19, 12% (10)
Retired[3]	12, 8% (0)	35, 21% (0)	27, 16% (0)
Household composition: mean, standard deviation (number of missing cases)			
Household size	1.83, 0.90 (0)	2.45, 1.09 (0)	2.72, 1.11 (0)
No. people 16 or over	1.61, 1.07 (6)	1.88, 0.90 (4)	2.16, 0.74 (0)
No. people under 16	0.11, 0.39 (5)	0.49, 0.89 (4)	0.57, 0.92 (1)
No. full-time workers	0.98, 0.70 (0)	0.94, 0.72 (0)	1.00, 0.71 (0)
No. workers (part- and full-time)	1.19, 0.61 (0)	1.12, 0.73 (0)	1.26, 0.77 (0)
Personal characteristics: mean, standard deviation (number of missing cases)			
Age	43.7, 14.2 (29)	54.1, 14.8 (48)	52.2, 13.9 (40)
Education[4]	4.32, 1.23 (1)	3.58, 1.24 (0)	3.82, 1.28 (3)
Female (=1, Male = 0)	0.53, 0.50 (3)	0.50, 0.50 (0)	0.48, 0.50 (0)
Household income[4]	6.14, 1.45 (3)	6.32, 1.26 (4)	6.44, 1.44 (3)
Years lived in Bay Area	19.7, 17.1 (0)	35.2, 18.2 (2)	32.3, 15.3 (1)

[1] The values given in this table are based on raw data (i.e., no imputed means).
[2] Not all job categories are presented, and thus, percentages do not sum to 100%.
[3] The number of missing cases for occupation and employment status differed since they were obtained from two different questions.
[4] Education and household income were entered as categorical data (e.g., 5 = income ranging from $20,001 to $35,000), but here averaged as if they were continuous. Respondents were on average well-educated (a

value of 4 represents completion of 4-year degree) with moderate income levels (a value of 6 represents income varying from $35,001 to $50,000).

Table 11: Travel and Residential Characteristics of the Binary Logit Model Sample[1] (N_2 = 492)

Variable	NSF (N=155) Traditional	CON (N=165) Suburban	SJ (N=172) Suburban
General travel information: mean, standard deviation (number of missing cases)			
No. of vehicles	1.35, 0.92 (0)	2.15, 1.01 (0)	2.37, 1.08 (0)
No. of vehicles / driver	0.91, 0.55 (0)	1.07, 0.50 (0)	1.09, 0.54 (0)
Commute distance (1-way, miles)	6.70, 11.10 (43)	15.98, 14.86 (66)	14.05, 11.76 (53)
Daily person trips	4.75, 3.08 (29)	4.16, 2.43 (22)	4.15, 2.14 (22)
Daily vehicle-miles traveled	21.68, 28.74 (30)	32.97, 35.23 (22)	34.04, 24.72 (22)
Daily transit-miles traveled	2.92, 8.36 (30)	5.62, 16.89 (22)	0.64, 4.58 (22)
Daily walk/bike-miles traveled	0.90, 2.07 (30)	0.40, 1.61 (22)	0.043, 0.34 (22)
Residential characteristics: mean, standard deviation (number of missing cases)			
Home size (square feet)	1304, 825 (28)	1527, 483 (8)	1678, 398 (6)
No. of bedrooms	2.02, 1.09 (4)	2.98, 0.70 (1)	3.51, 0.63
Home value category[2] (for the 50, 147, and 155 homeowners)	5.58, 1.25 (0)	3.76, 0.80 (0)	4.53, 0.62 (0)
Monthly rent category[2] (for the 105, 18, and 17 renters)	3.42, 1.15 (0)	3.06, 0.85 (0)	3.44, 1.26 (0)
Most important reasons for choosing current neighborhood: number, percent[3]			
Housing cost	86, 55%	101, 61%	104, 61%
Close to shops and services	57, 37%	32, 19%	43, 25%
Close to work	53, 34%	53, 32%	29, 17%
Good school	8, 5%	27, 16%	36, 21%

[1] The values given in this table are based on raw data (i.e., no imputed means).
[2] Home value and monthly rent were collected as ordinal categorical variables. Reference points for each category include: 4 (home value ranging from $180,001 to $250,000), 6 (home value ranging from $375,001 to $575,000), 3 (monthly rent ranging from $501 to $700), and 4 (monthly rent ranging from $701 to $1,000).
[3] There were no missing data on the "most important reasons" variables. Responses sum to more than 100% since more than one reason could be offered.

defined as unemployed, either retired (N = 131) or temporarily unemployed (N = 106).

The belief that retired people will have a different residential-choice decision process than employed respondents, and the fact that commute distance is a fundamental component of the conceptual model, led to this action. For example, retired people gave "close to shops and services" as an important reason for choosing their current neighborhood about 50% more often than did employed respondents.

The next modification made to the study sample was the removal of 100 respondents from the N_3 group to obtain the final group, N_4. This was done in an effort to develop a model that satisfies a key assumption for the validity of maximum likelihood estimation results, that model variables follow a multivariate normal distribution. Details of how these cases were identified for removal can be found in Chapter 8.

Though the original sample size was large (N_3 = 615), there was a concern that the removal of 100 outliers (about 17% of the original sample) would create a data set unrepresentative of the past data sets. An inspection of the sociodemographic and travel- related characteristics of the respondents from both data sets (see Tables 12 - 15) relieves this concern somewhat. The variables number of full-time workers and commute distance were the only variables that appeared to differ noticeably between the two samples. Further, regression models of residential choice developed on both data sets (N_3 and N_4) showed that significant explanatory variables were nearly all the same (including age, commute distance and culture lover) across models. The variables daily walk/bike-miles traveled and daily vehicle-miles traveled were the only variables that were only significant in the reduced data set (N_4).

Table 12: Sociodemographic Characteristics of the Cleaned Employed Sample[1] (N_3 = 615)

Variable	**5 Neighborhoods (N_3=615): NSF(N=121) SSF(N=119) CON(N=105) PH(N=140) SJ (N=130)**
Occupation[2]: number, percent cases	
Manager/administration	120, 19.5%
Professional/technical	274, 44.6%
Administrative support	93, 15.1%
Retired	Not Applicable
Household composition: mean, standard deviation	
Household size	2.33, 1.14
No. people 16 or over	1.89, 0.88
No. people under 16	0.40, 0.76
No. full-time workers	1.22, 0.60
No. workers (part- and full-time)	1.46, 0.50
Personal characteristics: mean, standard deviation	
Age	46.4, 10.1
Education[3]	4.16, 1.27
Female (=1, Male = 0)	0.52, 0.50
Household income[3]	6.64, 1.25
Years lived in Bay Area	24.5, 15.5

[1] The values given in this table are based on imputed means replacing missing data.
[2] Not all job categories are presented, and thus, percentages do not sum to 100%.
[3] Education and household income were entered as categorical data (e.g., 5 = annual income ranging from $20,001 to $35,000), but here averaged as if they were continuous. Respondents were on average well-educated (a value of 4 represents completion of 4-year degree) with moderate income levels (a value of 6 represents annual income varying from $35,001 to $50,000).

Table 13: Travel and Residential Characteristics of the Cleaned Employed Sample[1] (N_3 = 615)

Variable	**5 Neighborhoods (N_3=615): NSF(N=121) SSF(N=119) CON(N=105) PH(N=140) SJ (N=130)**
General travel information: mean, standard deviation	
No. of vehicles	1.95, 1.01
No. of vehicles / driver	1.05, 0.50
Commute distance (1-way, miles)	12.08, 11.82
Daily person trips	4.45, 2.22
Daily vehicle-miles traveled	30.13, 28.28
Daily transit-miles traveled	5.05, 13.36
Daily walk/bike-miles traveled	0.56, 2.68
Residential characteristics: mean, standard deviation	
Home size (square feet)	1495, 614
No. of bedrooms	2.71, 0.99
Home value category[2] (for the 435 homeowners)	4.39, 1.15
Monthly rent category[2] (for the 180 renters)	3.56, 1.03
Most important reasons for choosing current neighborhood: number, percent	

Housing cost	348, 56.6%
Close to shops and services	135, 21.9%
Close to work	175, 28.5%
Good school	65, 10.6%

[1] The values given in this table are based on imputed means replacing missing data.
[2] Home value and monthly rent were collected as ordinal categorical variables. Reference points for each category include: 4 (home value ranging from $180,001 to $250,000),
6 (home value ranging from $375,001 to $575,000), 3 (monthly rent ranging from $501 to $700), and 4 (monthly rent ranging from $701 to $1,000).

Table 14: Sociodemographic Characteristics of the Cleaned Reduced Sample[1] (N_4 = 515)

Variable	5 Neighborhoods (N_4=515): NSF (N=95) SSF (N=100) CON (N=87) PH (N=121) SJ (N=112)
Occupation[2]: number, percent cases	
Manager/administration	98, 19.0%
Professional/technical	238, 46.2%
Administrative support	82, 15.9%
Retired	Not Applicable
Household composition: mean, standard deviation	
Household size	2.31, 1.12
No. people 16 or over	1.85, 0.86
No. people under 16	0.42, 0.76
No. full-time workers	0.80, 0.40
No. workers (part- and full-time)	1.45, 0.50
Personal characteristics: mean, standard deviation	
Age	44.8, 9.4

Education[3]	4.15, 1.27
Female (=1, Male =0)	0.55, 0.50
Household income[3]	6.64, 1.20
Years lived in Bay Area	24.7, 15.6

[1] The values given in this table are based on imputed means replacing missing data.
[2] Not all job categories are presented, and thus, percentages do not sum to 100%.
[3] Education and household income were entered as categorical data (e.g., 5 = annual income ranging from $20,001 to $35,000), but here averaged as if they were continuous. Respondents were on average well-educated (a value of 4 represents completion of 4-year degree) with moderate income levels (a value of 6 represents annual income varying from $35,001 to $50,000).

Table 15: Travel and Residential Characteristics of the Cleaned Reduced Sample[1] (N_4 = 515)

Variable	**5 Neighborhoods (N_4=515): NSF(N=95) SSF(N=100) CON(N=87) PH(N=121) SJ (N=112)**
General travel information: mean, standard deviation	
No. of vehicles	1.94, 0.93
No. of vehicles / driver	1.05, 0.44
Commute distance (1-way, miles)	10.81, 9.62
Daily person trips	4.28, 1.80
Daily vehicle-miles traveled	28.41, 24.94
Daily transit-miles traveled	4.35, 11.79
Daily walk/bike-miles traveled	0.22, 0.53
Residential characteristics: mean, standard deviation	
Home size (square feet)	1488, 580

No. of bedrooms	2.71, 1.00
Home value category[2] (for the 368 homeowners)	4.39, 1.11
Monthly rent category[2] (for the 147 renters)	3.61, 1.01
Most important reasons for choosing current neighborhood: number, percent	
Housing cost	288, 55.9%
Close to shops and services	112, 21.7%
Close to work	152, 29.5%
Good school	54, 10.5%

[1] The values given in this table are based on imputed means replacing missing data.
[2] Home value and monthly rent were collected as ordinal categorical variables. Reference points for each category include: 4 (home value ranging from $180,001 to $250,000),
6 (home value ranging from $375,001 to $575,000), 3 (monthly rent ranging from $501 to $700), and 4 (monthly rent ranging from $701 to $1,000).

A review of Tables 12 and 13 shows that the typical employed respondent has a college education, a professional occupation, and takes 4.45 trips per day on average. Interestingly, the aggregate numbers for the full sample (N_1 = 852) are very similar to the aggregate numbers for the reduced, employed sample (N_3 = 615). A comparison of all the samples reveals that there are some strong similarities among the different groups. For example, housing cost was selected most frequently as a "most important reason for choosing current neighborhood", indicating that income was a key factor in individual residential choice.

Though it was noted earlier that sample N_4 has very similar sociodemographic characteristics as sample N_3, descriptives for N_4 are presented in Tables 14 and 15. This sample was used in the estimation of the final structural equation model presented in Chapter 8. The joint distribution of the variables from this data set satisfied the critical structural equation modeling assumption of multivariate normality. It is important to note that many other variables not shown in this table had similar (almost identical) means to the same variables in the N_3 data set.

The complexity of the empirical estimation of the fully-developed residential choice models presented in Chapter 8 led to the development of the N_3 and N_4 data sets with no missing data. First, despite the fact that AMOS (the software package utilized in estimating structural equations in Chapter 8) advertises that it can compute full-information maximum likelihood estimates "in the presence of missing data" (AMOS, 1997; pg. 500), models would not converge with data sets containing missing values. Discussion on a structural equation e-mail discussion group indicated that this was a common problem with AMOS 3.6. Second, it was desirable for the estimation to take place on the largest possible sample for maximum efficiency. Thus, listwise and pairwise deletion methods may be less effective than replacing missing data with mean values or imputed values.

Using the methods discussed in Section 4.4, all of the missing data in N_3 and N_4 was replaced with mean values. These data sets with no missing values (that is, the "cleaned" data sets) were used to develop all the statistical models in Chapters 7 and 8, and consequently, these cleaned versions of N_3 and N_4 are the basis for Tables 12-15. For completeness, the raw (uncleaned) versions of N_3 and N_4 are presented in Tables 16-19 in the appendix.

The differences between the raw and cleaned data sets are small. In fact, two of the variables with the most variation, professional/technical and commute distance, varied by about 4.5% and 1% respectively for N_3 and N_4.

4.4 Missing Data

Nearly all social science researchers must grapple with issues related to missing data (Little and Rubin, 1990). This study was no exception, and multiple methods were implemented to account for missing data. First, survey responses (or missing responses) on many important variables from one survey could be cross-checked with related variables based on data from another of the three surveys. One example that was cleaned was employment, a binary variable equal to one if the respondent checked that s/he had a job. Inspection of two other variables (job type and yearly personal income) allowed for a logical change in the employment variable value. For example, if a person had a value of zero (indicating not employed) or nothing (missing), but had responses to job type and personal income that were indicative of being employed, then the ostensible unemployment was taken to be temporary (or a response error) and the employment value was changed to a one. In the original data set containing 1331 individual cases, this variable was changed for only 42 (3.2%) of them. Other variables' values were changed significantly fewer times.

Many variables contained missing data that could not be cleaned by cross- checking. In particular, variables relating to travel, such as commute distance and daily vehicle-miles traveled, had the most missing data. Indeed, nonresponse in travel diary surveys is common, and failure to address the issue can lead to biases which can negatively impact the reliability of modeling results on such data (Polak and Han, 1997). In addition to this general concern, the importance of these variables in the residential choice conceptual model motivated the use of the method of regression imputation for their missing values.

A common strategy for handling missing values, regression imputation is the process of estimating a regression model for the dependent variable Y (the variable with the missing value problem) as a function of explanatory variables, using cases having no missing data on either the dependent or explanatory variables (Little and Rubin, 1990). The coefficient estimates obtained from the regression create a prediction equation that can be used to estimate the remaining missing Y values. The prediction equations used to impute missing data for commute distance, daily vehicle-miles traveled, and daily transit-miles traveled, are given in the Appendix, Tables 20-22. The higher the variation explained by the independent variables, the better the regression imputation effectiveness (Little and Rubin, 1990). It was found that commute distance was a significant variable in both regression imputation equations for travel demand. Given the explanatory power this variable had, it was concluded that the regression-based imputed means for commute distance would be used as input values into the prediction equations for daily vehicle-miles traveled and daily transit-miles traveled. Also, since the adjusted R-squared value for daily walk/bike-miles traveled was so poor (R^2-adjusted = 0.014), it was believed that naive imputation would be essentially as effective as regression-based imputation for missing values of this variable.

The regression model adjusted R-squared values ranged from 0.17 (for commute distance) to 0.27 (for daily vehicle-miles traveled), indicating that the regression imputation for vehicle miles was more successful than the regression imputation of missing data for commute distance. While the adjusted R-squared values are modest, these models represent an improvement over the common procedure of simply filling missing data with the mean.

For the remaining variables (such as gender and education), since they generally had few missing cases and were primarily exogenous variables less central to the model structure, simple or “naive” imputation (Polak and Han, 1997) was implemented. This method is standard practice and is generally considered satisfactory as a first approximation. Briefly summarized, neighborhood-specific mean values were imputed for cases with missing data on these variables. Thus, Concord respondents having missing data on income were imputed with the mean income value for Concord (based on the Concord participants who reported income), with the same procedure being performed for respondents in the other neighborhoods.

4.5 Chapter 4 Summary

The main purpose of this chapter was to describe the data that were used to develop the samples analyzed later in this dissertation. First, the background of the study for which this dissertation is an extension is given. A brief description of the design of the data collection immediately follows along with discussion of the generalizability of the data. Next, a description of the different respondent samples used in model development is given, including short discussions of why the samples were chosen. The chapter ends with a discussion of the methods used to handle missing data.

CHAPTER 5

ATTITUDE/LIFESTYLE MEASUREMENT

AND

DISCRETE CHOICE MODELS OF RESIDENTIAL NEIGHBORHOOD

5.1 Introduction

It seems self-evident that residential location decisions profoundly influence travel patterns, but the precise nature of that influence is not completely understood. For example, numerous empirical studies (see, e.g., Frank and Pivo, 1994; Rutherford *et al*., 1996; and Kitamura *et al*., 1997) have demonstrated that living in higher-density, traditional neighborhoods is associated with fewer vehicle trips and smaller distances traveled compared to living in typical low-density suburban environments. These encouraging results have supported a growing movement to use land use planning and design as a tool for reducing travel. This movement is likely to be successful if land use configuration (or residential choice) "causes" individual travel patterns. However, it is believed that other factors influence people's travel behavior, perhaps more strongly than does land use configuration, and it is important to investigate such variables to gain a better understanding of the interaction between spatial characteristics and individual travel.

It is expected that individual travel-related predispositions are one type of factor that strongly impacts both residential choice and travel demand. To investigate this belief, it is necessary to analyze data on individuals' attitudes and lifestyle preferences. In this interest, an extensive set of attitudinal and lifestyle variables was developed from the study data (see

Section 5.2).

To measure the explanatory power of these types of variables on residential choice, binary logit models were estimated and analyzed (see Section 5.3). The choice of a binary dependent variable structure was selected for the first set of residential choice models due to its simplicity. The development of an endogenous variable to represent residential preference/choice in the conceptual model is a complicated task, and the findings from the binary model were a first step towards achieving this goal. Interestingly, it turned out that even the simple binary structure of the neighborhood location dependent variable (suburb = 1 and traditional = 0) presented difficulties. For example, what characteristics truly represent a suburb, and to what extent must a neighborhood possess these characteristics? Further complicating the matter is the fact that neighborhoods can have both traditional and suburban characteristics. Indeed, this was especially the case for Pleasant Hill (PH) and South San Francisco (SSF), as these two neighborhoods had characteristics that were associated with each type of neighborhood. For example, PH was characterized by high residential density (commonly associated with traditional neighborhoods) and low population density (commonly associated with suburban neighborhoods). To establish a more contrasting endogenous variable (i.e., a larger variance between neighborhoods labeled traditional versus suburban), all cases from these neighborhoods were discarded from the binary choice analysis.

5.2 Attitude and Lifestyle Measurement

To measure attitudes and lifestyles, two factor analyses were performed on responses to numerous survey items related to personal views and activities. After experimenting with various factor extraction and rotation options, principal components analysis (PCA) and oblique rotation solutions were selected in both cases, with the number of factors chosen based on the eigenvalue-one and interpretability rules of thumb (Rummel, 1970). Tables 23 and 24 present the largest pattern matrix loadings for the final factor solutions. Mean values for each of the attitudinal and lifestyle factor scores by neighborhood are presented in Figures 2 and 3 respectively, where the factor scores are arranged roughly in order of degree of significant variation across neighborhood (based on a one-way analysis of variance on each factor score).

5.2.1 Attitude Measurement

Responses on a 5-point Likert-type (strongly disagree to strongly agree) scale to 39 statements relating to urban life (covering topics such as urban transportation, the environment, and housing) were factor analyzed with SPSS (Norusis, 1990). Table 23 shows the largest pattern-matrix loadings for the final 10-factor solution, which accounts for 49.1% of the total variance in the attitude data. One-way ANOVAS performed on each factor indicated that mean scores on all but the last (pro-transit) factor differed significantly across the three neighborhoods analyzed here. However, for brevity, only the four factors significant in the binary logit residential choice models presented in Section 5.3 will be discussed below. For a more detailed discussion of all ten factors, see

Kitamura *et al*. (1994 and 1997).

The pro-high density factor is based on statements such as, "I need to have space between me and my neighbors" (loading = -0.75) and "High-density residential development should be encouraged" (loading = 0.55). It is hypothesized that a person who has a high score on this factor will be more likely to prefer a residence in a high-density area. As expected, the mean score on this factor for North San Francisco was much higher (0.57) than for Concord (-0.49) or San Jose (-0.33), indicating that respondents in the traditional neighborhood are more favorable toward high-density development than respondents in the suburban neighborhoods of this study.

The pro-environment factor is identified by statements such as "environmental protection costs too much" (loading = -0.78) and "stricter vehicle smog control laws should be introduced and enforced" (loading = 0.47). An individual who is very environmentally sensitive may be more likely to live in a traditional neighborhood, as this type of neighborhood uses land more efficiently and facilitates the use of transportation modes other than the automobile. The mean factor scores shown in Figure 2 support this hypothesis, with the ranking by neighborhood the same as for the pro-high density factor although all means are less extreme for this factor than for the first.

Pro-pricing and pro-alternatives are factors relating to regulations and policies concerning transportation and the environment. The pro-pricing factor is characterized by statements such as "I would be willing to pay a toll to drive on an uncongested road" (loading = 0.76) and "We should raise the price of gasoline" (loading = 0.38). The pro-alternatives factor is somewhat heterogeneous, but generally relates to the provision of alternatives to gasoline-powered automobile travel, including statements such as “We should provide more incentives to people who use electric or other clean-fuel vehicles” (loading = 0.42) and “More lanes should be set aside for carpools and buses” (loading = 0.39). It is hypothesized that an individual who favors policies supporting more environmentally-efficient forms of travel will be more likely to live in a traditional neighborhood. Figure 2 supports this hypothesis, showing that, on average, NSF residents scored significantly more highly on both these factors than did residents of the two suburban neighborhoods.

These four attitudes collectively point to major differences between North San Francisco respondents’ and Concord and San Jose respondents’ views of land use and the environment. It is important to understand that this statistically significant difference in and of itself does not imply a particular direction of causality. Do people with different attitudes choose to live in different neighborhoods, or do the different neighborhoods in which people live engender different attitudes? Although the latter direction of influence (residential location causes attitudes) may well occur over the long run, the former relationship (attitudes cause residential location) is believed to be the stronger direction of influence in the short term. Resolving this question more completely would require a longitudinal study of how the attitudes of residents of different types of neighborhoods change with the length of time that they live in those neighborhoods.

5.2.2 Lifestyle Measurement

Lifestyle was measured based on the responses to three questions in the survey: 1) “What types of subjects did you read last month (check all that apply)?”, having 30 possible choices plus “other”; 2) “What best describes the way you spent last weekend (check as many as apply)?”, having 19 choices plus “other”; and 3) “From the following lists check all that you have done within the last 12 months”, having 57 possible responses in four categories labeled outdoors/sports, entertainment/events, travel, and do it yourself/education/hobbies, plus “other” responses for each category. Discarding the “other” responses resulted in a total of 106 binary variables representing a diverse set of lifestyle activities.

Factor analysis was performed on these 106 variables (although this procedure is more commonly conducted on variables that are at least approximately continuous, Rummel, 1970, points out that any data can be factor-analyzed). The final eleven-factor solution shown in Table 24 explains 29.4% of the total variance in the activity data, indicating that a considerable amount of the total variance in lifestyle indicators falls outside the 11-dimensional space spanned by the identified factors. Based on one-way ANOVAs on each factor, mean factor scores on the first six factors of Table 24 and Figure 3 differed significantly by neighborhood. Again, for brevity, only the three lifestyle factor scores that were significant in the models of Section 5.3 will be discussed below.

The lifestyle factor that differs most significantly across neighborhoods is strongly defined by activities such as: “attended a concert/symphony” (loading = 0.49), “attended the ballet” (loading = 0.46), and “attended the theater” (loading =0.39). Hence, this is labeled the “culture lover” factor. It is hypothesized that people with a cultural-oriented

lifestyle will be more likely to choose a residence that is accessible to many cultural activities (most likely in or near the high-density urban core). This hypothesis is supported by Figure 3, showing that North San Francisco residents have a much higher mean score on the culture-lover factor (0.65) than do residents of the two suburban neighborhoods Concord and San Jose (-0.35 and -0.33, respectively).

The next significant factor is characterized by activities such as: “read material on home improvement” (loading = 0.65), “made house improvements myself” (loading = 0.57), and “spent last weekend doing yardwork” (loading = 0.53). This factor has been named “nest builder” as it refers to a lifestyle that involves many home-related activities. A person who has a high score on this factor is hypothesized to be more likely to live in a low-density neighborhood, where homes and lots are larger and home-ownership is higher. As anticipated, respondents from Concord and San Jose had higher average scores on this factor than did respondents from North San Francisco.

The third factor found significant in the residential choice models was labeled “altruist”, based on statements such as: “Spent last weekend on religious activities” (loading = 0.58), “volunteered to help the community” (loading = 0.53), and “participated in community events” (loading = 0.43). Although we did not have a prior hypothesis about how this factor would differ by neighborhood, Figure 3 shows that San Jose residents scored most highly on this factor, NSF residents scored most negatively, and Concord residents were nearly neutral. While the differences are statistically significant according to the one-way ANOVA, the spread between the highest and lowest mean is smaller than for the other two lifestyles discussed. It may be that San Jose residents are marginally more conservative and hence (perhaps), more inclined to be religious; it may be that their marginally larger households, higher presence of children, and incomes gives them somewhat more motivation and means to participate in community activities.

In identifying differences in lifestyle by neighborhood, it is again important to examine the question of the direction of causality. The stronger direction of causality here may be more debatable than in the case of attitudes, since neighborhood type could clearly influence the activities undertaken. Do “culture-lovers” live in the urban core partly to have ready access to cultural events, or do they live there for other reasons but, after the fact, are induced to take advantage of their greater proximity to those events? Do “nest-builders” engage in home-improvement activities mainly because their large suburban home and yard (which they have chosen for other reasons) require them to do so, or do those who enjoy engaging in home-improvement activities choose to live in such a neighborhood while those who do not enjoy them choose to live in a higher-density, lower-maintenance residence such as a condo or apartment?

Although it is acknowledged that the “residential location causes lifestyle” link can be important, it is believed that the reverse direction is still quite plausible in this context. The specific definition of the lifestyle variables used here supports their interpretation as indicators of predisposition - that is, as causes rather than effects. For the most part, the variables represent activities which would be relatively accessible to everyone in a large metropolitan area, regardless of their specific neighborhood type. The fact that the time frame for 57 of the activities was "within the last year" even further allows for rough equality of opportunity across neighborhood types. The 30 variables identifying subjects the respondents read about within the last month are likely to reflect intrinsic interests, and again the subjects would be equally available to residents of all neighborhood types.

5.3 Binary Logit Residential Choice Models

The above discussion of attitudinal and lifestyle measures is an indicator of the complexity of developing variables for residential choice modeling. Another equally challenging task is defining endogenous variables to model. As noted earlier, for the first set of residential choice models, it was decided that a simple binary dependent variable structure would be implemented, where residential choice was defined as either a suburban (=1) or traditional neighborhood (=0).

To briefly summarize, binary logit models of residential choice were estimated on the data from three of the five neighborhoods: North San Francisco, Concord, and San Jose (N_2 = 492). More than 60 measures representing travel, residence, employment, attitude, lifestyle, and sociodemographic characteristics were evaluated for inclusion as explanatory variables in the model. In addition to t- and chi-square tests and analysis of variance, stepwise procedures of adding and removing variables (singly or in groups) were conducted to select the final, "best" model shown in Table 25.

With an adjusted-$?^2$ statistic of 0.52 (compared to 0.10 for the market share model containing only a constant), the overall model goodness-of-fit is respectable. The negative sign for the constant term indicates that unmeasured variables favor the choice of a traditional neighborhood on average. The remaining ten significant variables all had

the expected signs, and fall into three categories: sociodemographic variables, attitude factor scores, and lifestyle factor scores.

The three sociodemographic variables - number of people under age 16, number of vehicles, and number of years lived in the Bay Area - are all positively associated with choosing a suburban neighborhood. The appeal of the suburbs (larger homes and yards, perceived better schools and safer environment) as a place to raise a family needs no further explanation. The association of higher numbers of vehicles with a suburban residence is consistent with findings from the travel behavior/land use literature (see, e.g., Cervero, 1996a; Rutherford *et al*., 1996), although some mutual causality could certainly be at work here (to move to a suburban home with attached garage and less transit availability may necessitate and/or facilitate the acquisition of additional cars). The last

significant sociodemographic variable, "years lived in the Bay Area", can be considered a life cycle proxy. Specifically, it is hypothesized that people who have large values for this variable are older and more likely to have (or have had) children living at home. Even if a household is now in the empty nest stage and no longer needs the four bedroom home near good schools, inertia may keep it in that location which was optimal in the past. Nijkamp *et al*. (1993) hypothesized that life cycle was a key explanatory factor in household relocation decisions.

Seven out of the 10 significant variables in the model are attitudinal (4) and lifestyle (3) factor scores, demonstrating the considerable explanatory power of these types of variables. As would be expected from Figure 2 and the discussion in Section 5.2.1, the signs on the four attitudinal variables are all negative, indicating that people scoring highly on the pro-pricing, pro-environment, pro-high density, and pro-alternatives factors are significantly more likely to live in a traditional neighborhood. The signs for the three lifestyle factor scores are consistent with Figure 3 and the discussion in Section 5.2.2: those scoring high on the culture-lover factor are more likely to live in a traditional neighborhood, whereas those scoring highly on the nest-builder and altruist factors are more likely to be suburbanites.

It is of interest to quantify the specific contribution of sociodemographic and attitudinal/lifestyle variables to this particular model. This is done through their stepwise inclusion in the forward direction and exclusion in the backward direction, with the results shown in Tables 25 and 26. Including one block of variables in the forward direction from the market share model and comparing the difference in ?^{2}s provides an upper bound on the contribution of that (included) block (the italic numbers in Table 26), since the included variables are carrying part of the explanatory power of the excluded ones. Conversely, excluding a block of variables in the backward direction from the full model and comparing differences in ?^{2}s provides a lower bound on the contribution of that (excluded) block (the bold numbers in Table 26), again because the included block is carrying part of the explanatory power of the excluded block.

Table 26 demonstrates that the block of attitudinal/lifestyle factors carries greater explanatory power than the block of sociodemographic variables. The ?^{2} measure is substantially higher for the model containing only factor scores plus the constant than for the model containing only sociodemographic variables plus a constant. The incremental contribution of the attitudinal/lifestyle block to the model containing only the other block

(19.6 percentage points) is greater than the incremental contribution of the sociodemographic block to the model containing only the attitudinal/lifestyle variables (8.6 percentage points). A more stringent test would develop "best" sociodemographic-only and attitudinal/lifestyle-only models where additional variables in each block were tested for inclusion. But assuming the full model presented in Table 25 truly is the "absolute best" model of all feasible specifications, the qualitative outcome of such a test (i.e., that attitudinal/lifestyle variables are more powerful as a block) is unlikely to be different.

This model, together with the previous results of Kitamura *et al*. (1997), offers tentative support to the proposition that the main role of residential neighborhood type with respect to travel behavior is not one of direct causality. Kitamura *et al*. (1997) found that attitudinal factors (similar to the ones developed here on the same data set) carried much greater explanatory power in a model predicting fraction of auto trips than did neighborhood-related variables. The model presented here demonstrates that neighborhood choice itself is strongly associated with, and probably influenced by, attitudes and lifestyle. This suggests that much, if not most, of the relationship observed between land use configuration and travel behavior in previous studies can be explained by the influence of attitudes and lifestyle on both residential location and travel behavior. If this is true, then simply altering the land use configurations is unlikely to have the desired effect on travel behavior without also changing attitudes and lifestyle. The structural equations models developed in Chapter 8 will address this issue more rigorously, by simultaneously accounting for multiple relationships among these and other variables.

5.4 Chapter 5 Summary

This chapter investigated the importance of attitudinal and lifestyle variables on residential choice for residents in three Bay Area neighborhoods. First, the concept of attitudes and lifestyle in relation to residential choice was introduced. Next, the development and discussion of attitudinal and lifestyle measures was given. The chapter ended with an analysis of binary residential choice models, where the contributions of different types of variables (sociodemographic and attitudinal/lifestyle factors) to model variation were examined. In particular, it was noted that neighborhood choice is strongly associated with attitudes and lifestyle, and further, that the influence of attitudes and lifestyle on both residential choice and travel behavior may be stronger than that of land use configuration.

CHAPTER 6

DEFINING THE DEPENDENT VARIABLE, TRADITIONALNESS

6.1 Introduction

One of the most important steps in modeling any system is the identification and measurement of the endogenous variable(s). Defining the key dependent variable to be modeled for this study is a complicated task due to the many components associated with a household's or individual's choice of residential location. The "ideal" dependent variable may be defined by all the factors relevant to the choice of residential location such as dwelling unit characteristics, neighborhood characteristics, available public services, and housing supply. Unfortunately, though, this would result in a model with so many alternatives that empirical estimation could become a problem (Tu and Goldfinch, 1996).

One way to define residential choice is by the type of local area in which an individual or household lives. The binary dependent variable for the residential choice models in Chapter 5 was based on this definition, where each alternative was defined by the traditional and suburban character of neighborhoods in North San Francisco, Concord, and San Jose. A major motivation for selecting neighborhood as the geographic level of residential location choice was the desire to be able to compare results with other studies of residential location and travel demand that used neighborhood as their spatial designation. Many of these studies (see, e.g., Boehm and Ihlandfeldt, 1991; Prevedouros, 1992; Friedman *et al*., 1994; and Cervero and Radisch, 1996) classified neighborhoods into different types, such as traditional and suburban. By using neighborhood as the spatial scale in this study, and neighborhood type as the specific measure of interest, model estimation results are more comparable and more likely to be transferable and generalizable to other locations (as opposed to modeling the choice of specific geographic neighborhood - such as North San Francisco over Pleasant Hill - which would probably not be transferable).

The issue at hand is the importance endogenous variable development has in model estimation. In other words, does the dependent variable adequately capture the heterogeneity of the residential location choice? If not, model estimation results will likely be of little value. The next two sections present findings from the literature that pertain to residential choice endogenous variable formation. Specifically, discussion is given to how researchers have defined residential choice, along with a critical look at the strengths and weaknesses of these definitions.

6.2 Neighborhood Definitions in the Literature

The large number and variation among neighborhood definitions in the literature indicated the challenge of modeling residential choice. Characterizing a specific geographical area is not straightforward due to the fact that it can be described along more than one dimension, including physical, social, and psychological. Further, even if only one dimension is isolated, the task of defining a residential choice would be difficult; for example, Madanipour (1996) devoted an entire paper to defining the concept of physical space. Literature findings in two main areas of neighborhood definition development, geographic boundaries and geographic characteristics, are given next.

6.2.1 Spatial Scale and Boundaries in Neighborhood Definition Development

Physical space was one particular aspect of the complexity of defining residential choice, as researchers varied greatly in how they chose geographical boundaries for a neighborhood. Two different ways of viewing boundaries in residential choice appear in the literature, representing opposite ends of the precision spectrum.

First, some researchers did not define a neighborhood boundary in their studies (the least precise). For example, Lansing and Marans (1969) investigated respondents' perceptions of neighborhood quality and concluded that "while neighborhood was not defined in the interview, the context of the questions and the nature of the replies made it clear that the respondents were talking about the immediate vicinity of their homes" (pg. 196). The American Housing Survey, a national survey which collects data on items such as neighborhood land-use composition and household characteristics, does not provide a definition of neighborhood in its survey because of its designers' belief that people's notions of neighborhood vary (American Housing Survey, 1997). Lu (1998) supports this position with the following statement (pg. 1482): "Because a researcher's notion of neighborhood is likely to differ from a respondent's, the use of a predefined notion [boundary] of neighborhood may lead to distorted empirical results."

Second, many researchers have (precisely) defined neighborhoods by census tracts (see, e.g., Weisbrod, 1980; Heikkila *et al*., 1989; Waddell, 1993; Horowitz, 1995; and Cervero and Kockelman, 1997). One reason for using census tracts is that it is a relatively detailed but convenient geographical unit (compared to, e.g., the smaller census blocks) for which data is available (e.g., from the U.S. Census Bureau) for use in conjunction with household survey data. One drawback of the census tract is that it is generally arbitrarily determined. In other words, a census tract may be composed of greatly varying areas, and consequently, a census tract statistic such as residential density may be misleading due to being the average of high and low residential densities.

Lastly, although it is not clear whether one spatial scale is superior to another in land-use studies, the spatial scale chosen by a researcher for modeling is definitely important (Handy, 1993). For example, Boarnet and Sarmiento (1998) found that land-use variables measured at the zip code level were significant in a model of non-work automobile trips, while land-use variables at the census-tract level were not. Two relevant conclusions from their paper to the dependent variable development and modeling of this dissertation are (pg. 1166): 1) "controlling for residential location choice and using different levels of geographic detail when studying the link between land use and travel behavior [is important]", and 2) "more attention [should be] given to areas which are larger than what many New Urbanists consider the immediate neighborhood."

6.2.2 Physical Characteristics as Neighborhood Indicators

Some researchers have viewed neighborhoods in terms of their proximity to an urban city center, defining these neighborhoods with terms such as urban (located in or close to the central business district area) and suburban (see, e.g., Kain and Quigley, 1970; Aldana *et al*., 1973; Boehm and Ihlanfeld, 1991; and Prevedouros, 1992).

Other researchers extended the above definitions of urban and suburban neighborhoods by incorporating physical characteristics such as street network and dwelling-unit composition into the neighborhood categorization process. Neighborhoods defined with terms such as urban, traditional, neotraditional, and suburban that represent particular combinations of such underlying physical characteristics have been a key element in the land use and travel demand research literature (see, e.g., Ewing *et al*., 1994; Friedman *et al*., 1994; and Rutherford *et al*., 1996).

This last set of definitions of neighborhood comes close to denoting "neighborhood indices". Sawicki and Flynn (1996) describe the Quality of Life Project in Jacksonville, Florida, where 74 indices (such as number of households, percent of adults in the labor force, percent of population in poverty, and accessibility of supermarkets) were used to define Jacksonville. The authors note (pg. 179) that neighborhood indices are "just beginning to be used to make and evaluate policy, and to search for the causes of change in neighborhoods and in the lives of their residents" - such as changes in mobility. Sawicki and Flynn feel that indices are more useful when they are viewed separately (i.e., not part of an overall index), though little support was given for this opinion. Unfortunately, this author feels the potentially valuable concept of neighborhood indices is lost in their method of implementation. Specifically, though individual indicators provide richness of information, when comparing neighborhoods it is not reasonable to expect researchers to cognitively process 74 different indices, let alone prospective residents.

For the purpose of modeling residential choice, a continuous, disaggregate measure of neighborhood type is suggested to be preferable. As neighborhoods should be defined in terms of what they mean for residents (Handy, 1997), a disaggregate, individual-specific measure is more likely to capture the variation of individuals' perceptions of where they live. Further, neighborhoods possess characteristics that are continuous in nature. For example, population density can vary continuously across neighborhoods (e.g., see Carnahan, *et al.*, 1974), and thus, choosing an arbitrary cutoff point to define "high" density (which may later be used to classify a neighborhood as traditional) will result in a loss of valuable information. The loss of information becomes even greater when more than one characteristic is forced into a binary category. Lastly, a continuous dependent variable may be more tractable than a categorical one when estimating statistical models. The next section describes some of the challenges in creating a good neighborhood variable.

6.3 Challenges in Creating a Neighborhood Variable

A review of the land use and travel demand literature reveals many neighborhood characteristics that are or have been associated with neighborhood types, such as suburban or traditional. Friedman *et al*. (1994) categorized 550 San Francisco Bay Area communities geographically defined by census tracts as suburban if they (pg. 64): "[were] developed since the early 1950s with segregated land uses", "[had] a well-defined hierarchy of roads", "concentrate[d] site access at a few key points", and "[had] relatively little transit service". The authors established the following criteria for communities to be characterized as traditional (pg. 64): "were mostly developed before World War II", "had a mixed-use downtown commercial district with significant on-street curbside parking", and "had an interconnecting street grid and residential neighborhoods in close proximity to nonresidential land uses". Cervero and Kockelman (1997), in a study of how the built environment impacts travel demand, considered a large number of neighborhood variables, including: pedestrian-related factors such as sidewalk and bike path supply, automobile-related factors such as amount of parking and average arterial speed limits, and density-related factors such as nearness to stores and number of jobs per acre. Ryan and McNally (1995) presented design concepts for neotraditional neighborhoods (i.e., areas similar to traditional neighborhoods but built at a later time period), and noted that the main design goal of "neotraditionalists" was to implement neighborhood design characteristics that would create a "coherent neighborhood unit" that while still useable by car, would "de-emphasize and discourage its use". Design characteristics viewed as supporting this goal included: interconnected street networks, centralized retail and office space, and pedestrian and bicycle pathways.

Many researchers have studied the impact of urban form on travel using data on characteristics such as those previously mentioned (see, e.g., Boarnet and Sarmiento, 1998). Many (such as Friedman *et al*., 1994) have viewed residential location in terms of a binary variable -- either it is suburban or it is not. This definition may lead to ambiguous results as some residential locations (neighborhoods) can have a mix of characteristics that are found in both traditional and suburban locations. Thus, a person living in a high-density, transit-served corner of a census tract that otherwise appears to be a suburb (and is categorized as one by a researcher) may bias travel demand model results by increasing the average number of transit trips taken by a “suburban” respondent. Consequently, it may be more fruitful to model a dependent variable that better captures the potential heterogeneity of a residential neighborhood, and that is more sensitive to the specific characteristics faced by any particular resident. A variable that measures or defines residential location on a continuum may be able to improve modeling involving location variables.

Cervero and Kockelman (1997), for example, developed location variables that were continuous in nature. Using factor analysis, they uncovered two continuous dimensions that defined their study neighborhoods: “walking quality” (a factor based on attributes such as sidewalk availability and block length), and “intensity” (a factor based on attributes such as population density and retail store availability). To *et al*. (1983), in a similar effort, used principal components analysis to define a continuous housing quantity measure of residential location. It is believed that a continuous measure of neighborhood type will better represent (in terms of accuracy and model explanatory power) the residential location dependent variable for this dissertation. One reason for this view is that the residential location types represented in the available data are diverse and would not fit a discrete definition well.

In this study we have five geographical neighborhoods which could be considered measures of an individual’s residential choice. However, what generically characterizes a neighborhood is more of interest for residential choice modelers than a specific neighborhood itself. As discussed previously in Chapter 1, the trait of “traditionalness” is the defining dimension chosen for this study (though many other traits such as aesthetic beauty could be appropriate in other contexts). Many studies have defined geographical locations as being traditional or not traditional, but the author is not aware of any study that has developed a non-binary measure of traditionalness. One potential reason for this is the complexity of the definition of a traditional location. Unfortunately, though, ambiguous model results and/or erroneous findings can occur by trying to fit a rich, complex concept (such as traditionalness) into a simple, binary indicator (Etzioni and Lehman, 1967). A discussion of the background and development of this dissertation's neighborhood variable is given next.

6.4 Measuring Neighborhood Type

A first step toward defining a continuous, disaggregate measure of neighborhood traditionalness is to carefully analyze the characteristics that past researchers have identified with suburban and traditional neighborhoods (many of which were discussed in the previous section). Examination of these characteristics suggests that they could be categorized into three groups, with variables relating to: 1) density (such as population or dwelling-unit), 2) accessibility (such as pedestrian or work-related), and 3) pedestrian friendliness (such as walking safety). A traditional neighborhood would have high values on all three of these dimensions, and the degree to which a specific neighborhood possesses these characteristics defines the "level of traditionalness" it has. It is plausible that a neighborhood can have high values on one dimension but not the others, but it is more likely that the three dimensions will vary together (i.e., that density, accessibility, and pedestrian friendliness will be correlated). To investigate these ideas, key data were selected to be factor analyzed to see what dimensions and relationships would result.

In this study, careful consideration was given to finding an appropriate way to operationalize a continuous, disaggregate measure of traditionalness by incorporating a desirable level of complexity into the measure. From our available data, eighteen variables that are consistent with the literature and that collectively measure a range of relevant characteristics were identified. Values on these eighteen characteristics were obtained for the five neighborhoods from two sources: a comprehensive report on the neighborhoods' physical characteristics (Kitamura *et al.,* 1994) and individual responses to this study's survey questions. The average value by neighborhood for each of the characteristics is given in Table 27.

The mean neighborhood value for a particular characteristic is believed to represent (at least partly) the degree to which a neighborhood meets a particular traditional (or nontraditional) neighborhood characteristic/concept. For example, "number of parking spaces available for household use" is a proxy for residential density and/or household dependence on personal vehicles. A high mean value for this would more likely represent a nontraditional or suburban residential location. An example in the other direction is "good local public transit in your neighborhood", where a high mean value would be more indicative of a traditional neighborhood than a nontraditional neighborhood. Both of these examples support the prior field visit conclusions that the North San Francisco neighborhood is a good example of a traditional location (note the low mean value for parking, 1.43, and the high mean value for transit, 0.98), and that the

Table 27: Characteristics Used to Measure Traditionalness

Characteristic[1]	Data Type[2]	Mean Value of Characteristic (standard deviation within neighborhood)				
		NSF	SSF	CON	PH	SJ
Speed limit of road (S)	C, A	25.19 (0.00)	25.31 (0.00)	25.54 (0.00)	25.82 (0.00)	25.52 (0.00)
Grid-like street configuration (T) high = 1, medium = 0.5, low = 0	B, A	1 (0.00)	0 (0.00)	0.5 (0.00)	0 (0.00)	0.5 (0.00)
Population density (T) high = 1, low = 0	B, A	1 (0.00)	1 (0.00)	0 (0.00)	1 (0.00)	0 (0.00)
Size of home in square feet (S)	C, I	1366.6 (805.6)	1837.9 (834.4)	1551.5 (452.7)	1348.6 (608.4)	1687.2 (379.8)
Have a back yard (S) yes = 1, no = 0	B, I	0.47 (0.50)	0.93 (0.25)	0.97 (0.17)	0.54 (0.50)	0.97 (0.17)
Streets in neighborhood pleasant for walking/jogging (T) yes = 1, no = 0	B, I	0.84 (0.37)	0.90 (0.30)	0.91 (0.29)	0.86 (0.35)	0.95 (0.23)
Cycling is pleasant in your neighborhood (T) yes = 1, no = 0	B, I	0.63 (0.48)	0.49 (0.50)	0.90 (0.29)	0.94 (0.24)	0.84 (0.37)
Good local public transit in your neighborhood (T) yes = 1, no = 0	B, I	0.98 (0.14)	0.94 (0.23)	0.86 (0.34)	0.91 (0.28)	0.72 (0.45)
Enough parking space near your home (S) yes = 1, no = 0	B, I	0.48 (0.50)	0.77 (0.42)	0.89 (0.32)	0.76 (0.43)	0.91 (0.29)
Problems of traffic congestion in your neighborhood (T) yes = 1, no = 0	B, I	0.35 (0.48)	0.24 (0.43)	0.32 (0.47)	0.59 (0.49)	0.36 (0.48)
Distance in miles from your home to nearest public transit option (S)	C, I	0.24 (0.52)	0.28 (0.48)	0.25 (0.23)	0.35 (0.50)	0.51 (0.61)
Sidewalks are in your neighborhood (T) yes = 1, no = 0	B, I	1.00 (0.00)	1.00 (0.00)	0.60 (0.49)	0.76 (0.43)	0.99 (0.11)

Table 27 (continued): Characteristics Used to Measure Traditionalness

Characteristic[1]	Data Type[2]	Mean Value of Characteristic (standard deviation within neighborhood)				
		NSF	SSF	CON	PH	SJ
Bike paths are in your neighborhood (T) yes = 1, no = 0	B, I	0.31 (0.46)	0.04 (0.20)	0.79 (0.41)	0.91 (0.29)	0.44 (0.50)
Public transit is convenient in your neighborhood (T) yes = 1, no = 0	B, I	0.97 (0.18)	0.92 (0.27)	0.95 (0.23)	0.98 (0.14)	0.63 (0.49)
Number of parking spaces available for household use (S)	C, I	1.43 (1.04)	2.23 (1.15)	4.06 (2.08)	2.83 (3.24)	4.02 (1.48)
Distance in miles to nearest grocery store (S)	C, I	0.45 (0.50)	0.64 (0.56)	0.77 (0.57)	1.06 (0.79)	0.95 (0.66)
Distance in miles to nearest gas station (S)	C, I	0.47 (0.74)	0.91 (0.67)	0.72 (0.57)	0.82 (0.55)	0.90 (0.84)
Distance in miles to nearest park or playground (S)	C, I	0.51 (0.79)	0.65 (0.77)	0.70 (0.60)	1.45 (1.21)	0.68 (0.52)

[1] (T) indicates that a traditional location is hypothesized to have a higher mean value for this characteristic than a suburban (S) location.

[2] The characteristic data, being either effectively continuous (C) or binary (B), is taken from both aggregate (A, averages based on each neighborhood as a whole) and disaggregate sources (I, averages based on individual responses).

San Jose neighborhood is a good example of a suburban location (note the high mean value for parking, 4.02, and the low mean value for transit, 0.72).

A comparison of the mean values across neighborhoods shows that some neighborhoods have high values on some characteristics that are representative of traditional locations, and also have high values on some typical suburban characteristics. For example, Pleasant Hill has a high mean value for the traditional characteristic "good local public transit in your neighborhood" and a high value for the suburban characteristics "distance in miles to nearest park" and "grocery store". In essence, this is an indication that neighborhoods can have both traditional and nontraditional characteristics. This lends support to the contention that a continuous location measure is more appropriate for modeling than the common binary measures of location.

Both the aggregate and disaggregate data shown in Table 27 are important. First, aggregate data on neighborhood characteristics such as population density and street configuration are important as a base for neighborhood definition. The "unbiased" facts can help provide a valid picture of what the neighborhood is like on the whole. However, as noted earlier, not all neighborhoods are homogeneous, and two individuals living in different parts of the same neighborhood may experience very different situations. Further, two individuals may feel very differently about the exact same location within a neighborhood, which makes the disaggregate response valuable. For example, two people living in the same apartment complex may respond oppositely to the statement "there is good local public transit in your neighborhood." At first glance, this may appear to be undesirable. However, the modeling is at the individual level, and it is the individual's perception of the neighborhood that influences her/his choice. Thus, that choice will be more accurately modeled when it is measured in the terms that the individual uses rather than based on somewhat arbitrary distinctions imposed by the researcher.

6.5 Traditionalness, a Factor Analysis Approach

A factor analysis approach was taken to reduce the 18 interrelated characteristics shown in Table 27 into a smaller number of underlying dimensions. Various factor structures were hypothesized *a priori*. One hypothesis was that a single dimension of traditionalness would emerge, with the factor analysis essentially providing the "optimal" weights for combining the 18 variables into a single composite index (Rummel, 1970). Another hypothesis was that three dimensions might emerge along the lines of density, accessibility, and pedestrian friendliness. Multiple factor analyses were performed to determine what structures were most appropriate.

One important choice in the analysis is that between an aggregate versus disaggregate measure of traditionalness. In Table 27 the first three characteristics, speed limit of road, grid-like street configuration, and population density, are aggregate values (notice the standard deviation of zero) in that they are not differentiated by respondent. Though it is acknowledged that the values for these characteristics could be very different across participants in the same neighborhood, disaggregate data was not available, and consequently, in the disaggregate database, the mean value for each neighborhood was assigned to each respondent in the corresponding neighborhood. The remaining 15 characteristics, on the other hand, have values that vary from respondent to respondent. Hence, a decision must be made about whether to treat the characteristic data at the aggregate or disaggregate level when developing a measure of traditionalness for this study's empirical models.

Separate data sets with aggregate and disaggregate values for the 18 characteristics were developed for factor analysis. The primary justification for the aggregate data base is the fact that all of the research studies reviewed by the author looked at location characteristics in the aggregate, typically in terms of zonal averages. However, the aggregate measure has at least two weaknesses. First, reducing the individual respondents' responses to neighborhood means leaves a database that has five cases (each neighborhood being a case or sample point). Statistical inferences on such a small sample size may be biased, and should be viewed with caution (Guadagnoli and Velicer, 1990). Second, as Table 27 shows, most of the 15 disaggregate characteristics vary within each neighborhood, and using an aggregate measure may seriously misrepresent certain respondents. Both of these weaknesses are addressed by the development of a disaggregate measure of traditionalness. In the disaggregate database there are 852 cases. As this author is not aware of any research comparing the effectiveness of disaggregate and aggregate measures for location choice models, both methods were implemented. Models developed using both disaggregate and aggregate dependent variables are presented and compared in Chapter 7.

6.6 Factor Analysis Results

Analyses extracting 3, 2, and 1 factors, respectively, were performed using SPSS 8.0 on the disaggregate (N=852) data set, and a one-factor extraction was completed on the aggregate (N=5) data set (with so few cases, extracting more than one dimension was not appropriate). A variety of extraction (such as principal components and principal axis factoring) and rotation (such as varimax and oblique) methods were conducted in the factor analysis. Results were consistent among all combinations of methods, but the outcome used for the modeling in this dissertation is based on the principal components extraction and oblique rotation methods since this combination explained the most variation in the data and was the most interpretable.

6.6.1 The Aggregate and Disaggregate One-Factor Solutions

Tables 28 and 29 present the factor loadings for the 1-factor aggregate structure and 1-factor disaggregate structure, respectively. To assist in the interpretation of these results, the mean of each variable by neighborhood is presented alongside the factor loading for that variable. Both factor structures represent the measurement of the characteristic, level of traditionalness, along a single continuum. The single aggregate factor represented in Table 28 explained 44.7% of the total variation in the 18 neighborhood characteristics. Characteristics that are primary determinants of this factor include: "enough parking available near home" (loading = -0.95), "good public transit" (loading = 0.88), and "population density" (loading =

0.73). Neighborhoods that have high, positive scores for this factor are considered to be more

Table 28: Factor Loadings with Mean Neighborhood Values: 1-Factor, Aggregate Structure (Level of Traditionalness)

Characteristic[1]	Loading	Mean Value of Characteristic				
		NSF	SSF	CON	PH	SJ
Enough parking available near home [2]	-0.95	0.48	0.77	0.89	0.76	0.91
Number of parking spaces for HH use	-0.94	1.43	2.23	4.06	2.83	4.02
Good public transit	0.88	0.98	0.94	0.86	0.91	0.72
Distance to nearest grocery store (mi.)	-0.84	0.45	0.64	0.77	1.06	0.95
Streets are pleasant for walking	-0.81	0.84	0.90	0.91	0.86	0.95
Distance to nearest gas station (mi.)	-0.74	0.47	0.91	0.72	0.82	0.90
Population density (1 = high, 0 = low)	0.73	1.00	1.00	0.00	1.00	0.00
Distance to nearest public transit (mi.)	-0.73	0.24	0.28	0.25	0.35	0.51
Speed limit of roads (mph)	-0.70	25.2	25.3	25.5	25.8	25.5
Cycling is pleasant	-0.66	0.63	0.49	0.90	0.94	0.84
Have own backyard	-0.65	0.47	0.93	0.97	0.54	0.97
Public transit is convenient	0.63	0.97	0.92	0.95	0.98	0.63
Bike paths are present	-0.46	0.31	0.04	0.79	0.91	0.44
Level of grid-like street network (1 = high, 0 = low)	0.42	1.00	0.00	0.50	0.00	0.50
Sidewalks are present	0.37	1.00	1.00	0.60	0.76	0.99
Distance to closest park (mi.)	-0.34	0.51	0.65	0.70	1.45	0.68
Home size (1000 square feet)	-0.31	1.37	1.84	1.55	1.35	1.69
Traffic congestion is present	-0.21	0.35	0.24	0.32	0.59	0.36

[1] The characteristics are ranked by the magnitudes of their loadings on the single aggregate factor for neighborhood type, *Traditionalness*.

[2] Characteristics based on a statement like "enough parking available near home" have a value equal to 1 if the respondent answered yes, and a value equal to 0 if the respondent answered no (see Table

27).

Table 29: Factor Loadings with Mean Neighborhood Values: 1-Factor, Disaggregate Structure (Level of Traditionalness)

Characteristic[1]	Loading	Mean Value of Characteristic (standard deviation)[2]				
		NSF	SSF	CON	PH	SJ
Speed limit of roads[2] (mph)	-0.79	25.2	25.3	25.5	25.8	25.5
Bike paths are present	-0.56	0.31 (0.46)	0.04 (0.20)	0.79 (0.41)	0.91 (0.29)	0.44 (0.50)
Distance to nearest grocery store (mi.)	-0.53	0.45 (0.50)	0.64 (0.56)	0.77 (0.57)	1.06 (0.79)	0.95 (0.66)
Level of grid-like street network[2]	0.45	1.00	0.00	0.50	0.00	0.50
Cycling is pleasant	-0.44	0.63 (0.48)	0.49 (0.50)	0.90 (0.29)	0.94 (0.24)	0.84 (0.37)
Number of parking spaces for HH use	-0.41	1.43 (1.04)	2.23 (1.15)	4.06 (2.08)	2.83 (3.24)	4.02 (1.48)
Population density[2] (1 = high, 0 = low)	0.41	1.00	1.00	0.00	1.00	0.00
Sidewalks are present	0.36	1.00 (0.00)	1.00 (0.00)	0.60 (0.49)	0.76 (0.43)	0.99 (0.11)
Enough parking available near home	-0.36	0.48 (0.50)	0.77 (0.42)	0.89 (0.32)	0.76 (0.43)	0.91 (0.29)
Distance to nearest gas station (mi.)	-0.35	0.47 (0.74)	0.91 (0.67)	0.72 (0.57)	0.82 (0.55)	0.90 (0.84)
Distance to closest park (mi.)	-0.35	0.51 (0.79)	0.65 (0.77)	0.70 (0.60)	1.45 (1.21)	0.68 (0.52)
Good public transit	0.28	0.98 (0.14)	0.94 (0.23)	0.86 (0.34)	0.91 (0.28)	0.72 (0.45)
Distance to nearest public transit (mi.)	-0.26	0.24 (0.52)	0.28 (0.48)	0.25 (0.23)	0.35 (0.50)	0.51 (0.61)
Have own backyard	-0.21	0.47 (0.50)	0.93 (0.25)	0.97 (0.17)	0.54 (0.50)	0.97 (0.17)
Public transit is convenient	0.18	0.97 (0.18)	0.92 (0.27)	0.95 (0.23)	0.98 (0.14)	0.63 (0.49)
Streets are pleasant for walking	-0.16	0.84 (0.37)	0.90 (0.30)	0.91 (0.29)	0.86 (0.35)	0.95 (0.23)
Traffic congestion is present	-0.10	0.35 (0.48)	0.24 (0.43)	0.32 (0.47)	0.59 (0.49)	0.36 (0.48)

Home size (1000 square feet)	-0.05	1.37	1.84	1.55	1.35	1.69

[1] The characteristics are ranked by the magnitudes of their loadings on the single disaggregate factor for neighborhood type, *Traditionalness*.
[2] Site-based aggregate characteristics had a standard deviation of zero.

traditional than neighborhoods that have a low value for it. The standardized scores for the five neighborhoods on this aggregate factor are 1.51 for North San Francisco, 0.38 for South San Francisco, -0.29 for Pleasant Hill, -0.48 for Concord, and -1.13 for San Jose (see Figure 4). This measure of neighborhood type is the basis for the dependent variable of residential choice used in the regression model shown in Table 32, Chapter 7.

The single disaggregate factor for level of traditionalness, presented in Table 29, explained 15.2% of the total variation in the 18 neighborhood characteristics. The disaggregate data had far more variance to explain than did the aggregate data (N = 852 versus N = 5), and consequently, the fact that the disaggregate factor explained a far smaller proportion of that variance than did the aggregate factor is not viewed as an indicator that the aggregate factor is superior. Characteristics that are primary determinants of the single disaggregate factor include: "speed limits of roads" (loading = -0.79), "bike paths are present" (loading = -0.56), and "level of grid-like street network" (loading = 0.45). As before, neighborhoods that have high, positive scores for this factor are considered to be more traditional than neighborhoods that have a low value for it. The means and standard deviations on the disaggregate standardized factor score for the five neighborhoods are 1.47 (0.44) for North San Francisco, 0.63 (0.48) for South San Francisco, -0.46 (0.60) for San Jose, -0.55 (0.50) for Concord, and -0.85 (0.53) for Pleasant Hill (see Figure 4). This measure of neighborhood type is the basis for the dependent variable of residential choice used in the regression model shown in Table 33, Chapter 7.

Empirical findings generally matched expectations, as the two San Francisco neighborhoods clustered on the "traditional"side of the neighborhood measure with the only positive scores while the other three neighborhoods clustered on the suburban side with negative scores. The quintessentially traditional neighborhood of North San Francisco had the highest positive mean factor score on both the aggregate and disaggregate measures of level of traditionalness (having high values on traditional characteristics such as grid-like street networks and public transit accessibility), while the stereotypical suburban neighborhood San Jose had a negative mean factor score on both measures (having high values on suburban characteristics such as number of parking spaces and distance to shopping). While the ordering among the three suburban neighborhoods differs between the two solutions, each aggregate score falls within about one standard deviation of the corresponding mean disaggregate score.

Inspection of Tables 28 and 29 shows that the factor loadings for all characteristics have the same sign in each of the two structures, an indicator of some convergence between the two methods. However, the magnitudes of the factor loadings differ between the aggregate and disaggregate solutions. For example, the loading on the characteristic "enough parking available near home" is -0.95 for the aggregate solution (it is the characteristic with the highest loading), but only -0.36 for the 1-factor disaggregate solution. This discrepancy makes it difficult to identify confidently which characteristics are the most important determinants of a neighborhood's level of traditionalness.

The signs of the factor loadings (which represent the correlation between the characteristics and the level of traditionalness dimension) matched expectations for 15 of the 18 characteristics. For example, “enough parking available near home” and “distance to nearest grocery store” had large negative loadings, indicating that neighborhoods that have high mean values for these characteristics would align more on the suburban dimension than on the traditional dimension. The three characteristics with unexpected loadings (all negative) were “streets are pleasant for walking”, “cycling is pleasant”, and “bike paths are present”. These were expected to have positive loadings since previous research has shown that respondents in traditional neighborhoods are more likely to take non-motorized modes of travel than respondents from suburban neighborhoods (e.g., Kitamura *et al*., 1997). An inspection of Table 27 shows that the three neighborhoods categorized as suburban (Concord, Pleasant Hill and San Jose) had the highest neighborhood means for the characteristics “cycling is pleasant” and “bike paths are present” (while also having very high means on the characteristic “streets are pleasant for walking”). Thus, the negative factor loadings make sense given the data, though they do not conform to the romanticized image of traditional neighborhoods being the places for relaxed walk and bike trips.

It is important to note some qualifications on the use of these single-factor solutions. First, as mentioned earlier, the aggregate measure is based on a very small sample size (N=5), which has been hypothesized to bias statistical output (see, e.g., Guadagnoli and Velicer, 1990). However, it may be argued that using a small sample size for factor analysis is only a problem when making statistical inferences (such as assigning validity to the amount of variance explained), not when determining underlying dimensions. Second, unlike the two-factor disaggregate solution discussed next, the aggregate and disaggregate single factors are unrotated. Rotation in these cases, however, was not only unnecessary but undesirable, as the point was to create a single index incorporating the contribution of all the neighborhood characteristics to the traditionalness dimension. Rotating the axis would have increased the contribution of some characteristics while minimizing the contribution of others. An unrotated factor solution is just as valid as a rotated solution, with both methods explaining the same amount of variance in the data and delineating the same number of relevant dimensions (Rummel, 1970).

6.6.2 The Multi-Factor Disaggregate Solutions

Though the single factor solutions described above were conceptually interpretable, traditionalness could theoretically be a meta-scale composite of several subordinate dimensions. As noted earlier, possible dimensions such as pedestrian friendliness and accessibility were postulated for conceptual reasons. Inspection of the three-factor structure showed that three logical dimensions could not be identified with this study's data. The inability to identify a three-factor structure could have been the result of many things, including insufficient data variation (and type) and/or neighborhoods varying along one or two of the hypothesized dimensions but not all three. On the other hand, a review of the two-factor structure showed that the data could be usefully described by two different dimensions. Tables 30 and 31 contain the ranked pattern matrix loadings (along with the mean value by neighborhood for each characteristic) for each of the dimensions of the 2-factor disaggregate structure.

Table 30: Ranking of Factor Loadings with Mean Neighborhood Values,

Dependent Variable for 2 Factor, Disaggregate Structure (Suburban)

Characteristic[1]	Loading	Mean Value of Characteristic (standard deviation)[2]				
		NSF	SSF	CON	PH	SJ
Speed limit of roads[2] (mph)	0.84	25.2	25.3	25.5	25.8	25.5
Distance to nearest grocery store (mi.)	0.62	0.45 (0.50)	0.64 (0.56)	0.77 (0.57)	1.06 (0.79)	0.95 (0.66)
Distance to closest park (mi.)	0.58	0.51 (0.79)	0.65 (0.77)	0.70 (0.60)	1.45 (1.21)	0.68 (0.52)
Bike paths are present	0.57	0.31 (0.46)	0.04 (0.20)	0.79 (0.41)	0.91 (0.29)	0.44 (0.50)
Level of grid-like street network[2]	-0.56	1.00	0.00	0.50	0.00	0.50
Distance to nearest gas station (mi.)	0.38	0.47 (0.74)	0.91 (0.67)	0.72 (0.57)	0.82 (0.55)	0.90 (0.84)
Cycling is pleasant	0.36	0.63 (0.48)	0.49 (0.50)	0.90 (0.29)	0.94 (0.24)	0.84 (0.37)
Distance to nearest public transit (mi.)	0.26	0.24 (0.52)	0.28 (0.48)	0.25 (0.23)	0.35 (0.50)	0.51 (0.61)
Traffic congestion is present	0.26	0.35 (0.48)	0.24 (0.43)	0.32 (0.47)	0.59 (0.49)	0.36 (0.48)
Sidewalks are present	-0.26	1.00 (0.00)	1.00 (0.00)	0.60 (0.49)	0.76 (0.43)	0.99 (0.11)
Home size (1000 square feet)	-0.18	1.37 (0.81)	1.84 (0.83)	1.55 (0.45)	1.35 (0.61)	1.69 (0.38)
Have own backyard	-0.15	0.47 (0.50)	0.93 (0.25)	0.97 (0.17)	0.54 (0.50)	0.97 (0.17)
Enough parking available near home	0.12	0.48 (0.50)	0.77 (0.42)	0.89 (0.32)	0.76 (0.43)	0.91 (0.29)
Number of parking spaces for HH use	0.11	1.43 (1.04)	2.23 (1.15)	4.06 (2.08)	2.83 (3.24)	4.02 (1.48)
Good public transit	-0.10	0.98 (0.14)	0.94 (0.23)	0.86 (0.34)	0.91 (0.28)	0.72 (0.45)
Population density[2] (1 = high, 0 = low)	-0.05	1.00	1.00	0.00	1.00	0.00
Streets are pleasant for walking	0.03	0.84 (0.37)	0.90 (0.30)	0.91 (0.29)	0.86 (0.35)	0.95 (0.23)
Public transit is convenient	0.02	0.97 (0.18)	0.92 (0.27)	0.95 (0.23)	0.98 (0.14)	0.63 (0.49)

[1] The characteristics are ranked by the magnitudes of their loadings on the suburban dimension.
[2] Site-based aggregate characteristics had a standard deviation of zero.

Table 31: Ranking of Factor Loadings with Mean Neighborhood Values, Dependent Variable for 2 Factor, Disaggregate Structure (Traditional)

Characteristic[1]	Loading	Mean Value of Characteristic (standard deviation)[2]				
		NSF	SSF	CON	PH	SJ
Population density[2] (1 = high, 0 = low)	0.72	1.00	1.00	0.00	1.00	0.00
Have own backyard	-0.67	0.47 (0.50)	0.93 (0.25)	0.97 (0.17)	0.54 (0.50)	0.97 (0.17)
Number of parking spaces for HH use	-0.62	1.43 (1.04)	2.23 (1.15)	4.06 (2.08)	2.83 (3.24)	4.02 (1.48)
Enough parking available near home	-0.50	0.48 (0.50)	0.77 (0.42)	0.89 (0.32)	0.76 (0.43)	0.91 (0.29)
Home size (1000 square feet)	-0.39	1.37 (0.81)	1.84 (0.83)	1.55 (0.45)	1.35 (0.61)	1.69 (0.38)
Public transit is convenient	0.39	0.97 (0.18)	0.92 (0.27)	0.95 (0.23)	0.98 (0.14)	0.63 (0.49)
Good public transit	0.38	0.98 (0.14)	0.94 (0.23)	0.86 (0.34)	0.91 (0.28)	0.72 (0.45)
Distance to closest park (mi.)	0.33	0.51 (0.79)	0.65 (0.77)	0.70 (0.60)	1.45 (1.21)	0.68 (0.52)
Sidewalks are present	0.26	1.00 (0.00)	1.00 (0.00)	0.60 (0.49)	0.76 (0.43)	0.99 (0.11)
Streets are pleasant for walking	-0.25	0.84 (0.37)	0.90 (0.30)	0.91 (0.29)	0.86 (0.35)	0.95 (0.23)
Traffic congestion is present	0.25	0.35 (0.48)	0.24 (0.43)	0.32 (0.47)	0.59 (0.49)	0.36 (0.48)
Cycling is pleasant	-0.23	0.63 (0.48)	0.49 (0.50)	0.90 (0.29)	0.94 (0.24)	0.84 (0.37)
Bike paths are present	-0.10	0.31 (0.46)	0.04 (0.20)	0.79 (0.41)	0.91 (0.29)	0.44 (0.50)
Speed limit of roads[2] (mph)	-0.09	25.2	25.3	25.5	25.8	25.5
Level of grid-like street network[2]	-0.09	1.00	0.00	0.50	0.00	0.50
Distance to nearest public transit (mi.)	-0.06	0.24 (0.52)	0.28 (0.48)	0.25 (0.23)	0.35 (0.50)	0.51 (0.61)
Distance to nearest grocery store (mi.)	0.05	0.45 (0.50)	0.64 (0.56)	0.77 (0.57)	1.06 (0.79)	0.95 (0.66)
Distance to nearest gas station (mi.)	-0.02	0.47 (0.74)	0.91 (0.67)	0.72 (0.57)	0.82 (0.55)	0.90 (0.84)

[1] The characteristics are ranked by the magnitudes of their loadings on the traditional dimension. [2] Site-based aggregate characteristics had a standard deviation of zero.

Together, the two factors explain 28.2% of the variation in the data, indicating that most of the 18 traits analyzed have a sizeable amount of variation unique to that trait rather than common to the other traits. This two-factor solution is a rotated factor solution, as is common practice to improve interpretability. The oblique rotation option was selected as exhibiting the cleanest factor structure; the correlation between the 2 factors is -0.066. The suburban disaggregate factor presented in Table 30 explained 15.2% of the total variation in the 18 neighborhood characteristics. Characteristics such as "distance to nearest grocery store" and "distance to nearest park" had strong positive loadings on this factor, indicative of suburban neighborhoods with low mixed use. Further, "level of grid-like street network", a characteristic commonly associated with traditional neighborhoods, had a high, negative loading on the suburban disaggregate factor. In short, the traits loading positively on this factor are especially characteristic of suburban neighborhoods, and hence, the name suburban. As expected, the three suburban neighborhoods had highest positive factor score means, while North and South San Francisco (the traditional neighborhoods) had large, negative factor score means; this lends support to the validity of the suburban factor score. This measure of neighborhood type is the basis for one of the two dependent variables of residential choice used in the regression models of Chapter 7 and the structural equation models presented in Chapter 8. The traditional disaggregate factor presented in Table 31 explained 13.0% of the variance in the 18 neighborhood characteristics. Characteristics that are strongly positively associated with this factor include “population density” and “public transit is convenient”, both of which have been linked with traditional neighborhoods in other studies (see, e.g., Kitamura *et al*., 1997). Further, traits commonly associated with suburban neighborhoods such as “number of parking spaces” and “have own backyard” had large, negative loadings on the traditional factor. As expected, North San Francisco had the highest positive traditional factor score mean, while San Jose had the highest negative traditional factor score mean. This measure of neighborhood type is the basis for one of the two dependent variables of residential choice used in the regression models of Chapter 7 and the structural equation models presented in Chapter 8.

A look at both dimensions together is revealing. In particular, studying the mean factor scores (see Figure 5) by neighborhood shows that distinguishing neighborhood type may not be straightforward. First, San Jose (a neighborhood believed to be highly suburban) had a mean value near zero for the suburban dimension, a value indicative of a neighborhood that is a mixture of both traditional and suburban characteristics. Second, Pleasant Hill had the highest positive mean factor score for the suburban dimension while also possessing the second highest positive mean factor score for the traditional dimension.

To get a better understanding of the variation within and overlap between neighborhoods along these two factor dimensions, two plots were developed - Figure 5, a plot of the suburban and traditional mean factor scores by neighborhood, and Figure 6, a plot of disaggregate factor scores for each individual in the sample distinguished by their neighborhood of residence. For Figure 6, the factor score “centroids” for each neighborhood (i.e., an X,Y point where the horizontal coordinate X is the mean for the suburb factor score and the vertical coordinate Y is the mean for the traditional factor

score) are denoted by letter on the plot and are projected onto both axes to help see how the neighborhoods are ranked on each dimension. The plot illuminates several important points. First, one can see that North San Francisco aligns very clearly on both dimensions, indicating a strong level of traditionalness by both measures. South San Francisco is also traditional by both measures, although not as strongly as North San Francisco. There is no corresponding neighborhood that aligns as strongly on the suburban side of both dimensions as North San Francisco does on the traditional side. This suggests greater diversity than has previously been acknowledged in the literature as to what constitutes “suburbanness”. San Jose and Concord are about equal in terms of both representing a suburban measure of neighborhood, though neither comes close to the high mean factor score that Pleasant Hill has on the suburban dimension. Pleasant Hill also scores quite high on the positive side of the traditional measure, illustrating a neighborhood that is a blend of both dimensions, that is, possessing both traditional and suburban characteristics. This is also shown in the high variability of the individual factor scores shown in Figure 6. To summarize, Figure 6 shows quite clearly the folly of attempting to characterize the type of an entire neighborhood in terms of a single binary variable. First, at least two dimensions appear to be important, and neighborhoods can fall on each dimension independent of the other. Second, the range and variation of characteristics that define a neighborhood are more aptly modeled as continuous than binary. Third, individuals within the same neighborhood can have vastly different values for neighborhood type.

6.7 A Comparison of the Four Neighborhood Measures

Before introducing statistical models based on the neighborhood dimension variables (see Chapters 7 and 8), it is important to compare the similarities and differences among the four neighborhood measures presented thus far. The most logical place to start would be to look at what neighborhood characteristics are the primary determinants of each of the four factors. Figures 7-10 show the means for the five neighborhoods of the nine highest-loading (standardized) characteristics on each factor. The means plotted in Figures 7-10 are the same as those presented in Tables 28-31, except that, to facilitate the graphical presentation of variables measured on disparate scales, each value was standardized by subtracting the mean and dividing by the standard deviation of the five neighborhood means on that characteristic. An inspection of these figures can provide insight into what and how each neighborhood measures on those characteristics.

The mean values for the highest-loading characteristics for the single-factor structures (aggregate and disaggregate) are discussed first. For the aggregate solution, NSF had the lowest means on negatively-loading traits and the highest on positively loading traits for most of the top ranked characteristics (such as " enough parking available near home" and "good public transit"), giving them the highest magnitude factor score means, while the reverse tended to be true for SJ. In the disaggregate solution, however, while the pattern for NSF and SSF still holds on the positive side, it is now PH tending to have the highest means on negatively loading traits and the lowest means on positive ones. Thus, we see that while NSF, SSF, and Concord are fairly

consistent across the 18 traits, SJ and PH are more heterogeneous. SJ is more "suburban" than PH (see Figure 7) on traits such as parking availability, relative lack of transit services, and population density, while PH is more suburban than SJ on traits such as having higher speed limits and not having a grid-like street network.

It can be seen , then, that given the same neighborhoods and the same variable characteristics, the use of aggregate and disaggregate data yield different results. This finding has serious consequences for modeling residential choice. In Chapter 5 a binary model of residential choice was described, where NSF was the traditional alternative and SJ and CON were the suburban alternatives. This was at least partly supported by the 1-factor aggregate structure, as NSF had the highest mean factor score and SJ and CON had the lowest mean factor scores. On the other hand, the 1-factor disaggregate structure would suggest using PH or CON as one of the suburban alternatives. Either way, modeling results would likely be different, and thus, conclusions based on the models would need to be more cautiously viewed.

Turning now to the two-factor disaggregate structures, we find that objectively measured characteristics were dominant in the formation of the neighborhood dimensions, having at least the top three loadings for both the suburban and traditional factor dimensions (see Figures 9 and 10). For example, the neighborhood characteristic "speed limit of road" had the loading with the greatest magnitude for suburb (0.84) and the characteristic "population density" had the highest loading for traditional (0.72). Further, though subjective measures such as "cycling is pleasant" and "public transit is convenient" were important in the defining of the suburb and traditional dimensions (each being in the top 9 of the 18 factor loadings), they were few in number compared to the objective characteristics. Next, parking, transit, and distance to places were three main characteristics found to be heavily weighted in the creation of the neighborhood measures. This finding is significant in that it supports the main utility of using a data reduction technique such as factor analysis to group correlated characteristics into a representative dimension. For example, four characteristics relating to distance to a destination (such as a park or a grocery store) were in the top nine loadings for the suburban factor (all with a positive loading, indicating that greater distances are more representative of suburbs than of traditional neighborhoods), while two characteristics related to parking were in the top four loadings for traditional.

To conclude, an analysis of the mean factor score ordering is given. First, the traditional dimension of the 2-factor structure has a mean factor score neighborhood ordering (traditional - NSF, PH, SSF, CON, SJ - suburban) that is close to the same ordering as the 1-factor aggregate (traditional - NSF, SSF, PH, CON, SJ - suburban). The neighborhoods that represent the two extremes are the same (i.e., NSF is the most traditional neighborhood and SJ is the most suburban neighborhood), and only PH and SSF switch ordering. Next, the suburban dimension of the 2-factor structure has the exact same ordering as the 1-factor disaggregate structure (traditional - NSF, SSF, SJ, CON, PH - suburban). In this case, the neighborhood that is most identified with the suburban dimension is Pleasant Hill. Thus, the aggregate solution seems to have identified one dimension of neighborhood type, while the one-factor disaggregate solution identified the other. Both dimensions are identified by the two-factor disaggregate solution, though, and consequently, these are the factors used in the structural equation models of Chapter 8.

6.8 Summary of Chapter 6

Chapter 6 set the stage for the residential choice modeling conducted in the next two chapters. First, the concept of an endogenous variable for residential choice was introduced. Next, a description of neighborhood definitions found in the literature was presented. Specific challenges in developing a measure to represent a traditional neighborhood, such as choosing what type of data to use (aggregate versus disaggregate, micro scale versus macro scale) and identifying the appropriate characteristics immediately followed. The 18 characteristics used to define neighborhood type in this dissertation were presented next, along with a discussion of the results of three different factor analysis procedures. The ending section compared the four factors (a traditionalness aggregate single factor, a traditionalness disaggregate single factor, and traditional and suburban dimensions from a two-factor solution) obtained from the factor analyses, concluding that the concept of traditionalness versus suburbanness may be better viewed as two different dimensions instead of two ends of the same dimension.

CHAPTER 7

RESIDENTIAL CHOICE MODEL ESTIMATION:

SETTING THE STAGE

7.1 Introduction

Aggregate models of housing choice based on economic theory comprised much of the early work on residential location choice (e.g., Kain, 1962; Alonso, 1964; Muth, 1969). The methodologies for these models were based on the bid rent approach of consumer theory, in which an individual optimizes her or his utility under varying assumptions. The main assumption of the models was that individuals trade off housing costs with travel costs in determining their housing choice.

Much progress has been made since the 1960s on the economic theory described above, and further, different methodologies have been developed that can also be used to analyze residential choice. Advantages and disadvantages of the various research approaches are found in the literature, but the author is not aware of any systematic discussion of what method is superior under what circumstances. One goal of this dissertation is to gain insight into such issues. For example, would an aggregate measure of residential location be a better dependent variable than a disaggregate measure? Answers to these questions are likely to be difficult to obtain, which may be one reason behind the lack of such analysis in past research on residential location. The following sections describe the core modeling elements behind this dissertation, as well as address the question posed above. The second part of this chapter starts the operationalization of the conceptual model, with the estimation of single-equation regression models of the major components of the conceptual model. Issues surrounding the conceptual model specification such as endogeneity and efficiency are then described and addressed through various statistical model implementations. Specifically, the inclusion of endogenous variables as explanatory variables in single-equation models is undertaken, with the discussion of the estimation of a simultaneous two-equation system comprising the ending section.

7.2 Aggregate versus Disaggregate Residential Choice Measures

In Chapter 6, both aggregate and disaggregate measures of residential choice were developed. "Level of traditionalness" was defined both as a composite measure based on individual responses to a specific group of variables (the disaggregate measure), and as a composite measure based on the neighborhood mean responses to the same specific group of variables (the aggregate measure). Given that the analysis unit for our models of residential choice is in any case the individual (i.e., the models themselves are disaggregate, with individual-specific explanatory variables), it was anticipated that the disaggregate measure of the dependent variable would yield better results than the aggregate measure. Residential choice is an individual behavior, which would seem to correspond better with individual measures of chosen neighborhood type. Thus, averaging individual responses across a neighborhood can mask large variations in the level of traditionalness within a neighborhood, potentially making the aggregate measure a less accurate measure of the neighborhood type actually faced by the individual making the residential choice.

Daly (1982), in a study of the applicability of disaggregate models of behavior, concluded that “disaggregate statistical analysis is the most useful technique” in trying to model behavior. Richards (1982) also concludes that disaggregate models offer major improvements over aggregate models in the study of travel behavior. As the primary goals of this work are to understand individual behavior mechanisms related to travel and land use decisions, disaggregate modeling procedures were performed for nearly all of the empirical work in this dissertation. However, the interest in understanding the difference in explanatory power between aggregate and disaggregate measures of the dependent variable, residential choice, motivated the development of empirical models based on each type of measure.

To compare the performance of the two types of dependent variables, single-equation regression models were separately estimated on the N_3 data set (615 employed respondents). Tables 32 and 33 present the results for the aggregate dependent variable model and the disaggregate dependent variable model respectively.

A large number of variables (more than 10 in each model) representing three types of influences (sociodemographic, attitudinal, and lifestyle) were found to be significant in the first set of residential choice models based on the aggregate and disaggregate composite measures of traditionalness. Further, the factor score variables, both attitudinal and lifestyle, greatly contributed to both models. For example, “pro-alternatives” had a positive coefficient in both models, indicating that a person who feels more favorably toward policies that support environmentally-efficient forms of travel will be more likely to live in a traditional neighborhood. Lastly, the coefficients had the expected signs.

Table 32: Residential Choice Model, Aggregate Traditionalness Dependent Variable

Variable	Coefficient	t-statistic (p-value)
Intercept	1.27	7.99 (0.00)
Sociodemographic		
Age	-0.015	-4.90 (0.00)
Professional	-0.19	-3.12 (0.00)
Under Sixteen (number of children in HH)	-0.14	-3.10 (0.00)
Vehicles	-0.16	-5.06 (0.00)
Years Lived in Bay Area	-0.0054	-2.55 (0.01)
Attitudinal		
Pro-Alternatives	0.15	5.00 (0.00)
Pro-Pricing	0.11	3.53 (0.00)
Work-Driven	-0.066	-2.36 (0.02)
Lifestyle		
Adventurer	-0.078	-2.56 (0.01)
Culture-Lover	0.22	6.92 (0.00)
Fun-Seeker	-0.063	-1.98 (0.05)
Homebody	-0.12	-3.74 (0.00)
Nest-Builder	-0.11	-3.49 (0.00)
Traveler	0.086	2.98 (0.00)
Model Statistics		
R-squared	0.380	
Adjusted R-squared	0.366	
Standard Error of Estimation	0.714	
F-Statistic	26.30	
Degrees of Freedom	614	

P-value of F-Statistic	0.000

Table 33: Residential Choice Model, Disaggregate Traditionalness Dependent Variable

Variable	Coefficient	t-statistic (p-value)
Intercept	0.77	3.38 (0.00)
Sociodemographic		
Age	-0.015	-4.23 (0.00)
Education	0.070	2.27 (0.02)
Professional	-0.21	-2.72 (0.01)
Vehicles	-0.13	-3.52 (0.00)
Attitudinal		
Pro-Alternatives	0.13	3.68 (0.00)
Pro-Drive Alone	0.074	2.13 (0.03)
Pro-Pricing	0.13	3.74 (0.00)
Work-Driven	-0.093	-2.79 (0.01)
Lifestyle		
Culture-Lover	0.25	6.59 (0.00)
Fun-Seeker	-0.099	-2.83 (0.01)
Hobbyist	0.099	2.72 (0.01)
Homebody	-0.17	-4.49 (0.00)
Nest-Builder	-0.084	-2.34 (0.00)
Model Statistics		
R-squared	0.275	
Adjusted R-squared	0.259	
Standard Error of Estimation	0.860	
F-Statistic	17.49	

Degrees of Freedom	614
P-value of F-Statistic	0.000

Vehicles, for example, had a negative sign in both models, suggesting that people who owned more vehicles were less likely to choose a traditional neighborhood. While "best" models were developed independently for each type of dependent variable, the two final models shared most of the same explanatory variables. Ten variables were common to both (such as age, pro-pricing, and nest-builder), four variables were unique to the aggregate model (under sixteen, years lived in the Bay Area, traveler, and adventurer), and three variables were unique to the disaggregate model (education, pro-drive alone, and hobbyist).

A comparison of Tables 32 and 33 also showed some unexpected outcomes. It turned out that the residential choice model based on the aggregate dependent variable had a better model fit. Specifically, the explanatory variables explained about 37% of the variation in the aggregate model, while the explanatory variables in the disaggregate model explained only 26% of the variance. One possible explanation for this is the fact that the aggregate measure has less variability to explain, with only 5 different values (a single factor score for each of the five neighborhoods) compared to the 615 different disaggregate values. The possibility was considered that the assumption of normally-distributed disturbances was not met for the aggregate model, which would cause the model results and the comparison to be invalid. However, an inspection of the standardized residuals (both a histogram and normal probability plot) indicated that the aggregate model met this key assumption.

In addition to having a better model fit, the explanatory variable structure based on the aggregate measure appeared superior. Models were estimated on the "best" explanatory variable structure found from the aggregate and disaggregate models with the dependent variables switched (i.e., the aggregate dependent variable was regressed against the statistically significant explanatory variables found from the disaggregate model, and conversely). The aggregate dependent variable model was more robust in the sense that only two variables from the disaggregate dependent variable model (pro-drive alone and hobbyist) became insignificant when regressed against the aggregate dependent variable, while four variables from the aggregate dependent variable model (under sixteen, years in Bay Area, traveler, and adventurer) became insignificant when regressed against the disaggregate dependent variable.

The fact that the proportion of variance explained (R^2) for each model is probably confounded with the differing amounts of variability in their respective dependent variables illustrates that a comparison of the aggregate and disaggregate approaches to the dependent variable measurement is not straightforward. While this author is not prepared to pronounce the aggregate approach unequivocally superior due to its higher R^2 and apparently more robust structure, it does appear that the aggregate measurement of neighborhood type is not as deficient as had been feared.

Nevertheless, disaggregate measures of residential choice will be used for the remainder of this dissertation, for both conceptual and statistical reasons. First, it was found that residential neighborhood type comprised two dimensions, suburban and traditional, which was only captured via disaggregate measures. More importantly, it is believed that residential choice is innately individual rather than aggregate behavior, and thus, better modeled with individual measures. As argued before, the traditional or suburban nature of a neighborhood can vary widely over a relatively small area, and the individual's choice is likely to be based on the specific characteristics s/he faces, not a general measure for the entire area. The next section presents the first step toward the operationalization of the residential choice conceptual model. Specifically, single-equation regression models for each of the available endogenous variables of the conceptual model will be developed, initially using only exogenous variables as predictors.

7.3 Exogenous-Explanatory-Variable-Only Single-Equation Models

A logical first step toward developing an interdependent system of equations relating to residential choice/preference was the estimation of single-equation regression models of endogenous variables derived from this dissertation's conceptual model. This modeling strategy was chosen due to its simplicity and ability to effectively uncover variables that are significantly associated with a dependent variable. In essence, this was the first empirical attempt of this study to identify variables that would be significant in the full-scale conceptual model that represented multiple interdependent relationships.

After a fair amount of trial-and-error (primarily relating to the requirements for identification of a system of SEMs, as discussed in Chapter 8), nine endogenous variables representing different elements of the conceptual model were identified for estimation purposes: traditional and suburban factor scores (representing residential location), pro-high density, pro-driving, and pro-transit factor scores (representing respondent attitudes), vehicle miles, transit miles, and walk/bike miles (representing travel demand), and commute distance (representing job location). The following subsections discuss the results of each regression estimation. The model assumptions for regression analysis (such as homoskedasticity and normality of e) were checked (via methods such as residual analysis) and were met for each of the following models, indicating that the estimated coefficients possess OLS properties such as unbiasedness. Further, all of the models were estimated on the cleaned (no missing observations) data sample $N_3 = 615$, using procedures of stepwise regression and manual entering via SPSS 9.0 for Windows. Chapter 5 contains descriptions of all of the variables seen below.

7.3.1 Regression Model of Residential Choice - Traditional

A measure of residential choice/preference representing how closely a neighborhood is characterized by traditional qualities, the "traditional" factor score from the two-factor solution (see Section 6.6.2), was first modeled. Table 34 contains the model output for this estimation.

Inspection of the table shows that nearly 40% of the variation in the endogenous variable traditional is explained by the exogenous variables - a respectable proportion for a disaggregate model such as this, and incidentally, higher than the percent explained by the aggregate measure modeled in Section 7.2. Especially in view of the observation that the aggregate measure has the advantage of having less variability to explain, the fact that a model with one of the two disaggregate dependent variables (from the two-factor solution) has a higher R^2 suggests that the more fine-grained measurement permitted by the disaggregate approach (the only one that allowed more than one factor to be

Table 34: Single-Equation Regression Model - Traditional

Variable	Coefficient	t-statistic (p-value)
Intercept	1.39	6.28 (0.00)
Sociodemographic		
Age	-0.013	-3.59 (0.00)
Education	0.050	1.88 (0.06)
Household Size	-0.15	-4.48 (0.00)
Number of Vehicles	-0.254	-6.57 (0.00)
Years Lived in the Bay Area	-0.0052	-2.20 (0.03)
Attitudinal		
Pro-Pricing	0.071	2.14 (0.03)
Lifestyle		
Culture-Lover	0.19	5.34 (0.00)
Fun-Seeker	-0.081	-2.42 (0.02)
Nest-Builder	-0.23	-7.00 (0.00)
Model Statistics		
R-squared	0.396	

Adjusted R-squared	0.387
Standard Error of Estimation	0.797
F-Statistic	44.12
Degrees of Freedom	614
P-value of F-Statistic	0.000

identified) does have a statistical advantage over the aggregate approach. Common sociodemographic variables such as household size and number of vehicles played a large role in the model, but attitudinal and lifestyle variables were also well-represented. The signs of the estimated coefficients matched prior hypotheses. For example, the negative sign for household size indicated that respondents from larger households were less likely to have chosen a traditional neighborhood. This outcome makes sense in that traditional represents a higher-density area, a type of location that may not be as desirable for large families.

7.3.2 Regression Model of Residential Choice - Suburban

A measure of residential choice/preference representing how closely a neighborhood is characterized by suburban qualities, the “suburban” factor of the two-factor disaggregate solution, was modeled next. Table 35 contains the results of this estimation.

The much lower adjusted R-squared value, 0.105, reflects how much weaker the exogenous variables are at explaining the variation in the residential choice/preference conceptual variable suburban as compared to traditional. Interestingly, only one sociodemographic variable, age, was significant in this model. As before, the signs of the estimated coefficients matched prior expectations. For example, the negative signs for pro-alternatives and pro-pricing indicated that individuals who chose suburban neighborhoods were less likely to support policies related to reducing the impact of driving.

Table 35: Single-Equation Regression Model - Suburban

Variable	Coefficient	t-statistic (p-value)
Intercept	-0.50	-2.63 (0.01)
Sociodemographic		
Age	0.011	2.69 (0.01)
Attitudinal		
Pro-Alternatives	-0.15	-3.62 (0.03)
Pro-Pricing	-0.11	-2.75 (0.01)
Work-Driven	0.070	1.91 (0.06)

Lifestyle		
Athlete	0.079	2.13 (0.03)
Culture-Lover	-0.15	-3.71 (0.00)
Hobbyist	-0.11	-2.75 (0.01)
Model Statistics		
R-squared	0.115	
Adjusted R-squared	0.105	
Standard Error of Estimation	0.953	
F-Statistic	11.25	
Degrees of Freedom	614	
P-Value of F-Statistic	0.000	

7.3.3 Regression Model of Attitudes & Lifestyle - Pro-High Density

The next three estimated regression models were based on individual attitudinal measures: pro-high density, pro-driving, and pro-transit. These measures represent the attitudes and lifestyle component of the conceptual diagram. Table 36 contains the results of the model for pro-high density.

Like the suburban regression model, pro-high density had a relatively low adjusted R-squared value (0.136). Unfortunately, it had the best model fit among the three regression models based on attitudinal variables, reflecting the complexity of explaining the variation of individual attitudes. Sociodemographic and lifestyle variables composed most of the exogenous variables. The signs of the estimated coefficients generally matched expectations. Household income's positive impact on pro-high density suggested that people with higher incomes were more likely to approve policies supportive of high-density development; this is potentially a reflection of the ability to pay for a high-density residence, a residence which generally will be more costly than a similar residence in a low-density suburb. The opposite signs of adventurer and outdoor enthusiast in this model provided more information about the difference between these two lifestyle measures. The negative coefficient for adventurer suggests that people who enjoy the act of traveling (supported by the positive coefficient in the pro-driving model presented next) in low-density space (such as using an off-road vehicle, or driving motor cross) will be less likely to support policies encouraging high-density development. The positive coefficient for outdoor enthusiast (defined by statements such as "visited a local or state park" and "went hiking") may be a reflection of the interest of people who choose Table 36: Single-Equation Regression Model - Pro-High Density

Variable	Coefficient	t-statistic (p-value)
Intercept	0.16	0.69 (0.49)
Sociodemographic		
Female	-0.23	-2.82 (0.01)
Household Income	0.11	3.29 (0.00)
Household Size	-0.23	-6.30 (0.00)
Years Lived in the Bay Area	-0.0066	-2.54 (0.01)
Attitudinal		
Time-Satisfied	0.096	2.18 (0.03)
Lifestyle		
Adventurer	-0.12	-2.97 (0.00)
Nest-Builder	-0.19	-4.72 (0.00)
Outdoor Enthusiast	0.12	3.00 (0.00)
Model Statistics		
R-squared	0.148	
Adjusted R-squared	0.136	
Standard Error of Estimation	0.959	

F-Statistic	13.11
Degrees of Freedom	614
P-Value of F-Statistic	0.000

to live in high-density areas to be able to enjoy outdoor activities that are an escape from day-to-day life. These findings suggest that individuals who score highly on adventurer are less likely to enjoy high-density living (either in day-to-day activities like work, or in leisure), while outdoor enthusiasts are people who can and do enjoy living in high-density areas.

7.3.4 Regression Model of Attitudes & Lifestyle - Pro-Driving

An individual's attitude toward driving was modeled next. The regression model based on the endogenous variable pro-driving had the second highest fit of the three models, explaining nearly 10% of the variance in the dependent variable. Table 37 contains the results of this estimation.

The model for pro-driving was similar to the one for pro-high density in that both only had one attitude variable and a few variables of the lifestyle and sociodemographic types.

Although the model goodness-of-fit was low, the six significant variables it contained were conceptually sound. For example, the positive sign of the adventurer variable (a lifestyle measure) suggests that, at least in this California sample, achievement of an adventurous lifestyle is most typically associated with the appreciation of the benefits of driving an automobile. Thus, this result also supports the idea that lifestyle traits can influence (or be influenced by) individual attitudes.

7.3.5 Regression Model of Attitudes & Lifestyle - Pro-Transit

The last attitudinal endogenous variable to be modeled was pro-transit. It is

Table 37: Single-Equation Regression Model - Pro-Driving

Variable	Coefficient	t-statistic (p-value)
Intercept	-0.48	-5.38 (0.00)
Sociodemographic		
Female	0.47	5.81 (0.00)
Years Lived in the Bay Area	0.0095	3.69 (0.00)
Attitudinal		
Time-Satisfied	-0.112	-2.55 (0.01)
Lifestyle		

Adventurer	0.15	3.58 (0.00)
Nest-Builder	0.077	1.96 (0.05)
Relaxer	0.082	2.18 (0.03)
Model Statistics		
R-squared	0.095	
Adjusted R-squared	0.086	
Standard Error of Estimation	0.967	
F-Statistic	10.69	
Degrees of Freedom	614	
P-Value of F-Statistic	0.000	

expected that people who support policies designed to reduce the impact of human activity on the environment will be more likely to be pro-transit. Table 38 presents the results of this estimation.

Only three exogenous variables were found to be significant in explaining the pro-transit attitude. The signs of two of the three variables are intuitive: a pro-drive alone attitude is negatively associated with pro-transit, and a pro-environmental attitude is positively associated. The positive association of manager status with a pro-transit attitude may be logical if managers are more likely to use upscale forms of transit (e.g., commuter trains) in their commutes. Manager was not significant in the structural equation models of Chapter 8, suggesting that the indirect relationship it is representing here (in Chapter 7) may be accounted for more directly. Overall, however, the model fit was poor (adjusted R-squared of 0.062); this was more evidence of the complexity of modeling individual attitudes.

7.3.6 Regression Model of Travel Demand - Vehicle-Miles

The next three regression models were based on travel demand endogenous variables: vehicle-miles per day, transit-miles per day, and walk/bike-miles per day. Though travel demand is a complex subject, it was expected that better results would be obtained for these three models than for the previous three attitude models. Table 39 presents the model for daily vehicle-miles.

The exogenous variables, all of which had the expected sign, explained about 13% of the variance in a respondent's demand for vehicle miles. The sociodemographic

Table 38: Single-Equation Regression Model - Pro-Transit

Variable	Coefficient	t-statistic (p-value)
Intercept	-0.057	-1.28 (0.20)
Sociodemographic		
Manager	0.24	2.30 (0.02)
Attitudinal		
Pro-Drive Alone	-0.15	-3.89 (0.00)
Pro-Environment	0.18	4.43 (0.00)
Lifestyle		
No Significant Variables		
Model Statistics		
R-squared	0.067	
Adjusted R-squared	0.062	
Standard Error of Estimation	0.982	
F-Statistic	14.53	
Degrees of Freedom	614	
P-Value of F-Statistic		

	0.000

Table 39: Single-Equation Regression Model - Vehicle-Miles

Variable	Coefficient	t-statistic (p-value)
Intercept	9.36	1.54 (0.13)
Sociodemographic		
Female	-7.33	-3.30 (0.00)
Household Income	2.01	2.26 (0.02)
Number of Vehicles	5.42	4.88 (0.00)
Attitudinal		
Pro-Alternatives	-2.63	-2.36 (0.02)
Lifestyle		
Adventurer	4.11	3.65 (0.00)
Model Statistics		
R-squared	0.135	
Adjusted R-squared	0.128	
Standard Error of Estimation	26.41	
F-Statistic	19.07	
Degrees of Freedom	614	

P-Value of F-Statistic	0.000

variables that were significant are quite typical of those in trip generation models; other typical variables such as household size, number of children, and number of workers were also tried and not found to be significant. It is likely that the number of vehicles variable that is in the final model is capturing the explanatory power offered by those other indicators of household size. Supporting the previous model (pro-driving), people who had high values for the adventurer factor score were more likely to have high values for vehicle miles.

7.3.7 Regression Model of Travel Demand - Transit-Miles

The next travel demand regression model was estimated on the endogenous variable transit-miles. This regression model had the second highest fit of the three travel-demand models. Table 40 presents the results for this estimation.

Four variables were found to be significant in this model. Pro-growth was negatively related with transit miles; the idea being that people who want growth are more interested in the low-density development (i.e., sprawl) that is associated with lower levels of transit use. Similarly, the more vehicles possessed by the HH, the fewer the transit-miles traveled by the respondent. The manager coefficient was not expected to be positive a priori, but the same argument applies here as for its appearance with a positive sign in the pro-drive alone attitude model (Section 7.3.5). Pro-drive alone was expected to have a negative sign, indicating that people who prefer to drive alone would be less likely to use transit. The fact that it has a positive sign suggests an interaction effect with another variable. One possible explanation is that pro-drive alone respondents

Table 40: Single-Equation Regression Model - Transit-Miles

Variable	Coefficient	t-statistic (p-value)
Intercept	9.42	8.13 (0.00)
Sociodemographic		
Manager	2.90	2.15 (0.03)
Number of Vehicles	-2.50	-4.81 (0.00)
Attitudinal		
Pro-Drive Alone	1.32	2.57 (0.01)
Pro-Growth	-1.89	-3.61 (0.00)
Lifestyle		
No Significant Variables		
Model Statistics		
R-squared	0.071	
Adjusted R-squared	0.065	
Standard Error of Estimation	12.91	
F-Statistic	11.67	

Degrees of Freedom	614
P-Value of F-Statistic	0.000

may be heavy users of both line haul transit for commuting long distances, and primarily automobile users for work-related and non-work travel. It should be noted that no lifestyle variables were found to be significant in this model. This may mean that the lifestyle activities identified in this study are more likely to involve auto or non-motorized modes of travel than transit.

7.3.8 Regression Model of Travel Demand - Walk/Bike-Miles

The last travel demand regression model was estimated on the endogenous variable walk/bike-miles. Walk/bike-miles is a measure of how much an individual uses these non-motorized modes of travel. Table 41 presents the model for walk/bike-miles. The four explanatory variables that were significant in the model had logical signs. In particular, a pro-alternatives attitude was positively associated with walk/bike-miles, indicating that people who supported transportation alternatives in fact used some of these alternative modes of travel. The walk/bike regression model had the lowest adjusted R-squared value out of all of the models (0.023), meaning that many important predictors of this indicator were unmeasured and possibly idiosyncratic. Further, though this dependent variable "passed" the regression assumptions, it definitely had the largest skewness of the dependent variables used in this study (many zero values were present and some very large values).

7.3.9 Regression Model of Job Location - Commute Distance

Only one variable for job location was available for this dissertation. Commute

Table 41: Single-Equation Regression Model - Walk/Bike-Miles

Variable	Coefficient	t-statistic (p-value)
Intercept	1.41	2.80 (0.01)
Sociodemographic		
Age	-0.018	-1.68 (0.10)
Attitudinal		
Pro-Alternatives	0.272	2.47 (0.01)
Pro-Growth	-0.211	-1.96 (0.05)

Lifestyle		
Nest-Builder	-0.25	-2.32 (0.02)
Model Statistics		
R-squared	0.029	
Adjusted R-squared	0.023	
Standard Error of Estimation	2.64	
F-Statistic	4.63	
Degrees of Freedom	614	
P-Value of F-Statistic	0.001	

distance was used as a proxy for job location in the conceptual model. Table 42 contains the results of the model for commute distance.

The exogenous variables explained about 6% of the endogenous variable variance. Only attitudinal and sociodemographic variables were significant in this model; all with the expected signs. Results showed that females, on average, had lower commute distances (consistent with the literature, see, e.g., White, 1977), and individuals who were pro-drive alone were more likely to have longer commute distances than those who had low scores on pro-drive alone.

7.4 Regression Models with Endogenous Explanatory Variables

Up to this point, the explanatory variables selected for inclusion in each model have been assumed to be independent of each other and of the error term for that model. Failure to meet this requirement violates the assumptions necessary for ordinary least squares (OLS) estimates to be statistically consistent (Greene, 1997). The inclusion of endogenous variables on the right-hand-side of an equation is one common way in which this requirement fails to be met. However, since a major thesis of this study is that the endogenous variables are interrelated through the conceptual model shown in Figure 1, it is clearly incomplete to exclude endogenous variables as predictors. Hence, as the second stage of identifying significant variables for the structural model, models for the previously identified endogenous variables will now include endogenous variables

as predictors. Since this procedure clearly violates OLS assumptions, the specific results obtained for these models should not be overemphasized; they are primarily a

Table 42: Single-Equation Regression Model - Commute Distance

Variable	Coefficient	t-statistic (p-value)
Intercept	5.19	1.97 (0.05)
Sociodemographic		
Female	-2.79	-2.97 (0.00)
Household Income	1.25	3.31 (0.00)
Attitudinal		
Pro-Alternatives	-1.11	-2.32 (0.02)
Pro-Drive Alone	1.20	2.60 (0.01)
Pro-Pricing	-1.12	-2.48 (0.02)
Lifestyle		
No Significant Variables		
Model Statistics		
R-squared	0.061	
Adjusted R-squared	0.054	
Standard Error of Estimation	11.50	

F-Statistic	7.95
Degrees of Freedom	614
P-Value of F-Statistic	0.000

stepping-stone to the more appropriate structural equation models presented in Chapter 8. However, it should be noted in passing that many single-equation models published in the literature are subject to the same endogeneity bias. Tables 43 through 51 present the best models for the same nine endogenous variables, where each of those variables is now allowed to enter each of the other eight equations as a potential predictor. A general comparison of the two sets of models is given next.

Table 43[1]: Regression Model with Endogenous Explanatory Variables - Traditional

Variable	Coefficient	t-statistic (p-value)
Intercept	1.78	9.94 (0.00)
Sociodemographic		
Age	-0.014	-3.98 (0.00)
Household Size	-0.18	-5.43 (0.00)
Number of Vehicles	-0.23	-5.82 (0.00)

Years Lived in the Bay Area	-0.0058	-2.50 (0.01)
Attitudinal		
Pro-High Density (endogenous)	0.19	5.93 (0.00)
Pro-Pricing	0.11	3.30 (0.00)
Lifestyle		
Nest-Builder	-0.18	-5.55 (0.00)
Travel Demand		
Vehicle Miles (endogenous)	-0.0033	2.77 (0.01)
Model Statistics		
R-squared	0.400	
Adjusted R-squared	0.392	

[1] Statistics for this model are presented for completeness, but cannot be considered reliable since OLS assumptions are violated by the inclusion of endogenous variables as explanatory.

Table 44[1]: Regression Model with Endogenous Explanatory Variables - Suburban

Variable	Coefficient	t-statistic (p-value)
Intercept	-0.69	-3.71 (0.00)
Sociodemographic		
Age	0.011	2.85 (0.00)
Attitudinal		
Pro-Alternatives	-0.12	-3.06 (0.00)
Pro-High Density (endogenous)	-0.044	-1.69 (0.09)
Pro-Pricing	-0.010	2.62 (0.01)
Work-Driven	0.079	2.23 (0.03)
Lifestyle		
Athlete	0.068	1.90 (0.06)
Culture-Lover	-0.15	-3.69 (0.00)
Hobbyist	-0.12	-3.08 (0.00)
Homebody	0.17	4.30 (0.00)
Job Location		
Commute Distance (endogenous)	0.017	5.26 (0.00)
Model Statistics		
R-squared	0.180	
Adjusted R-squared	0.166	

[1] Statistics for this model are presented for completeness, but cannot be considered reliable since OLS assumptions are violated by the inclusion of endogenous variables as explanatory.

Table 45[1]: Regression Model with Endogenous Explanatory Variables- Pro-High Density

Variable	Coefficient	t-statistic (p-value)
Intercept	0.17	2.89 (0.00)
Sociodemographic		
Female	-0.14	-1.82 (0.07)
Number of People in Home Under 16 Years of Age	-0.18	-3.61 (0.00)
Attitudinal		
Pro-Driving (endogenous)	-0.099	-2.63 (0.01)
Work-Driven	0.15	4.25 (0.00)
Time-Satisfied	0.092	2.20 (0.03)
Lifestyle		
Culture-Lover	0.19	4.60 (0.00)
Nest-builder	-0.13	-3.24 (0.00)
Outdoor-Enthusiast	0.11	2.84 (0.01)
Residential Choice		
Traditional (endogenous)	0.21	5.10 (0.00)
Travel Demand		
Walk/Bike Miles	0.043	3.06 (0.00)
Model Statistics		
R-squared	0.235	
Adjusted R-squared	0.222	

[1] Statistics for this model are presented for completeness, but cannot be considered reliable since OLS assumptions are violated by the inclusion of endogenous variables as explanatory.

Table 46[1]: Regression Model with Endogenous Explanatory Variables - Pro-Driving

Variable	Coefficient	t-statistic (p-value)
Intercept	-0.56	-5.61 (0.00)
Sociodemographic		
Female	0.43	5.40 (0.00)
Years Lived in Bay Area	0.084	3.24 (0.00)
Attitudinal		
Pro-High Density (endogenous)	-0.11	-2.94 (0.00)
Time-Satisfied	-0.11	-2.45 (0.02)
Lifestyle		
Relaxer	0.088	2.35 (0.02)
Travel Demand		
Vehicle Miles	0.0052	3.74 (0.00)
Model Statistics		
R-squared	0.101	
Adjusted R-squared	0.092	

[1] Statistics for this model are presented for completeness, but cannot be considered reliable since OLS assumptions are violated by the inclusion of endogenous variables as explanatory.

Table 47[1]: Regression Model with Endogenous Explanatory Variables - Pro-Transit

Variable	Coefficient	t-statistic (p-value)
Intercept	-0.11	-2.67 (0.01)
Sociodemographic		
No Significant Variables		
Attitudinal		
Pro-Drive Alone	0.12	3.24 (0.00)
Pro-Environment	0.17	4.49 (0.00)
Lifestyle		
No Significant Variables		
Travel Demand		
Transit Miles	0.020	6.78 (0.00)

[1] Statistics for this model are presented for completeness, but cannot be considered reliable since OLS assumptions are violated by the inclusion of endogenous variables as explanatory.

Table 48[1]: Regression Model with Endogenous Explanatory Variables - Vehicle Miles

Variable	Coefficient	t-statistic (p-value)
Intercept	14.08	5.41 (0.00)
Sociodemographic		
Female	-6.56	-3.33 (0.00)
Number of Vehicles	4.30	4.56 (0.00)
Attitudinal		
Pro-Driving (endogenous)	1.59	1.70 (0.09)
Lifestyle		
Adventurer	2.56	2.66 (0.01)
Traveler	1.82	2.02 (0.04)
Job Location		
Commute Distance (endogenous)	1.22	14.56 (0.00)
Travel Demand		
Transit Miles (endogenous)	-0.69	-9.29 (0.00)
Walk/Bike Miles (endogenous)	-0.78	-2.27 (0.02)
Model Statistics		
R-squared	0.101	

Adjusted R-squared	0.092

[1] Statistics for this model are presented for completeness, but cannot be considered reliable since OLS assumptions are violated by the inclusion of endogenous variables as explanatory.

Table 49[1]: Regression Model with Endogenous Explanatory Variables - Transit Miles

Variable	Coefficient	t-statistic (p-value)
Intercept	3.84	3.00 (0.00)
Sociodemographic		
Number of Vehicles	-1.75	-3.26 (0.00)
Attitudinal		
Pro-Driving (endogenous)	-0.98	-2.09 (0.04)
Pro-Growth	-1.39	-2.91 (0.00)
Pro-Transit (endogenous)	2.81	5.94 (0.00)
Lifestyle		
No Significant Variables		
Job Location		
Commute Distance (endogenous)	0.38	9.40 (0.00)
Residential Location		
Traditional (endogenous)	-0.69	2.03 (0.04)
Model Statistics		

R-squared	0.234
Adjusted R-squared	0.23

[1] Statistics for this model are presented for completeness, but cannot be considered reliable since OLS assumptions are violated by the inclusion of endogenous variables as explanatory.

Table 50[1]: Regression Model with Endogenous Explanatory Variables - Walk/Bike Miles

Variable	Coefficient	t-statistic (p-value)
Intercept	0.84	5.40 (0.00)
Sociodemographic		
No Significant Variables		
Attitudinal		
Pro-Alternatives	0.25	2.25 (0.03)
Pro-Driving (endogenous)	-0.19	-1.76 (0.08)
Pro-High Density (endogenous)	0.34	3.32 (0.00)
Lifestyle		
No Significant Variables		
Travel Demand		
Vehicle Miles (endogenous)	-0.090	-2.35 (0.02)
Model Statistics		

R-squared	0.047
Adjusted R-squared	0.041

[1] Statistics for this model are presented for completeness, but cannot be considered reliable since OLS assumptions are violated by the inclusion of endogenous variables as explanatory.

Table 51[1]: Regression Model with Endogenous Explanatory Variables - Commute Distance

Variable	Coefficient	t-statistic (p-value)
Intercept	4.19	6.97 (0.00)
Sociodemographic		
No Significant Variables		
Attitudinal		
No Significant Variables		
Lifestyle		
No Significant Variables		
Residential Choice		
Suburban	1.33	3.53 (0.00)

Traditional	-0.82	-2.19 (0.03)
Travel Demand		
Transit Miles	0.39	13.70 (0.00)
Vehicle Miles	0.20	14.33 (0.00)
Model Statistics		
R-squared	0.392	
Adjusted R-squared	0.388	

[1] Statistics for this model are presented for completeness, but cannot be considered reliable since OLS assumptions are violated by the inclusion of endogenous variables as explanatory.

7.5 Comparison of Regression Models - Exogenous-Only Explanatory Variables versus Endogenous Explanatory Variables

A comparison of the two sets of tables is quite revealing. First, many variables that were significant in the exogenous-variable-only models were also significant in the exogenous/endogenous-variable models; when this happened, reassuringly, the coefficient signs were always in the same direction and the coefficient magnitudes similar. Second, the endogenous variables both replaced and complemented the previous exogenous variables. For instance, the endogenous model of pro-high density had three new exogenous variables that the exogenous-variable-only model of pro-high density did not have (culture lover, work-driven, and number of people in home under 16 years of age), while losing four exogenous variables. One common reason for this type of result is interaction among variables; one such example suspected is the inclusion of traditional as an explanatory variable and the resulting significance of culture lover, a lifestyle variable found earlier to be strongly associated with traditional neighborhoods. Of necessity, since the previous models were the "best" exogenous-variable-only models and since all of those variables were still available for inclusion in these models, the model fits for every model improved with the addition of endogenous explanatory variables, with adjusted R-squared values ranging from 0.041 to 0.392. However, it is important to remember that the statistical results are suspect due to the violation of OLS regression assumptions.

The travel demand endogenous variables frequently appeared as significant variables in this set of models (e.g., vehicle miles was significant in four of the eight regression models other than its own). Commute distance, a proxy for job location, also turned out to be important in the regression models; it was statistically significant in three of the models. These findings support the belief that the complex interdependent relationships defining the conceptual model need to be addressed as a system (as the endogenous variables are not affected independently of each other, but in an interactive manner). The last section in this chapter presents the first effort at modeling a simple system, with the estimation of the two residential choice/preference variables (the traditional and suburban factor scores) using a seemingly unrelated regression procedure (SUR).

7.6 Seemingly Unrelated Regression Modeling (SUR)

When a researcher is faced with more than one regression equation where the errors are correlated across equations, seemingly unrelated regression (SUR) modeling may be performed. Even if the equations have no observed explanatory variables in common (hence, seemingly unrelated), there is a gain in efficiency from their joint estimation (Greene, 1990). In our context, it is logical to expect that the two measures of residential choice/preference, the traditional and suburban factor scores, would have unobserved predictors in common and hence that their error terms would be correlated. We wished to compare the two equations estimated separately with the SUR results of estimating them together (other subsets of the nine equations may also be viewed as potentially seemingly unrelated, and estimated using this technique; however, for simplicity we focused only on the two residential choice/preference models, representing the heart of the dissertation).

Hence, as a first step toward simultaneous estimation of the conceptual model equations, the two measures of residential choice/preference were estimated using SUR procedures. It is important to note that a key difference between SUR and SEM (the procedure used in Chapter 8) is that in SUR there are no endogenous variables on the right-hand-side of the equation (that is, no endogenous variables acting as predictors). Generalized least squares is often the method of estimation for SUR, but maximum likelihood estimation was used here as it is the preferred method for the later models in Chapter 8. The results of the SUR are shown in Table 52.

Inspection of Table 52 shows that simultaneous estimation yielded model fits approximately equal to the ones obtained individually (though the model fit for suburban was better with the SUR method). Interestingly, there were slight differences in the number and type of significant explanatory variables between the single and simultaneous models, reflecting that interaction is occurring among the variables in the simultaneous model. The variables that were significant in both models uniformly had higher t-statistics in the SUR model, reflecting the greater efficiency (smaller standard errors of estimation) alluded to above. The fact that interaction is present supported the decision to use a structural equation framework for the estimation of the final conceptual model.

Table 52: Seemingly-Unrelated-Regression (SUR) Model - Residential Choice

Endogenous Variable - **Traditional**	Coefficient	t-statistic (p-value)
Intercept	1.34	6.09 (0.00)
Sociodemographic		
Age	-0.015	-4.41 (0.00)
Education	0.60	2.29 (0.02)
Household Size	-0.15	-4.38 (0.00)
Number of Vehicles	-0.27	-7.10 (0.00)

Attitudinal		
Pro-Pricing	0.082	2.50 (0.01)
Lifestyle		
Culture-Lover	0.19	5.57 (0.00)
Fun-Seeker	-0.088	-2.68 (0.01)
Nest-Builder	-0.23	-7.03 (0.00)
Model Fit		
Squared Multiple Correlation (degrees of freedom)	0.390 (39)	
Endogenous Variable - **Suburban**	Coefficient	t-statistic (p-value)
Intercept	-0.62	-3.28 (0.00)
Sociodemographic		
Age	0.012	3.03 (0.00)
Professional	0.18	2.31 (0.02)
Attitudinal		
Pro-Alternatives	-0.15	-3.79 (0.00)
Pro-Drive Alone	-0.068	-1.80 (0.09)
Pro-Pricing	-0.13	-3.44 (0.00)
Lifestyle		
Athlete	0.067	1.85 (0.09)
Culture-Lover	-0.17	-4.32 (0.00)

Hobbyist	-0.11	-2.90 (0.00)
Homebody	0.18	4.48 (0.00)
Model Fit		
Squared Multiple Correlation (degrees of freedom)	0.147 (39)	

7.7 Chapter 7 Summary

Chapter 7 laid the groundwork for the estimation of the residential choice/preference conceptual model. The first section presented single-equation regression models of nine endogenous variables representing four components of the conceptual model; a first step at identifying potential explanatory variables for the full conceptual model estimation in Chapter 8. The next section contained models that also allowed endogenous variables to be used as explanatory variables; the second step in the framework development. A seemingly-unrelated regression model followed (to test a simultaneous estimation procedure), with the results supporting the need to more fully model the interactive affects among the endogenous variables. The significant variables found here were the primary variables used in specifying the full conceptual model, which is presented in the next chapter.

CHAPTER 8

OPERATIONALIZING THE RESIDENTIAL CHOICE CONCEPTUAL MODEL: A STRUCTURAL EQUATION ANALYSIS

8.1 Introduction

Chapter 7 presented models for many aspects of an individual's residential choice/preference. Specifically, each endogenous variable modeled represented a box (primary concept) in the conceptual model of residential choice/preference visually depicted by Figure 1 and described in Chapter 3. A major goal of this dissertation was to operationalize this residential choice/preference conceptual model. The single-equation model findings from Chapter 7 provided a foundation for the development of a more comprehensive model, a simultaneous system of equations that represents an individual's choice/preference of residential location. Specifically, the statistically significant exogenous variables found in the single-equation models were used as a base for the development of the system of equations modeled later in this chapter. Variables not found to be significant in the single-equation models, but felt to be conceptually sound, were also tried in the models of Chapter 8.

It is important to note that not every relationship defined in the conceptual model was modeled, neither in Chapter 7, nor in Chapter 8. There are many reasons for this. First, mathematically estimating a simultaneous model with so many interdependent relationships requires a substantial amount of data, the right data (i.e., careful measurement of each variable in the model), and a very complex model specification structure. Though the data set for this dissertation is quite rich and robust (see Chapter 4), the large number of nonrecursive relationships (variables that simultaneously influence each other) prevented the estimation of the full conceptual model due to underidentification (see Section 8.3.2.3).

The following section describes why a simultaneous-equation model structure was chosen for the conceptual model, and includes a brief investigation into the idea of causality. Next, theory for structural-equation modeling (SEM) is presented, followed by the specification of the conceptual model. A description of model results concludes the chapter.

8.2 The Advantages of Simultaneous-Equation Models

The conceptual residential choice/preference model developed in this dissertation is built on sets of relationships. For example, residential location is a function of variables such as household characteristics and job location, making it an endogenous variable in this study. Similarly, job location, though hypothesized to influence residential choice (i.e., an explanatory "right-hand-side" variable), is also an endogenous variable in this study since it is also modeled as a function of household characteristics. These complex interrelationships suggest the use of simultaneous equation systems that allow a researcher to study the behavioral system as a whole. Further, even if the only interest were in a small part of the system such as travel demand, the exclusion of the interaction with sub-elements of the whole system would have serious implications for both the estimation and interpretation of any models from the system (Greene, 1997). In fact, one major advantage of using SEM procedures is that in addition to obtaining coefficients for the direct relationships of a model system (which is all OLS regression can do, and badly, given the violation of assumptions in our model), the direction and extent of the combined impacts of interaction relationships can also be calculated. Formally, these results are called direct effects and total effects, which are the sum of direct and indirect effects (see Section 8.4.1.1). Total effects can differ substantially from direct effects, so it is important to have the complete picture that can be obtained only from SEM.

Another important reason for choosing a simultaneous model structure was its positive statistical properties. The residential choice conceptual model in this dissertation (see Figure 1) reflects many interrelationships among variables. It was shown in Chapter 7 that many of the variables selected to model these relationships were not independent of each other (as seen by the significant t-statistics for several endogenous variables that were used as explanatory variables in Section 7.4), and further, that error terms were not uncorrelated across equations. Thus, modeling the endogenous variables one equation at a time violates the assumptions required for OLS to be valid. Hence, it is imperative to estimate the system of equations comprising the conceptual model together. The best way to do that is using full-information maximum likelihood estimation, as we have done here. The resulting coefficients possess all of the desirable asymptotic properties of maximum likelihood estimators (e.g., consistency and efficiency).

In addition to the above-mentioned positive aspects of structural equation modeling, the approach has been considered a tool to investigate causality. Mulaik (1987) points out that many researchers have a limited understanding of the concept of causality. Indeed, since understanding the causality underlying the relationships among people's choice of residence, job location, and travel is a major goal of this dissertation, a brief look into the definition of causality and how it may be interpreted in terms of modeling is given next.

A common theory of residential choice is that individuals choose a job first, and then, based on the job location, choose a home (see, e.g., Verster, 1985). In other words, this theory assumes that job location, at least partially, *causes* residential choice. For this case and others, it is important to be aware of what an empirical model can and cannot say about causality. But what is causality? Webster defines causality as "the relation between a cause and its effect or between regularly correlated events or phenomena." We refer to causal modeling as a set of techniques to investigate causes, providing explanations of effects as the result of previous causes. For example, structural equation modeling has been known to many investigators as "causal modeling". But in reality, statistical methods based on correlational data cannot "establish or prove causal relationships between variables." (Mueller, 1997, pg. 355). As Mulaik (1987) puts it:

> "...researchers take correlational data obtained for a number of variables, specify a general model of proposed causal relations, and then fix and free parameters in trial-and-error fashion until they find a specific model that fits the data. The resulting model may or may not show much similarity to their original model, but they will believe that in their finally obtained model they have discovered real causal connections." (pg. 19)

He warns that this method has some flaws and that a researcher should be aware of the potential shortcomings. Most importantly, he states that the validity of a model that has few fixed parameter values left from the original design will be suspect, despite what the goodness-of-fit test may indicate. Mulaik suggests that a researcher must use intuition (conceptual reasoning) to select a causal direction hypothesis to implement and test for fit. He further notes that "time ordering" is likely to make many of the specifications conceptually unsound. Mulaik states that causal model coefficients are uniquely defined when "the effects of changes in exogenous variables [have had time] to work themselves through to stable values in the endogenous variables." This study complied with the former recommendation, in that every relationship defined was based on conceptual reasoning, but it was unable to follow the latter suggestion since only cross-sectional data were available.

In conclusion, it is believed that the SEM framework can help a researcher to gain an understanding of potential causal mechanisms that lie behind the data. However, it is recommended that terms such as cause and effect be replaced with words like predictor and outcome when forming conclusions from model estimations based on correlational data, especially cross-sectional data.

8.3 Structural-Equation Modeling (SEM) Theory

Researchers have found that structural (or *simultaneous*) equation systems can be effective in modeling behavioral phenomena involving relationships among multiple endogenous variables (Greene, 1990). Structural equation models can be used to develop and test conceptual behavioral theories (Mulaik, 1987). Thus, generally, each equation in a structural equation model "represents hypothesized causal links and not just empirical associations" (Anderson, 1987, pg. 49). Further, by capturing more of the context in which a single relationship occurs, the empirical results are more likely to represent true causality than undirected associations, compared to a single equation regression. An overview of the specification and identification of a general structural equation model is provided next.

8.3.1 The SEM Framework

Using the notation of Mueller (1996), a structural model can be described by the following matrix representation:

$$\mathbf{Y = ?Y + GX + ?}\,,$$

where:

Y = (NY x 1) column vector of endogenous variables (NY = # of endogenous variables),

X = (NX x 1) column vector of exogenous variables (NX = # of exogenous variables), **B** = (NY x NY) matrix of coefficients representing the direct effects of endogenous variables on other endogenous variables,

G = (NY x NX) matrix of coefficients representing the direct effects of exogenous variables on endogenous variables, and lastly,

? = (NY x 1) column vector of errors.

This above system of equations is defined as a set of **structural** equations in that it is based on a conceptual structure (Cooley and LeRoy, 1985). In terms of this study, each equation represents a relationship relevant to the concept of residential choice/preference. Section 8.4 illustrates this point.

In addition, a (NX x NX) variance/covariance matrix **F** for the exogenous variables **X** and a (NY x NY) variance/covariance matrix ? for the error terms **?** must be known for a complete model. This is part of the reason structural equation modeling is also commonly known as covariance structure analysis. To estimate the structural equation model defined above, S, the population covariance matrix of observed variables X and Y will be described in terms of the unknown parameters ? that comprise the **B**, **G, F ,** and ? matrices. In other words, each value within the population covariance matrix will be defined as a function of one or more model parameters. The unknown parameters will then be estimated mathematically by algorithms that attempt to minimize the difference between the sample covariance matrix (calculated from the observed data) and the population covariance matrix (Joreskog, 1970). Indeed, a structural equation model's goodness of fit is based on how well its model-implied variances and covariances of variables (population statistics) compare to the actual variances and covariances of variables calculated from the data used to estimate the model (Hayduk, 1987).

8.3.2 SEM Estimation Issues

In our context, five issues associated with structural equation modeling are important: 1) endogenous variable type, 2) endogenous variable measurement, 3) achieving identification, 4) multivariate normality, and 5) model fit. The last four are always important, while the first is an issue in some contexts including ours. Each of these areas is related to the others, and they are key to successfully estimating a structural equation model.

8.3.2.1 Endogenous Variable Type

A variable that is "directly caused or influenced" by another variable is classified as endogenous (Hayduk, 1987, pg. 89). An important issue in SEM is that the endogenous variables may not be continuous. Indeed, many behavioral phenomena are measured in discrete terms; choice of neighborhood type is one example. A considerable amount of research involving the use of qualitative dependent variables in a structural equation system has been documented by Maddala (1983). Modelers must make modifications to the general structural equation framework if they want to explain discrete phenomena. One approach that has been taken in the case of binary or ordinal discrete variables is to view the observed choice as the outcome of an unobserved ("latent") continuous variable crossing successive thresholds, and modeling the continuous latent variable (see, e.g., Heckman, 1978).

In this study, using the factor analysis procedure described in Chapter 6 to define the variables representing neighborhood type (choice) resulted in continuous measures (the factor scores) for these important endogenous variables. Consequently, the need to model discrete choice variables was avoided, a key advantage of the factor analysis measurement approach over a more typical categorical formulation of the neighborhood choice variable (e.g., traditional versus suburban).

8.3.2.2 Endogenous Variable Measurement

The residential choice/preference conceptual model (see Figure 1) is a visual illustration of many hypothesized relationships among variables. Theoretically, one could take each relationship depicted on the figure and embed it in a structural equation. Unfortunately, each element in the conceptual diagram can be represented by many different variables (see Tables 1 through 6), making it much more complex to specify an empirical model. For example, travel may be measured in several different ways (e.g., trip frequency and percent of trips driving alone), each of which may have some independent explanatory power. This indicates that many travel behavior variables could be placed in the model structure, and determining how each of those possible variables affects, and is affected by, other variables, is non-trivial.

The variables chosen to represent each of the endogenous variables in the conceptual model were: traditional and suburban factor scores (Residential Choice/Preference), daily vehicle-mile rate, daily transit-mile rate, and daily walk/bike- mile rate (Travel Demand), pro-driving, pro-high density, and pro-transit factor scores (Attitudes and Lifestyle), and commute distance (Job Location). Other variables could have reasonably been chosen to represent conceptual model elements (such as daily vehicle-trips or the pro-low density attitude factor). However, based on data availability, preliminary exploration, and conceptual reasoning, the above selected variables were believed to best represent the conceptual model elements. For example, though good data were available for both daily vehicle-mile rate and daily vehicle-trip rate for individuals, and both variables represent one aspect of travel demand, models with daily vehicle-mile rate consistently produced better model fits than daily vehicle-trip rate models.

A few other important decisions need to be discussed. First, since much of the data representing Neighborhood Characteristics and Dwelling-Unit Characteristics was used in establishing measures for Residential Choice/Preference, it was decided that no endogenous (nor exogenous) variables would be used to represent these two conceptual model elements. It was believed that any variable measures (especially based on this study's data set) for these two elements would be too highly correlated with the main conceptual model element, Residential Choice/Preference, potentially leading to multi-collinearity problems. Second, since variables representing Sociodemographic and Life-Cycle variables are more commonly used as exogenous variables in behavioral models, it was decided that no endogenous variables would be created for this conceptual model element. The choice not to model these three elements as endogenous variables reduces the complexity of the model structure by removing multiple nonrecursive relationships (i.e., two-way relationships between variables). This decision was very important to achieving model identification, the topic that follows. Section 8.4 contains a description of the modeling results based on the above endogenous variable decisions.

8.3.2.3 Identification

Identification, in essence, means that all of the coefficients in the structural equations can be determined. The task here is to determine whether or not there is adequate variance and covariance information from the observed variables to estimate a model's unknown coefficients. A simple analogy is the case where two unknowns in an equation are to be found. If there is only one equation and two unknowns, an infinite number of solutions are available, and hence, it is unidentifiable. However, if two independent equations are given a unique answer to the two unknowns can be found (i.e., they can now be identified).

In structural equation modeling, all of the equations have to be identifiable for the empirical estimates to be valid (Berry, 1984). In other words, an identifiable system of equations is one in which each parameter can be written as a function of the variances and covariances of the observed variables (Mueller, 1996). Sample size is not an issue for identification, but what is important is the “ratio of the number of variables in the model to the number of unknown parameters” (Mueller, 1997, pg. 357).

Two methods are commonly used to determine if a system of equations is identifiable: the order and rank conditions. The order condition applies to one equation at a time, and can be described as follows (Griffiths *et al.*, 1992, pg. 606):

1. “When the number of endogenous and exogenous variables [on the right hand side] of the equation minus 1 is less than the total number of exogenous variables in the system, then the equation is over identified.”
2. “When the number of endogenous and exogenous variables [on the right-hand-side] of the equation minus 1 is equal to the total number of exogenous variables in the system, then it is identified.”
3. “ When the number of endogenous and exogenous variables [on the right-hand-side] of the equation minus 1 is greater than the total number of exogenous variables in the system, then it is underidentified.”

The order condition is a necessary, but not sufficient, condition for identification (Griffiths *et al.*, 1992). Due to its simplicity though, it is typically used first to screen for identification before going to the more arduous tests like the rank condition.

The rank condition is a necessary and sufficient condition for identification, that is adapted from a procedure of testing a system of equations one equation at a time for identification. Using the notation shown above, **Y = ?Y + GX + ?** , the jth equation of the system may be represented as follows: $\mathbf{y}^t\mathbf{G_j} + \mathbf{x}^t\mathbf{?}_j = \mathbf{e_j}$, where t is the transpose operator. For each such equation, the reduced-form coefficient matrix (**?** $_j$) is found by solving for **y** in terms of only $\mathbf{x_s}$ and **e** (i.e., reduced form equations are equations re-written with the endogenous variables on the right side of the equation expressed as linear combinations of the exogenous variables **x**). Greene (1990, pg. 592) defines the rank condition for identification as: rank[**?** $_j$] = M_j, where M_j is the number of included endogenous variables in equation j. Specifically, if the rank of the reduced-form matrix for equation j is equal to the number of included endogenous variables for equation j, then equation j is identified. In other words, the rank condition is a test to see whether there is one solution for each of the structural parameters given the reduced-form parameters. Each equation is analyzed in a step-wise procedure to see if the rows and columns of the matrix can be reduced to a form that represents identification (see Berry, 1984).

Identification is typically not easily achieved. Researchers often have to make restrictions on their equations to meet identification requirements. In simultaneous equation models, identification is most often achieved by restricting selected elements of ? and G to 0 (Long, 1983). If an element of ? is zero, say $ß_{ij}$, it means that the endogenous variable Y_j does not directly affect the endogenous variable Y_i. A researcher should be very careful about the types of restrictions placed on an equation, as estimation biases result from missing or improper variable specification. Clif (1983) warns researchers to avoid this type of problem altogether, by establishing theories prior to survey design, data collection, and modeling efforts (which ideally would facilitate exact, identifiable equations). In reality however, when theory itself suggests a system of equations which is not identifiable (as it often seems to do), the researcher has the unattractive task of trying to decide which restrictions impose the least damage on the theoretical model. Further, exploratory studies such as the one providing the data for this dissertation, while ideally grounded in theory and solid survey design and data collection, are unlikely to perfectly develop all of the right variables and the right way to measure them.

This dissertation has succeeded in specifying a conceptual model that is both acceptably realistic and statistically identifiable. The nine-equation simultaneous system of equations presented in Section 8.4 is the result of conceptual reasoning and statistical modification. The number and type of variables used to represent conceptual model elements were determined (see Section 8.3.2.2) in such a manner that every free parameter could be identified.

Modifying models based on hand calculations of the rank and order conditions would have been very tedious. Fortunately, AMOS 3.6 (the software used to estimate the structural equation models for this study) automatically determines whether or not a specified model is identifiable, providing output that suggests which parameters are likely not identified. Unfortunately, AMOS does not report whether a model is overidentified or just identified (both conditions are considered "identifiable"). A manual check of the order conditions of this dissertation's final model (Section 8.4) indicated that the system of equations was overidentified (i.e., the number of endogenous and exogenous variables on the right-hand-side of each equation [subtract one] was less than the total number of exogenous variables in the system, 38). This is actually a desirable outcome. MacCallum (1995) states that overidentified models are better than just-identified models because overidentified models will allow there to be a meaningful correspondence between the data and the model, while just-identified models always fit exactly, making inferences about model fit meaningless.

When an equation is overidentified it is suggested that its likelihood function be maximized with respect to all of its unknown parameters (Greene, 1990); the full-information maximum likelihood procedure used in AMOS satisfied this suggestion. Further, parameter estimates are still fully efficient if the error terms are normally distributed (which was found to be the case for every endogenous variable except daily walk/bike mile rate).

8.3.2.4 Multivariate Normality

A key assumption in the estimation of structural equation models is that the observed variables follow a multivariate normal distribution. When this assumption is true, the variance of the estimated parameters is consistently estimated by sample variances, but when it is false, the standard errors of parameter estimates can be significantly underestimated, leading to false model conclusions (West *et al.*, 1995). A review of the literature reveals that meeting this condition is a problem in many studies. Bentler and Dudgeon (1996, pg. 566) stated that "in practice [for structural equation models], the normality assumption will often be incorrect." Micceri (1989) reviewed data that was used in numerous journal articles and found that a majority of the conclusions were based on data that were nonnormally distributed. Other researchers (e.g., Breckler, 1990; Gierl and Mulvenon, 1995) have noted that it is very common for practitioners to ignore the assumption of normality and to make conclusions as if the assumption were met.

8.3.2.5 Model Fit

As mentioned earlier, a structural equation model's goodness-of-fit is based on how well its model-implied variances and covariances of variables (population statistics) compare to the actual variances and covariances of variables calculated from the data (Hayduk, 1987). The ?^{2}-test used to be the primary test of fit in covariance structure analysis, where a large ?^{2} indicated a large discrepancy between the sample and fitted covariance matrices (Hu and Bentler, 1995). Many other tests of fit were developed later as problems associated with goodness-of-fit ?^{2} tests were found. In particular, two issues were especially noteworthy: 1) the assumption of a chi-square distribution is less accurate as the sample size increases (Hu and Bentler, 1995), and 2) the ?^{2} statistic can be made smaller simply by reducing a study's sample size (Bentler and Bonett, 1980). As sample size is important to structural equation modeling (Bentler, 1993, recommends a ratio of sample size to parameters estimated [such as variable coefficients, but not including covariance estimates] to be around 10; our study's smallest sample (N_4 = 515) amply meets this suggested requirement, since a total of 47 parameters are estimated), removal of cases to improve the ?^{2} statistic was not considered a credible solution to these issues.

8.4 Conceptual Model Specification

Based on conceptual reasoning and statistical considerations, a system of nine equations was developed to empirically represent the residential choice/preference conceptual model of Chapter 3. The single-equation full (i.e., both endogenous and exogenous variables on the right-hand side of the regression equation) models presented in Chapter 7 were the basis of the structural equations modeled here (though changes in variable significance due to interaction in the simultaneous estimation has made each equation different than in Chapter 7). The final (untransformed) structural equations model specification is given by the following equations:

Traditional = f(constant, age, household size, no. of vehicles, nest-builder, pro-pricing, pro-HD, error1)

Suburban = f(constant, age, culture-lover, hobbyist, homebody, pro-alternatives, pro-pricing, commute distance, error2)

Pro-HD = f(constant, no. of people under age 16, culture-lover, nest-builder, outdoor enthusiast, work-driven, pro-driving, error3)

Pro-driving = f(constant, female, years lived in Bay Area, time-satisfied, vehicle miles, error4)

Pro-transit = f(constant, pro-drive alone, pro-environment, error5)

Vehicle miles = f(constant, female, no. of vehicles, adventurer, walk/bike miles, error6)

Transit miles = f(constant, no. of vehicles, pro-growth, pro-transit, commute distance, error7)

Walk/bike miles = f(constant, pro-alternatives, pro-driving, error8)

Commute distance = f(constant, suburban, vehicle miles, transit miles, error9)

In particular, the first two equations represent the Residential Choice/Preference element in the conceptual diagram (see Figure 1), the next three equations represent the Attitudes and Lifestyle element, followed by three equations that represent the Travel Demand element, and the last equation is used to represent the Job Location element. The above equations re-written in terms of structural equation notation are (suppressing the individual specific subscripts on the left hand side variables, Xs, and ?s for simplicity):

$$R_t = ?_{11} + ?_{12}X_2 + ?_{14}X_4 + ?_{15}X_5 + ?_{1,12}X_{12} + ?_{1,18}X_{18} + ß_{13}A_h + ?_1$$

$$R_s = ?_{21} + ?_{22}X_2 + ?_{29}X_9 + ?_{2,10}X_{10} + ?_{2,11}X_{11} + ?_{2,14}X_{14} + ?_{2,18}X_{18} + ß_{29}J + ?_2$$

$$A_h = \ ?_{31} + ?_{36}X_6 + ?_{39}X_9 + ?_{3,12}X_{12} + ?_{3,13}X_{13} + ?_{3,20}X_{20} + ß_{34}A_d + ?_3$$

$$A_d = \ ?_{41} + ?_{43}X_3 + ?_{47}X_7 + ?_{4,19}X_{19} + ß_{46}T_v + ?_4$$

$$A_{pt} = \ ?_{51} + ?_{5,15}X_{15} + ?_{5,16}X_{16} + ?_5$$

$$T_v = \ ?_{61} + ?_{63}X_3 + ?_{65}X_5 + ?_{68}X_8 + ß_{68}T_w + ?_6$$

$$T_{tr} = \ ?_{71} + ?_{75}X_5 + ?_{7,17}X_{17} + ß_{75}A_{pt} + ß_{79}J + ?_7$$

$$T_w = \ ?_{81} + ?_{8,14}X_{14} + ß_{84}A_d + ?_8$$

$$J = \ ?_{91} + ß_{92}R_s + ß_{96}T_v + ß_{97}T_{tr} + ?_9$$

where:

R = residential choice/preference - t is traditional, s is suburban;
A = attitudes - h is pro-high density, d is pro-driving, pt is pro-transit;
T = travel demand - v is daily vehicle miles, tr is transit miles, and w is for walk/bike miles;
J = job location;
? = coefficients expressing exogenous variable direct effects;
ß = coefficients expressing endogenous variable direct effects;
X = exogenous variable values, with X_1 always equal to one; and
$?_i$ = the error term for equation i, i = 1, ..., 9.

Note that this set of structural equations is essentially a simplified version of the original conceptual model of Figure 1. The main difference is the number of endogenous concepts to be modeled. Figure 1 contains 7 endogenous concepts, whereas the empirical structural equation system models 4 endogenous concepts: residential choice/preference, travel demand, attitudes and lifestyle, and job location. The other three types of variables (socio-demographics/life cycle, neighborhood and dwelling unit) are assumed to be exogenous, that is, predetermined and not influenced by the other four endogenous variables. This assumption of exogeneity is reasonable in the present context, where we are estimating a static model (at one point in time). When the process is looked at in a dynamic sense, however, factors like sociodemographics and life cycle may change over time due to factors such as residential choice, and hence, should be modeled as endogenous to the system.

The empirical estimation of structural equations requires methods more sophisticated than ordinary least squares. Using AMOS 3.6, the matrices of coefficients B and G were estimated via the full-information maximum likelihood approach. These results are analyzed next.

8.4.1 Conceptual Model (Untransformed) Results

After numerous model modifications and estimations, an intuitively sound conceptual model was finally obtained that was identifiable. A look at the direct and total effects coefficients of the structural equation model (see Tables 53 and 54) shows that many of the expected relationships within the conceptual choice/preference model were confirmed. In other words, the coefficient values imply a strength and direction of the (direct and total, respectively) relationship between two variables. Chapter 5 contains detailed descriptions of all of the variables discussed below.

8.4.1.1 Direct Effects

First, it should be noted that direct effects of the attitudes and lifestyle variables constituted many of the significant effects of the structural model, indicating the explanatory power that attitudinal and lifestyle variables have in this residential choice/preference model. Pro-pricing, an attitudinal variable representing a respondent's willingness to pay for traffic control measures, had a positive effect on traditional (and a negative effect on suburban), indicating that people who score highly on this factor are significantly more likely to choose/prefer a traditional neighborhood. Another significant attitudinal variable, pro-alternatives, had an intuitive negative relationship with suburban and a positive relationship with walk/bike miles. All of these findings supported the initial expectations that people who are environmentally conscious and supportive of policies and actions aimed at reducing negative impacts on the environment would be more likely to choose to live in a traditional neighborhood.

Individual lifestyle variables also played a strong role in the model. Two examples are adventurer and nest-builder. People scoring highly on adventurer, a variable representing a person's participation in activities such as driving off-road vehicles, traveled longer distances by vehicle, on average. Nest-builders, people who participated often in home-based activities such as gardening, were less likely to choose traditional neighborhoods and less likely to be pro-high density. This finding was particularly interesting in that it supported the idea of lifestyle activities influencing a

Table 53: Direct Effects (t-stat.) of Nontransformed Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location

Explanatory Variable ®	Trad-itional	Sub-urban	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Constant	1.512 (8.06)	-0.733 (-3.12)	0.145 (3.03)	-1.041 (-5.47)	-0.010 (-0.26)	28.668 (6.96)	5.241 (2.22)	0.608 (5.60)	1.798 (1.02)
Sociodemographic									
Age	-0.015 (-4.63)	0.008 (2.44)							
Female				0.550 (5.64)		-9.322 (-3.91)			
Household Size	-0.128 (-3.63)								
Number of People Under Age 16			-0.225 (-4.69)						
Number of Vehicles	-0.247 (-6.47)					5.121 (4.92)	-2.19 (-4.95)		
Years Lived in Bay Area				0.007 (3.36)					
Lifestyle									
Adventurer						4.16 (4.30)			
Culture-Lover		-0.158 (-4.41)	0.251 (6.83)						
Hobbyist		-0.100 (-2.96)							
Homebody		0.175 (4.81)							
Nest-Builder	-0.141 (-3.79)		-0.182 (-4.90)						
Outdoor Enthusiast			0.091 (2.57)						

Table 53: Direct Effects (t-stat.) of Nontransformed Res. Choice SEM - continued

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urban	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Attitudes									
Pro-Alternatives		-0.105 (-2.73)						0.275 (2.59)	
Pro-Drive Alone					0.161 (4.16)				
Pro-Environment					0.167 (4.25)				
Pro-Growth							-1.082 (-2.86)		
Pro-Pricing	0.109 (3.42)	-0.123 (-3.58)							
Time-Satisfied				-0.119 (-3.28)					
Work-Driven			0.119 (3.63)						
Pro-High[1] Density	0.498 (5.21)								
Pro-Driving[1]			-0.643 (-4.40)					-0.566 (-1.60)	
Pro-Transit[1]							3.192 (2.25)		
Residential Choice/Preference									
Suburban[1]									2.692 (2.89)
Travel Demand									
Vehicle Miles[1]				0.020 (4.06)					0.223 (5.33)
Transit Miles[1]									0.712 (4.89)
Walk/Bike[1] Miles						-7.441 (-2.02)			
Job Location									

Commute[1] Distance		0.030 (1.97)					0.341 (1.94)		

[1] Endogenous variable.

Table 54: Total Effects of Nontransformed Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad- itional	Sub- urban	Pro- HD	Pro- Driving	Pro- Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Constant	0.070	-0.007	0.062	0.042	0.001	30.134	5.055	0.561	12.084
Sociodemographic									
Age	-0.015	0.009					0.011		0.033
Female	-0.128	-0.076	-0.257	0.400		-7.637	-0.857	-0.226	-2.514
Household Size	-0.128								
Number of People Under Age 16	-0.112		-0.225						
Number of Vehicles	-0.282	-0.014	-0.070	0.110		5.583	-2.348	-0.062	-0.468
Years Lived in Bay Area	-0.003		-0.005	0.008		0.034	0.004	-0.005	0.011
Lifestyle									
Adventurer	-0.028	0.045	-0.057	0.089		4.536	0.509	-0.050	1.493
Culture-Lover	0.125	-0.177	0.251				-0.215		-0.630
Hobbyist		-0.112					-0.136		-0.399
Homebody		0.196					0.238		0.698

Nest-Builder	-0.232		-0.182						
Outdoor Enthusiast	0.045		0.091						

Table 54: Total Effects of Nontransformed Residential Choice SEM - continued

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urban	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Attitudes									
Pro-Alternatives	0.014	-0.140	0.028	-0.044		-2.229	-0.393	0.300	-1.151
Pro-Drive Alone		0.016			0.161		0.698		0.541
Pro-Environment		0.017			0.167		0.724		0.561
Pro-Growth		-0.034					-1.471		-1.140
Pro-Pricing	0.109	-0.138					-0.167		-0.490
Time-Satisfied	0.042	-0.005	0.084	-0.130		-0.547	-0.061	0.073	-0.180
Work-Driven	0.059		0.119						
Pro-High[1] Density	0.498								
Pro-Driving[1]	-0.349	0.046	-0.701	0.090		4.589	0.515	-0.617	1.511
Pro-Transit[1]		0.101					4.339		3.362
Residential Choice/Preference									

Suburban[1]		0.120					1.358		3.982
Travel Demand									
Vehicle Miles[1]	-0.007	0.011	-0.014	0.021		0.090	0.122	-0.012	0.359
Transit Miles[1]		0.032					0.359		1.053
Walk/Bike[1] Miles	0.051	-0.080	0.102	-0.159		-8.111	-0.911	0.090	-2.670
Job Location									
Commute[1] Distance		0.045					0.505		0.479

[1] Endogenous variable.

person's attitudes and choice of residence. However, as noted in Section 8.2, causality inferences even for a structural equation model must be made with caution. In particular, causal relationships in the opposite direction were not tested in order to maintain the identifiability of the model by focusing on the direction of causality expected to be the strongest.

Sociodemographic variables, the variables usually found in studies of residential choice, were also valuable. Number of vehicles was especially important, being significant in three of the nine equations. As expected, number of vehicles was positively associated with vehicle-miles traveled, and negatively associated with the amount of travel by transit and the choice/preference of living in a traditional neighborhood. Variables that were expected to be significant but were not included household income, education, and occupation. These variables were significant in the single-equation models of Chapter 7, indicating that simultaneous-equation estimation captures their effects through other variables within the system of equations.

It is also important to analyze the impact that endogenous variables have on other endogenous variables (i.e., the coefficients in the ? matrix). Vehicle miles, commute distance, and pro-driving are the three endogenous variables that were most often significant in explaining residential choice/preference model components. Vehicle miles was positively associated with each of the other two, supporting the belief that the more an individual travels by car the more likely the person will have a positive view of driving and a longer commute. Commute distance and suburban were each significant in the equation for the other, illustrating the symbiotic association of suburban locations with longer commutes (compared to those for traditional neighborhood dwellers). Further, commute distance was positively related with transit miles, supporting the idea that people would be more likely to commute via transit if the distance between home and work was long enough to be worth the effort (probably also reflecting the availability of high-quality transit service -- express bus and rail -- for longer-distance commutes). The pro-driving attitude showed a negative relationship with the pro-high density attitude and with walk/bike miles, both logical results.

Before analyzing the total effects and error correlations from the structural equation results, it is valuable to focus on the heart of the model, the two residential choice equations. The traditional and suburban equations had significant explanatory variables from the sociodemographic, attitudes, lifestyle, and job location categories. As expected, household size and number of vehicles were negatively associated with the choice of a traditional neighborhood, while age was positively associated with the choice of a suburban neighborhood. Pro-pricing was positively associated with traditional and negatively associated with suburban, reflecting that respondents who support monetary policies that encourage alternatives to the automobile are more likely, on average, to live in traditional neighborhoods. Other attitudinal variables such as pro-alternatives and pro-high density also impacted residential choice, but residential choice does not directly affect attitudes (at least the three we tested as endogenous), which is consistent with our expectation of the primary direction of impact, and increases our confidence in the realism of this operationalization of the model.

The lifestyle variable, culture-lover, was positively associated with the choice of living in a traditional neighborhood, confirming earlier model findings (see Chapter 5 and 7) that respondents who participate in cultural activities were more likely on average to live in traditional neighborhoods.

As mentioned earlier, the job location variable, commute distance, impacted residential choice in the expected direction, indicating that people with longer commutes were more likely to live in suburban neighborhoods than in traditional neighborhoods. Looking at the relationship in the other direction, the biggest effect of residential choice was on commute distance. Though the strength of the relationship appears to be much stronger for this latter direction (as seen by the magnitudes of the direct effects), this does not necessary imply that residential choice preceding and influencing job location (please refer to Sections 3.18 and 3.19) is the dominant effect. Rather, it is quite likely a consequence of the specific way the job location and residential choice variables are measured in this study. However, it may in fact reflect the reality that in a large metropolitan area and with two-career households, often residential choice does precede the current job held, and hence, determines commute distance.

It is particularly noteworthy that no travel variables are directly significant to residential choice, nor are any residential choice variables directly significant to travel. By contrast, five attitude and lifestyle variables significantly impact travel demand. However, the relative power of the direct effects may be very misleading due to interaction effects between related variable relationships (i.e., indirect effects), and consequently, conclusions should be based on both direct and indirect effects. The following section addresses this important topic.

8.4.1.2 Total Effects

One of the main advantages of structural equation modeling is that the impact of variable interactions can be found in terms of total effects. The total effect of a variable on another variable is the sum of the direct effect and all of the indirect effects between the two variables. For example, suppose exogenous variable X_1 affects endogenous variable Y_3 directly, and also indirectly by affecting Y_1 and Y_2, which both affect Y_3 but not each other. Mueller (1996, pg.144) describes the following process of finding the total effect of X_1 on Y_3: 1) first, find all the variables that directly affect the endogenous variable Y_3; 2) second, record the structural coefficients that link (either directly or indirectly) the variables identified in the first step to Y_3; and 3) third, multiply each structural coefficient identified in the second step by the covariance of X_1. An example of what the equation would look like given three variables affecting Y_3 (i.e., endogenous Y_1, Y_2 and exogenous X_1) is:

Total effects of X_1 on Y_3 = $s_{y3,x1}$ = $ß_{31}s_{y1,x1} + ß_{32}s_{y2,x1} + ?_{31}s^{2}_{x1}$,

where s_{ij} = Cov(i,j).

When variables are standardized, total effects can be simply found by multiplying direct effect coefficients, and the above equation changes to:

Total effects of X_1 on Y_3 = $?_{y3,x1} = ß_{31}\,?_{11} + ß_{32}?_{21} + ?_{31}$,

where $?_{i1}$ is the structural coefficient for X_1 in the equation for Y_i (i.e., the direct effect of X_1 on Y_i, for i = 1, 2, 3).

As noted earlier, the direct effect values imply a strength and direction of a direct relationship between two variables. On the other hand (when the variables are standardized), the indirect effect of one variable on another (through one or more intervening variables) is the product of the corresponding coefficients (i.e., direct effect coefficients) that link the variables together.

For example, consider the effect of the variable female on pro-driving for the standardized model. First (see Table 58 in the Appendix), female has a positive direct effect on pro-driving (structural coefficient of 0.275) and a negative direct effect on vehicle miles (structural coefficient of -0.165). Second, vehicle miles has a positive direct effect on pro-driving (structural coefficient of 0.554). Vehicle miles, then, is an intervening variable between female and pro-driving, creating a negative indirect effect of female on pro-driving through vehicle miles (indirect effect = -0.165 * 0.554 = -0.091). Thus, the total effect of female on pro-driving (0.200), shown in Table 59 (see Appendix), is less than the direct effect of female on pro-driving (0.275). In other words, all else equal, women have higher pro-driving attitude scores than men. But because those who drive less tend to have lower pro-driving attitude scores, and because women tend to drive less than men, the direct effect is somewhat attenuated when that additional relationship is taken into account. The difference between the total effect of 0.200 and the sum of the direct (0.275) and indirect (-0.091) effects mentioned here, is the sum of all the remaining indirect effects of female on pro-driving (accounting for, e.g., the effect of female through the indirect effects of vehicle-miles and pro-driving on themselves, as shown in Table 59). It can be seen that if only direct effects are taken into consideration, researchers are likely to make inaccurate conclusions about the impacts of explanatory variables on endogenous variables (Lu and Pas, 1999).

Comparison of Tables 53 and 54 shows that indirect effects play an important role in explaining model relationships. Note that many more boxes are filled (demonstrating an effect) in the total effects table than in the direct effects table. For example, the variables pro-alternatives and number of vehicles each had direct effects on only two endogenous variables, while they had total effects on every endogenous variable but one. This finding reflects the importance that multi-variable interactions (i.e., indirect effects) have in representing a model's structure. It was not surprising to see the many indirect effects, as they simply confirm the prior belief that an individual's residential choice/preference is based on many complex interrelationships.

The total effects of the travel demand variables on residential choice were largely as expected. For example, walk/bike miles was positively associated with the traditional neighborhood variable and negatively associated with the suburban neighborhood variable, reflecting the typical results reported in the literature, that alternative modes of travel are more often used by residents in traditional neighborhoods. Vehicle miles was positively associated with the choice of a suburban neighborhood and negatively associated with the choice of a traditional neighborhood, supporting findings from the literature that indicate lower vehicle-miles are associated with traditional neighborhoods and higher vehicle-miles are associated with suburban neighborhoods.

These results suggest that travel demand has an influence on residential choice (assuming that the SEM structure does provide insight into causal structures). But in what sense does (current) vehicle miles or walk/bike miles cause an individual or household to have chosen a certain type of residential neighborhood? It is possible that travel demand variables are a proxy for travel predisposition (i.e., how much a person naturally desires to travel by various modes), and that an individual's choice of residence will be based on its ability to meet this predisposition.

It is important to note that attitudinal and lifestyle variables had the greatest total effect on travel demand (both in terms of direct and total effects). By contrast, residential choice had little total effect on travel demand.

The conspicuous lack of effect of neighborhood type on travel demand constitutes perhaps the strongest evidence to date supporting the speculation that the association commonly observed between land use configuration and travel patterns is not one of direct causality, but due to correlations of each of those variables with others. In particular, these results suggest that when attitudinal, lifestyle, and sociodemographic variables are accounted for, neighborhood type has very little influence on travel behavior.

All of the conclusions presented are based on the results presented in the direct and total effects tables given in this chapter, as well as a set of standardized direct and total effects tables given in the appendix (see Tables 58 and 59). Coefficients for standardized explanatory variables were obtained to confirm the strength of the relationships implied by the magnitudes of the variable coefficients. It was important to do this to ascertain the relative influence that variables had on one another, correcting for differences in scale across variables.

8.4.1.3 Cross-Equation Error Correlations

The correlation structure for the error terms in a structural equation model is specified by the researcher. Just as the observed variables comprising the system of equations were related, it was expected that the error terms (unobserved variables) among equations would be correlated, and thus, no a priori restrictions were placed on the error-correlation structure (i.e., all error terms were allowed to correlate with all others). Table 55 presents the error-correlation results for the untransformed model.

The largest-magnitude correlation (-0.618) was between the equations for transit miles and commute distance, reflecting the conceptual overlap between those two variables. Since transit miles and commute distance are positively related in our model, this finding suggests that the observed variables captured the positive relationship between the two variables, whereas the unobserved variables affecting both are acting in opposite directions. The next largest-magnitude correlation (-0.556) was between vehicle miles and walk/bike miles, indicating that much of the unexplained variation in these two variables is due to sources common to both. The negative correlation makes sense, in that when vehicle-miles is high, walk/bike miles will tend to be low, so that a variable affecting both is likely to do so in opposite directions. Looking at the other end, one of the smallest correlations was between the error variables for suburban and vehicle miles, reflecting a low level of commonality among unobserved variables for these two equations.

Table 55: Cross-Equation Error Correlations of Nontransformed Model

Dimensions:	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Endogenous Variables ®	Trad-itional	Sub-urban	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Traditional	1.0								
Suburban	0.11	1.0							
Pro-HD	-0.03	-0.02	1.0						
Pro-Driving	0.09	-0.05	0.04	1.0					
Pro-Transit	-0.30	-0.07	0.06	0.41	1.0				
Vehicle Miles	0.06	0.02	-0.02	0.07	-0.02	1.0			
Transit Miles	-0.13	-0.01	0.01	-0.35	0.13	-0.21	1.0		

Walk/Bike Miles	-0.04	-0.02	0.06	0.15	0.17	-0.56	0.06	1.0	
Commute Distance	-0.08	-0.38	-0.09	0.04	-0.02	0.001	-0.62	0.01	1.0

8.4.1.4 Addressing the Normality Assumption

Though many of the findings reported above supported past research (from other authors and earlier work from this study), the validity of the claims is theoretically dependent upon the structural equation models meeting its assumptions. Unfortunately, the assumption that the observed variables follow a multivariate normal distribution was not met. The Mardia statistic (a measure of multivariate kurtosis) calculated by the statistical software used in this study, AMOS 3.6, was 313.15 with a critical ratio of 72.28 (a critical ratio above 1.96 would signify departure from multivariate normality with 95% confidence).

This significant failure to meet the normality assumption could jeopardize the validity of the model results, and thus, modifications were made. The first step was to transform (taking the natural log or square root) the six variables (daily vehicle-miles, daily transit-miles, daily walk/bike-miles, adventurer, number of persons under the age of sixteen, and number of vehicles) that had high kurtosis values, as such transformations have been found to be potentially effective in making the distribution of a variable more normal (West *et al*., 1995). After re-estimating the previous model with the newly transformed variables, the resulting Mardia statistic was significantly better, but still a little over 100.

Though there is "currently little empirical and theoretical guidance available as to when a statistically significant variation from normality becomes large enough to affect structural modeling conclusions" (Bentler, 1989, pg. 227), this degree of departure was still deemed large enough to warrant further corrective action. The next step was to find and remove outliers, as "extreme data points may affect the results of structural equation modeling" (West *et al*., 1995, pg. 61). AMOS provided the Mahalanobis distance (see Everitt, 1993 for details) for each case in the data set, where the greater the Mahalanobis distance, the greater the contribution to the departure from multivariate normality. Based on this information, cases were removed ten at a time (to minimize the number of cases discarded) until the remaining data set had a multivariate normal distribution. The removal of 100 cases led to a sample of 515 respondents for which a structural equation model was estimated, resulting in a Mardia statistic of 0.55 with a critical ratio of 0.12. While the removal of apparent outliers (especially so many of them) was not an appealing step, the alternative of egregiously violating the model assumptions was even more unattractive. In support of this step, it can be noted that the reduced sample differed little from the larger sample in terms of mean values on key variables (as indicated in Section 4.3), and further, the findings from the model that met the assumption of multivariate normality were very similar to the results of the earlier model that did not meet the assumption. The results of this transformed model (i.e., the model that met the multivariate normality assumption) are presented and discussed next.

8.4.2 Transformed Model Results

As noted above, the originally estimated structural equations model did not meet the multivariate normality condition. Consequently, a second structural equations model estimation was carried out on a transformed data set (i.e., on that met the normality assumption). Care was taken to find a new, optimal model (i.e., an intuitively sound model with the best fit). Specifically, variables were added to and removed from each equation (based on conceptual reasoning and findings from Chapter 7) until the best identifiable model was found. Surprisingly, only one change needed to be made to the original system of structural equations, and that was the addition of the lifestyle variable relaxer to the equation for pro-driving. The directions and strengths of relationships among the variables were very similar (allowing for the effects of the transformations). Indeed, even the error correlation relationships yielded the same conclusions (e.g., that commute distance and transit miles had a great deal of commonality among their unobserved variables). Thus, the discussion of the substantive effects given for the earlier model applies here as well. However, for completeness, Tables 56 and 57 present the direct and total effects, respectively, for this transformed model.

8.4.3 Inspection of Model Fit

Interestingly, both models estimated in this study would be considered bad fits using the ?^{2} measure-of-fit, possibly due to the large sample sizes (N_3 = 615 and N_4 = 515). Table 58 presents a selection of model-fit indices for the two models (see, e.g., Arbuckle, 1997 and Hoyle, 1995 for a more complete discussion of fit measures). An inspection of the table shows that, by the measures in common use, both models fit the data well, with the original model (that did not meet the multivariate normality assumption) ironically having a slightly better fit.

Table 56: Direct Effects (t-stat.) of Transformed[1] Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urban	Pro-HD	Pro-Driving	Pro-Transit	Ln Vehicle Miles	Ln Transit Miles[1]	Ln Walk/ Bike Miles[1]	Commute Distance
Constant	1.822 (8.3)	-1.034 (-3.8)	0.204 (3.9)	-1.407 (-4.7)	-0.046 (-1.1)	3.838 (4.3)	0.823 (3.1)	0.160 (11.6)	-3.313 (-1.2)
Sociodemographic									
Age	-0.018 (-5.06)	0.009 (2.61)							
Female				0.543 (6.03)		-0.812 (3.63)			
Household Size	-0.081 (-1.92)								
Ln Number of People Under Age 16			-0.328 (-4.86)						
Sq Number of Vehicles	-0.761 (-5.44)					1.652 (5.17)	-0.870 (-6.30)		
Years Lived in Bay Area				0.004 (2.21)					
Lifestyle									
Ln Adventurer						0.554 (2.40)			
Culture-Lover		-0.178 (-4.48)	0.230 (5.80)						
Hobbyist		-0.099 (-2.75)							
Homebody		0.217 (5.50)							
Nest-Builder	-0.143 (-3.53)		-0.140 (-3.61)						
Outdoor Enthusiast			0.106 (2.87)						
Relaxer			0.042 (1.63)						

[1] Ln = Natural log transformation, Sq = Square root transformation

Table 56: Direct Effects (t-stat.) of Transformed[1] Res. Choice SEM - continued

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urban	Pro-HD	Pro-Driving	Pro-Transit	Ln Vehicle Miles[1]	Ln Transit Miles[1]	Ln Walk/ Bike Miles[1]	Commute Distance
Attitudes									
Pro-Alternatives		-0.070 (-1.57)						0.026 (2.13)	
Pro-Drive Alone					0.178 (4.21)				
Pro-Environment					0.161 (3.73)				
Pro-Growth							-0.058 (-2.07)		
Pro-Pricing	0.108 (3.11)	-0.074 (-2.23)							
Time-Satisfied				-0.109 (-3.19)					
Work-Driven			0.092 (2.55)						
Pro-High[2] Density	0.591 (5.00)								
Pro-Driving[2]			-0.702 (-4.41)					-0.082 (-2.00)	
Pro-Transit[2]							0.286 (2.56)		
Residential Choice/Preference									
Suburban[2]									2.149 (2.84)
Travel Demand									
Ln Vehicle Miles[2]				0.226 (4.07)					2.213 (4.89)
Ln Transit									5.732

Miles[2]									(4.96)
Ln Walk/Bike Miles[2]						-7.201 (-2.04)			
Job Location									
Commute[2] Distance		0.059 (2.62)					0.065 (3.24)		

[1] Ln = Natural log transformation, Sq = Square root transformation.
[2] Endogenous variable.

Table 57: Total Effects of Transformed[1] Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urban	Pro-HD	Pro-Driving	Pro-Transit	Ln Vehicle Miles[1]	Ln Transit Miles[1]	Ln Walk/ Bike Miles[1]	Commute Distance
Constant	0.052	0.015	0.053	0.101	-0.037	4.802	0.616	0.149	10.88
Sociodemographic									
Age	-0.018	0.011					0.002		0.038
Female	-0.172	-0.148	-0.292	0.416		-0.563	-0.161	-0.035	-2.489
Household Size	-0.081								
Ln Number of People Under Age 16	-0.194		-0.328						
Sq Number of Vehicles	-0.940	-0.091	-0.303	0.432		1.910	-0.969	-0.036	-1.523
Years Lived in Bay Area	-0.002		-0.003	0.005		0.003	0.001	-0.005	0.013
Lifestyle									
Ln Adventurer	-0.060	0.045	-0.102	0.145		0.641	0.184	-0.012	2.832

Culture-Lover	0.136	-0.224	0.230				-0.050		-0.765
Hobbyist		-0.124					-0.028		-0.425
Homebody		0.272					0.060		0.930
Nest-Builder	-0.226		-0.140						
Outdoor Enthusiast	0.063		0.106						
Relaxer	-0.020	0.008	-0.034	0.048		0.029	0.008	-0.004	0.127

[1] Ln = Natural log transformation, Sq = Square root transformation.

Table 57: Total Effects of Transformed[1] Res. Choice SEM - continued

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urban	Pro-HD	Pro-Driving	Pro-Transit	Ln Vehicle Miles[1]	Ln Transit Miles[1]	Ln Walk/ Bike Miles[1]	Commute Distance
Attitudes									
Pro-Alternatives	0.021	-0.146	0.035	-0.050		-0.220	-0.082	0.031	-1.272
Pro-Drive Alone		0.035			0.178		0.088		0.581
Pro-Environment		0.031			0.161		0.080		0.528
Pro-Growth		-0.040					-0.101		-0.664
Pro-Pricing	0.108	-0.093					-0.021		-0.318
Time-Satisfied	0.052	-0.020	0.089	-0.126		-0.076	-0.022	0.010	-0.334
Work-Driven	0.054		0.092						
Pro-High[2] Density	0.591								

Pro-Driving[2]	-0.479	0.182	-0.811	0.156		0.692	0.198	-0.096	3.056
Pro-Transit[2]		0.195					0.498		3.271
Residential Choice/Preference									
Suburban[2]		0.255					0.278		4.293
Travel Demand									
ln Vehicle Miles[2]	-0.108	0.304	-0.183	0.262		0.156	0.331	-0.022	5.111
ln Transit Miles[2]		0.681					0.742		11.450
ln Walk/Bike[2] Miles	0.780	-2.189	1.321	-1.883		-8.327	-2.385	0.156	-36.804
Job Location									
Commute[2] Distance		0.119					0.129		0.997

[1] Ln = Natural log transformation, Sq = Square root transformation.
[2] Endogenous variable.

Table 58: Comparison of Measures of Fit Between Structural Equation Models

	Original Model (N_3 = 615)	Transformed Model (N_4 = 515)
Degrees of freedom	214	213
?^{2}: measures discrepancy between the sample and population covariance matrices; the smaller the better.	417.524	554.100
?^{2}/d.f.: a "relative" chi-square value corrected for degrees of freedom; values in the range 1 to 3 are indicative of an acceptable fit, with values closer to 1 being better.	1.951	2.601
Normed Fit Index (NFI): proportion of worst (independence) model ?^{2} explained by the model of interest; varies between 0 and 1, with 1 being the best.	0.974	0.968
Relative Fit Index: NFI corrected for degrees of freedom; values close to one represent a good fit.	0.913	0.893
Incremental Fit Index: the incremental improvement of the model of interest over the worst (independence) model; values close to 1 indicate a good fit.	0.987	0.980
Comparative Fit Index: assumes a noncentral ?^{2} distribution for the worst (independence) model discrepancy; values close to 1 represent a good fit.	0.987	0.979
Akaike Information Criterion (AIC): balances discrepancy against complexity; in comparing two models, the smaller the better.	1469.524	1608.100
Brown-Cudeck Criterion: penalizes complexity more heavily than the AIC; in comparing two models, the smaller the better.	1538.927	1692.243

For brevity, only the results of the incremental fit indices (i.e., the relative fit index [R.F.I.], normative fit index [N.F.I.], and the comparative fit index [C.F.I.]) will be presented. Incremental fit indices compare a researcher-generated model to both a perfectly fitting model (i.e., the saturated model in AMOS; $?^2 = 0$) and a terribly fitting model (i.e., the independence model in AMOS) "to see how large the discrepancy function becomes between the saturated and independence model" (AMOS, 1997, pg. 563). For example, a model that has a small discrepancy (e.g., 3%) from the saturated model will have a high goodness-of-fit statistic (e.g., N.F.I. = 0.97), whereas a model that has a small discrepancy from the independence model will have a small goodness-of-fit statistic. A look at the range of values for the incremental fit indices, 0.893 for the relative fit index to 0.987 for the comparative and normative fit indices, shows that, in general, the model discrepancy functions are small (in comparison to the saturated model), indicating that the structural equation models estimated have reasonable fits. Ironically, the original model outperforms the transformed model on every single goodness-of-fit indicator. Nevertheless, the transformed model performs well in absolute terms, and is quite close to the original model on several indicators. These results, plus the structural similarity between the untransformed and transformed models suggest that the sizable departure from normality exhibited by the original model does not, in fact, materially affect the outcome, and that the relationships discussed previously are robust with respect to this empirical context.

8.5 Chapter 8 Summary

Chapter 8 was a critical component of this dissertation, presenting both the theory and results of the estimation of the residential choice/preference conceptual model on which this dissertation is based. The section following the introduction reintroduced the goal of operationalizing the conceptual model and discussed why the structural equation model (SEM) framework was chosen (including a brief discussion on the concept of causality in modeling). In particular, it noted how SEM procedures can obtain the coefficients for the direct relationships of a model system, as well as calculate the direction and extent of the combined impacts of interaction relationships. Next, theoretical and implementation issues of SEM were discussed, including achieving identification and meeting the multivariate normality assumption.

Specification and results were presented

for a modified conceptual model, including a discussion on the value of total effects analysis and cross-equation error correlation findings. Three key conclusions from these sections were: 1) no travel variables were directly significant to residential choice (and vice versa), 2) total effects must be taken into consideration for researchers to make accurate judgements about the impacts of explanatory variables on endogenous variables (indeed, it was found that if only direct effects are taken into consideration, researchers are likely to make inaccurate conclusions about the impacts of explanatory variables on endogenous variables -- as seen in the example of gender's effect on the attitude pro- driving), and 3) attitudinal and lifestyle variables were the most powerful explanatory variables in the structural equations models. The final chapter sections contained model fit results, and a brief mention of how a transformed conceptual model yielded similar results to the original residential choice/preference conceptual model. The major conclusions were that incremental fit indices indicated that both models fit the data well, and that the substantive effects from the original model also applied to the transformed model.

CHAPTER 9

CONCLUSIONS AND RECOMMENDATIONS

9.1 Introduction

The main objectives of this chapter are to synthesize the important ideas and findings of this work, and to discuss some productive directions for future research. For that purpose, the first eight chapters have been grouped into four sections: literature review foundation (Chapters 1 and 2), conceptual model creation (Chapter 3), data background and variable development (Chapters 4, 5, and 6), and modeling methodology and estimation (Chapters 5, 7, and 8). A brief summary of the major ideas and findings from each of the four sections is given next, followed by suggestions for future research work in this area.

9.2 Literature Review -- the Foundation (Ch. 1, 2)

The literature review for this study was instrumental in identifying advantages and disadvantages of various modeling methodologies, potential significant relationships among variables, core concepts, and suggestions for avoiding pitfalls (such as creating an endogenous variable with too many alternatives, choosing an inappropriate modeling strategy for study objectives, and claiming causality when it is actually association). The literature review was composed of two parts, one involving in-depth reviews of landmark studies and one involving short reviews of numerous inter-related studies.

The flagship study by Kain (1962a) was truly a model to follow. By coincidence, both

Kain’s study and this one had a system of 9 structural equations to define residential choice and travel behavior. By choice, this study, like Kain’s, used an econometric framework to study interdependent relationships between elements such as travel and residential choice. However, unlike Kain, this study did not assume a causal ordering of the equations (e.g., that a person first chooses a neighborhood, and then chooses a house). The issue of causality was important in this dissertation, and thus, a model structure that allowed some simultaneous causality (a more realistic assumption) was implemented.

Another landmark study was Tu and Goldfinch (1996). Unlike Kain (1962a), though, their study methodology was quite different from the methodology in this dissertation. Tu and Goldfinch pursued a market segmentation approach, whereas we did not allow coefficients to differ across households. However, it is believed that their approach has merit and some of the directions for future extensions of this dissertation are based on ideas from their work. Further, their conclusions that different households have different decision making processes was instrumental in the decision to permit bi-directional relationships among endogenous variables (rather than assuming the unidirectional causality embodied in Kain’s recursive structure).

The classic economic residential location models (see, e.g., Alonso, 1964) were introduced from the literature. These models can be viewed as a consumer allocation problem. In short, the models are based on the assumption that households have a fixed annual income which they use to buy land (a residence), travel (based mainly on commute distance), and other commodities. This constrained utility maximization process is generally described in terms of the relationship between rent or housing prices and distance from the central business district

(CBD). Also, various definitions of residential location used by previous researchers (such as census tracts and dwelling-unit categories), were presented along with discussion on how a measure of neighborhood type could be developed. These fundamental ideas were the basis for the development of the measures of traditionalness (endogenous variables) found in Chapter 6, and selection of explanatory variables used in the model specifications of Chapters 5, 7, and 8.

9.3 Creating the Conceptual Model (Ch. 3)

Findings from the literature review suggested that the residential choice process involves many interdependent relationships. A conceptual model of residential choice/preference was developed that was a comprehensive reflection of those relationships, supported by the literature and by informed judgement. No other single study was found by the author that collectively presented each of the relationships shown in the conceptual model. The resulting model is one of the key original contributions of this dissertation, and is its conceptual heart.

An important topic for the conceptual model section was the distinction between residential preference and choice. Though it was acknowledged that there was a conceptual difference between the two (primarily due to constraints), the difficulty of distinguishing between them with this study's data set prevented separate empirical analysis of them. It was found that most researchers did not model both separately in one study, and it is believed that the combination of the two concepts into one category does not significantly weaken this dissertation's empirical results and conclusions. In fact, one study that did look at both (see Louviere and Timmermans, 1990) concluded that individuals both form preferences and make choices by integrating information about attributes of residences. Thus, the significant explanatory variables in this study's residential choice/preference model are likely to shed insight into both preference and choice behavior.

The important next step in the conceptual model was the identification of numerous variables that were expected to influence residential choice. Specifically, variables were selected from the literature review of travel behavior/land use and spatial interaction studies, including previous models of residential choice, and presented in several tables found in Chapter 3. Many of these variables, such as age and household size, were found to be significant in the models developed in this study.

A visual illustration of the conceptual model was given in Figure 1, with arrows representing hypothesized relationships between the rectangles (which represented sets of variables defining a particular category such as *Neighborhood Characteristics*). The arrows implied a direction of causality, and, in modeling terms given in Chapter 8, each represented a direct effect of one type of variable on another. Numerous variables were interdependent, and consequently, many of the connecting arrows had heads at both ends. The conceptual model was explained by presenting examples that illustrated the various hypotheses in Figure 1. In addition, a majority of the examples had literature citations that supported the hypothesized relationship.

9.4 Data and Variable Development (Ch. 4, 5, and 6)

Chapters 4, 5 and 6 of this dissertation presented information on data background and variable development that were instrumental in the development and analysis of residential choice models.

Micro-scale data on land use, roadway networks, and public transit were collected in a 1992 land use-travel behavior study of 5 San Francisco (California) Bay Area neighborhoods. In addition, sociodemographics, attitudinal, lifestyle, and travel-related data were obtained from respondent surveys mailed to a random sample of residents in these neighborhoods. Since the primary goal of the original 1992 study was the analysis of travel behavior rather than residential choice, a notable weakness of the current study was that data on people's residential location decision-making processes, such as how important job location was in the selection of a specific residence and neighborhood, was not collected. The sections describing the data that were used (four primary study samples, N_1 - N_4) included a discussion on how missing data were handled through conditional mean and regression-based imputation.

Attitudinal, lifestyle, and residential choice variables were developed through factor analysis (see Chapter 5). Attitudinal and lifestyle variables are often missing in the literature on residential choice, but they were found to be highly significant predictors in this dissertation. Ten attitudinal and eleven lifestyle factor score measures were defined and discussed in relation to residential choice modeling.

The residential choice variables were unique to this study, and another useful contribution. Many researchers have defined residential location to be geographically specific (either micro or macro, such as census tracts, housing bundles, distance from central business district, and neighborhoods). This study focused on residential choice in terms of neighborhood type, specifically in terms of the traditional versus suburban nature of the neighborhood. Neighborhoods can be characterized along other dimensions (e.g., level of crime, level of wealth, etc.), but the dimension of traditionalness is much more common in studies of travel behavior, allowing the results of this study to fit into that context. For the regression and structural equation models developed here, factor analysis was used to define continuous measures of neighborhood type based on 18 characteristics identified from the literature and available in our data. It was felt that this type of residential choice measure was superior since it would be more transferable to other study situations. In other words, what generically characterizes a neighborhood is more of interest for residential choice modelers than a specific neighborhood itself. Further, using continuous measures of residential choice obviated the need to employ the more complex econometric methods required to handle discrete dependent variables in single equations or multiple-equation systems.

Several residential choice variables were developed (aggregate data versus disaggregate data, 1-factor structure versus 2-factor structure) in the study. The literature offered no guidance on which measure would be superior, and consequently, exploratory analysis was performed to try to identify which measures were more useful. Ultimately, all of the factor solutions made conceptual and empirical sense, but the two-factor disaggregate solution was preferred on both grounds. Further, the results showed that the concept of traditionalness is too complex to be captured by a single factor dimension. Figure 6, a plot of the individual factor scores (distinguished by neighborhood) for the two-factor disaggregate solution illustrated quite clearly that considerable variation on neighborhood type could occur within a neighborhood, and that a single neighborhood could strongly possess both traditional and suburban characteristics (Pleasant Hill, in particular, possessed both types of traits). Consequently, the two-factor disaggregate solution (traditional and suburban endogenous variables) was used in the structural equation model estimation of Chapter 8.

9.5 Modeling Methodology and Estimation (Ch. 5, 7, 8)

The eventual estimation of a modified residential choice structural equations model was the final step in a strategized sequence of simpler preliminary models. These simpler models, presented in Chapters 5 and 7, provided crucial insight into the specification and estimation of the more complex multi-equation structural models described in Chapter 8.

9.5.1 Preliminary Single-Equation Models (Ch. 5, 7)

As noted earlier, categorizing neighborhoods as traditional or suburban was not straightforward (despite the fact that a great deal of rich data was available for each of the neighborhoods, some based on site visits). Thus, for the first residential choice model, two of the neighborhoods (South San Francisco and Pleasant Hill) that were suspected to be more of a blend of both traditional and suburban types were removed, to help clarify the choice between the two extreme types. A binary dependent variable for residential choice (equal to zero if a respondent lived in traditional North San Francisco and equal to one if a respondent lived in suburban Concord or San Jose) was used in a logit model of residential choice. A major finding from this model was that attitudinal and lifestyle factors possessed greater explanatory power than sociodemographic variables for the variation in residential choice. Further, expected variable relationships from the conceptual model were supported by the model coefficients. As an example, it was hypothesized that larger families would be more likely to choose suburban neighborhoods over traditional neighborhoods (as larger homes would tend to be more affordable in the suburbs); the variable "number of people under 16 " had a positive sign, indicating that, indeed, larger families were more likely to choose a suburb. Though convincing results were obtained from the binary logit model, such a simplistic measure of residential choice would not be useful in estimating the more complex, multi-equation structural model. First, as noted earlier, a categorical endogenous variable would have created statistical difficulties in the structural equation modeling of Chapter 8. Second, measuring neighborhood type in dichotomous terms was conceptually less realistic than capturing the continuous variation that was observed to exist in our data, and offered little insight into the choice of the hybrid neighborhoods that were dropped from this portion of the analysis (thus, discarding valuable data).

Chapter 7 contained numerous single-equation regression equations based on continuous endogenous variables. The first set of nine equations did not allow any endogenous variables on the right hand side (and thus, met OLS assumptions), while the next set of nine equations did (thus violating OLS assumptions). Each of these equations represented a relationship from the conceptual model, and it was expected that significant variables from these models would likely be important in the final, full model (which, in fact, they generally were). The model for traditionalness had the best fit (R^2-adj. = 0.39) of the nine exogenous-only single-equation model estimations, while the model for daily walk/bike-miles rate had the poorest fit (R^2-adj. = 0.03). Common sociodemographic variables such as household size and number of vehicles were significant in many of the single-equation models. However, the attitudinal and lifestyle variables also played a major role in explaining endogenous variable variation.

A new finding was that endogenous variables were highly significant in the single-equation models, supporting the characterization of the conceptual model in terms of interdependent relationships. Given this, the first step toward simultaneous estimation of the conceptual model was the estimation of the two measures of residential choice (traditional and suburban) using seemingly unrelated regression (SUR) procedures. Two main results of the SUR estimation were: 1) there were slight differences in the number and type of significant explanatory variables between the single and simultaneous models, reflecting that interaction is occurring among the variables in the simultaneous model, and 2) the variables that were significant in both models uniformly had higher t-statistics in the SUR model, reflecting the greater efficiency of SUR over single-equation procedures.

9.5.2 Structural Equations Models (Ch.8)

Using the significant relationships found in Chapter 7, a system of nine equations was simultaneously estimated using a maximum likelihood procedure. The insights obtained from this system of equations represent another contribution of this study. A reassuring finding was that many of the variables that were significant in prior models were also significant in the simultaneous model system. However, many variables did lose their power (t-statistic went below 1.64) and were removed from a particular equation; it is important to note that they did not, in general, become insignificant in the entire model system, but just were not significant in as many equations as at first. On the other hand, a small number of variables that were not significant before did become significant in the simultaneous model structure. These two results confirm that the simultaneous structure does involve interactions that impact model specification, and hence that modeling system equations one at a time is inferior.

A major advantage of the structural equation models was their ability to capture interaction effects among multiple variables simultaneously. Total effects, the sum of direct and indirect effects, of the explanatory variables on the endogenous variables were computed from the coefficient matrices representing direct effects. The arrows in the conceptual model (Figure 1) are intended to represent direct effects only. Indirect effects, on the other hand, occur when a variable impacts a variable through an intervening variable. Specifically, when all variables are standardized, the indirect effect of one variable on another (through one or more intervening variables) is the product of the corresponding coefficients (i.e., the direct effect coefficients) that link the two variables in question together.

It was found that the apparent power of the direct effects may be very misleading due to interaction effects among variables (i.e., indirect effects), and consequently, it was important to consider indirect effects when making any conclusions. For example, consider the effect of the variable female on pro-driving for the standardized model discussed in Chapter 8. First, female has a positive direct effect on pro-driving (structural coefficient of 0.275) and a negative direct effect on vehicle miles (structural coefficient of -0.165). Second, vehicle miles has a positive direct effect on pro-driving (structural coefficient of 0.554). Vehicle miles, then, is an intervening variable between female and pro-driving, creating a negative indirect effect of female on pro-driving through vehicle miles (indirect effect = -0.165*0.554 = -0.091). Thus, the total effect of female on pro-driving is less than the direct effect of female on pro-driving. Clearly, such impacts from indirect effects can impact researcher conclusions.

A particularly noteworthy result from the multi-equation model was that no travel variables were directly significant to residential choice, nor were any residential choice variables directly significant to travel (see Table 53). By contrast, attitude and lifestyle variables had a large, direct impact on travel demand. When total effects were analyzed (see Table 54), it was found that travel variables and residential choice variables do influence each other through indirect effects. However, both the direct and total effects results indicate that attitude and lifestyle variables play the greatest role in explaining residential choice and travel demand. Taken together, this supports the speculation that the association commonly observed between neighborhood type and travel patterns is not one of direct causality, but primarily due to correlations of each of those variables with others. In particular, these results suggest that when attitudinal, lifestyle, and sociodemographic variables are accounted for, neighborhood type has very little influence on travel behavior.

9.6 Recommended Directions for Future Research

Several directions for research appear promising. With the currently available data, future work could be done on defining different measures of residential choice (using other combinations of neighborhood characteristics or even other location definitions such as type of dwelling unit – single family home, etc. – and type of area – countryside, small city, etc.) and testing how well these fit as dependent variables in residential choice models. Analyzing different location measures would be an improvement over focusing on only one aspect of a neighborhood (in this case, its degree of suburbanness versus traditionalness). The explanatory variables measured in this study may point to one type of preferred neighborhood, while a respondent could live in a different-than-predicted type of neighborhood due to other positive characteristics of it that we did not address, and/or constraints that were not measured or properly incorporated. This supports the need for more future efforts on characterizing a neighborhood more completely.

Sample segmentation by variables such as number of children, number of workers, and vehicle availability may improve future models of residential choice by increasing the homogeneity of the resulting segments. This was already done on a small scale by the removal of the retired and unemployed respondents. On that note, a study of the residential choice of the retired sample may yield interesting, and possibly quite different, results. Powerful explanatory variables like job location are likely to be replaced by variables more pertinent to non-working individuals, such as degree of safety and access to recreation. On the other hand, the history dependence or inertia effect (living where one lived before retirement, because one hasn't gotten around to moving yet) is likely to be strong and cannot be neglected in such a study.

The collection of new data would increase the possibilities of extending this work. It would be desirable for a rigorous research plan to be designed to capture a household's decision making structure, ideally while in the process of choosing a residential location. Such a plan would require the identification of households that are planning to relocate in a year or less. All members (of such households) over 16 would be surveyed to collect information on the factors that were important in the decision to choose a new residential location (as well as sociodemographic and travel diary information). Factors that may be found significant could include job re-location, family-size change and school quality. Ideally, questions would be designed to measure the relative influence individual household members had in the final choice of residence; constraints may even be identified that help distinguish preference and choice. For example, questions could include: "If you had the final decision, would this have been your residential choice?", (if no) "On a scale of 1 to 10 (10 being the highest rating), how much influence did you have in choosing your residence?", and "What constraints, if any, prevented you from selecting your first choice of residence?" Though this type of study design may add to the understanding of an individual's or household's residential choice/preference, it is essentially a one-move study, and hence limited in its ability to show the dynamics and causality of residential choice.

The collection of longitudinal data would be the best approach for resolving causality in an individual's or household's residential choice process. The incorporation of panel data on attitudes, lifestyles, travel, sociodemographic, and residential variables into a structural equation model would allow the dynamics of the interactions between the previously mentioned variables to be studied. For example, over a period of years one may be able to study whether individual attitudes and lifestyle changed, whether residential location influenced any changes in attitudes and lifestyle, and whether behavior (such as travel demand) changed along with them. A longitudinal study with a focused research design like the one mentioned above could offer the best hope for disentangling relationships of mutual causation found in the residential choice conceptual model of this dissertation.

9.7 Contributions of this Work

There were three main interrelated objectives for this dissertation: 1) to present a realistic conceptual model of the relationships among residential choice, job location, travel behavior, and other major components involved with spatial interaction modeling at a disaggregate level, 2) to provide a better understanding, through empirical models, of what motivates an individual to choose a certain type of residential location, and of the relationship between residential location and travel behavior, and 3) to identify the role that attitudinal and lifestyle variables play in people's residential location choice and travel behavior. Each of these goals has been advanced, and has resulted in some useful contributions to the land use and travel research area.

The conclusions and implications of a model fully depend on the specification of the relationships involved in the process or behavior under consideration. One unexpected finding from the literature review was the absence of a study that fully specified the many interdependent relationships involved with residential choice. In the papers reviewed, the authors only looked at portions of the residential choice process defined by the conceptual model in this dissertation. Indeed, the completeness of the model, along with its literature support of hypotheses, is one accomplishment of this dissertation.

The measurement of neighborhood type as a continuous variable through factor analysis was another important part of this work. Many researchers have used geographically-specific measures of residential location (such as census tracts and neighborhoods), but these measures have limited transferability to other geographic contexts. Other researchers have used discrete indicators such as suburban versus traditional, but this is limiting in that there can be considerable variation within such types of residential location. A two-factor disaggregate solution representing traditional and suburban neighborhood dimensions was used in this study. This approach allowed for a single area to possess attributes of both types of neighborhoods, and allowed individuals within the same area to face different neighborhood characteristics – a flexibility amply justified by the empirical results. Further, it has the statistical advantage of producing continuous measures of endogenous variables, a trait that is desirable in both regression and structural equation models.

A large amount of the research on transportation/land use interactions to date has mainly consisted of pointing out correlations between travel patterns and characteristics of urban form. An important next step taken in this study is the development of insight into the behavior underlying these correlations (i.e., the causal relationships). The structural-equation methodology used here is noted for being able to provide insight into causality due to its ability to measure interaction effects among variables. Specifically, SEM procedures allow the total effects, the sum of direct and indirect (interaction) effects, of the explanatory variables on the endogenous variables to be estimated. The arrows in the conceptual model (Figure 1) are intended to represent direct effects only. Indirect effects, on the other hand, occur when a variable impacts a variable through an intervening variable. The difference between direct and total effects can be substantial, and consequently, it is important to consider both effects when making research conclusions. The value of the SEM approach was readily seen in the analysis of the direct and total effects results of the models, where inferences based only on the direct effects were sometimes different than those based on the total effects (showing that single-equation model conclusions can be erroneous) due to not incorporating interaction effects.

A particularly noteworthy result from the multi-equation model was that no travel variables were directly significant to residential choice, nor were any residential choice variables directly significant to travel (see Table 53). By contrast, attitude and lifestyle variables had a large, direct impact on travel demand. When total effects were analyzed (see Table 54), it was found that travel variables and residential choice variables do influence each other through indirect effects. The direct and total effects results indicate that attitude and lifestyle variables play the greatest role in explaining residential choice and travel demand. In particular, these results suggest that when attitudinal, lifestyle, and sociodemographic variables are accounted for, neighborhood type has very little influence on travel behavior.

The measurement and incorporation of attitudinal and lifestyle measures into residential choice models was another accomplishment. Given that a person's perceptions, attitudes, lifestyles and preferences have a significant impact on all of her or his decisions, the use of attitudinal and lifestyle variables in the estimation of models seems natural. However, few residential choice studies have incorporated attitudinal variables (see for example, Prevedouros, 1992), with the present study advancing the state of the art in that regard. Attitudinal and lifestyle variables were found to have the largest explanatory power of all the variables included in the discrete choice models of Chapter 5, the regression models of Chapter 7, and the structural equation models of Chapter 8.

In conclusion, this dissertation has pointed to several ways in which current residential choice models can be improved. Most important, improved models can lead to a better understanding of the underlying factors motivating an individual or household to select a particular type of residential location. In turn, transportation and urban planners can better understand the market for various neighborhood types, and improve their ability to forecast the travel impacts of different land use configurations.

Bibliography

Aldana, E., Neufville, R., and Stafford, J. (1973). Microanalysis of urban transportation demand. *Highway Research Record* 446, pp. 1-11.

Alonso, W.A. (1964). Location and Land Use. Harvard University Press, Cambridge, Massachusetts.

American Housing Survey (1997). Appendix A: Definitions. URL: http://www.census.gov/pub/hhes/housing/ahs/appa.html.

AMOS (1997). AMOS version 3.6. SmallWaters Corporation. Chicago.

Anas, A. (1982). Residential Location Markets and Urban Transportation: Economic Theory, Econometrics and Policy Analysis with Discrete Choice Models. Academic Press, New York.

Ansah, J. A. (1977). Destination choice set definition in travel behavior modeling. *Transportation* **11**, pp. 127-140.

Asher, Herbert (1976). Causal Modeling. Sage University Papers. Series: Quantitative Applications in the Social Sciences. Number 07-003.

Atash, Farhad (1993). Mitigating traffic congestion in suburbs: An evaluation of land-use strategies. *Transportation Quarterly* **47** (4), pp. 507-524.

Ben-Akiva, Moshe and Bowman, John L. (1998). Integration of an activity-based model system and a residential location model. *Urban Studies* **35** (7), pp. 1131-1153.

Bentler, P. and Speckart, G. (1981). Attitudes “cause” behaviors: A structural equation analysis. *Journal of Personality and Social Psychology* **40** (2), pp. 226-238.

Berry, William (1984). Nonrecursive Causal Models. Sage University Papers. Series: Quantitative Applications in the Social Sciences. Number 07-037.

Blumenberg, Evelyn and Ong, Paul (1997). Can welfare recipients afford to work far from home? *Access* **10**, pp. 15-19.

Boarnet, Marlon G. and Sarmiento, Sharon (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.

Boehm, Thomas and Ihlanfeldt, Keith (1991). The revelation of neighborhood preferences: An n-chotomous multivariate probit approach. *Journal of Housing Economics* **1**, pp. 33-59.

Carnahan, D., Galle, O., and Guest, A. (1974). Congestion, concentration and behavior: Research in the study of urban population density. *The Sociological Quarterly* **15**, pp. 488-506.

Cervero, Robert (1996a). Traditional neighborhoods and commuting in the San Francisco Bay Area. *Transportation* **23**, pp. 373-394.

Cervero, Robert (1996b). Mixed land-uses and commuting: Evidence from the American Housing Survey. *Transportation Research* **30A** (5), pp. 361-377.

Cervero, R. and Radisch, C. (1996). Travel choices in pedestrian versus automobile oriented neighborhoods. *Transport Policy* **3** (3), pp. 127-141.

Cervero, R. And Kockelman, Kara (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research* **2D** (3), pp. 199-219.

Cho, Cheol-Joo (1997). Joint choice of tenure and dwelling type: A multinomial logit analysis for the city of Chongju. *Urban Studies* **34** (9), pp. 1459-1473.

Christensen, Ronald (1990). Log-Linear Models. Springer Texts in Statistics, Springer-Verlag.

Clif, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research* **18**, pp. 115-126.

Crane, Randall (1996). Cars and drivers in the new suburbs: Linking access to travel in neotraditional planning. *Journal of the American Planning Association* **62** (1), pp. 51 - 65.

Cullinane, Sharon (1992). Attitudes towards the car in the U.K.: Some implications for policies on congestion and the environment. *Transportation Research* **26A** (4), pp. 291-301.

Daly, Andrew (1982). Applicability of disaggregate models of behavior: A question of methodology. *Transportation Research* **16A** (5-6), pp. 363-370.

Dasgupta, Indraneel (1996). Distributive implications of women's employment in a gender-segmented labor market with intra-household bargaining. Working paper in economics, University of California-Riverside, no. 96-15.

DeSalvo, Joseph and Huq, Mobinul (1996). Income, residential location, and mode choice. *Journal of Urban Economics* **40**, pp. 84-99.

Dobson, R., Dunbar, F., Smith, C., Reibstein, D., and Lovelock, C. (1978). Structural models for the analysis of traveler attitude-behavior relationships. *Transportation* **7**, pp. 351-363.

Dubin, Robin (1985). Transportation costs and the residential location decision: A new approach. *Journal of Urban Economics* **17**, pp. 58-72.

Dussault, Raymond (1997). Best places to live. *Sacramento Magazine* **23** (3), pp. 35-44.

Edin, Nancy (1966). Residential Location and Mode of Transportation to Work: A Model of Choice. Chicago, Chicago Area Transportation Study. Sponsored by the State of Illinois, Cook County, and city of Chicago, in cooperation with the U.S. Bureau of Public Roads.

Epstein, Lee (1996). Integrating the environment and land use into transportation decision making. *TR News* 187.

Etzioni, A. And Lehman, E. (1967). Some dangers in "valid" social measurement. *The Annals* 373, pp. 1-15.

Ewing, R., Haliyur, P., and Page, G. (1994). Getting around a traditional city, a suburban planned unit development, and everything between. *Transportation Research Record* 1466, pp. 53-62.

Forkenbrock, D. and Foster, N. (1990). Economic benefits of a corridor highway investment. *Transportation Research* **24A** (4), pp. 303-312.

Frank, Lawrence and Pivo, Gary (1994). Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit, and walking. *Transportation Research Record* 1466, pp. 44-52.

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

Galster, G., and Hesser, G. (1981). Residential satisfaction: Compositional and contextual correlates. *Environment and Behavior* **13** (6), pp. 735-758.

Gilbert, G., and Foerster, J. (1977). The importance of attitudes in the decision to use mass transit. *Transportation* **6**, pp. 321-332.

Giuliano, Genevieve (1989). New directions for understanding transportation and land use. *Environment and Planning* **21A**, pp. 145-159.

Giuliano, Genevieve (1996). Transportation, land use, and public policy. *TR News* 187.

Golob, T. (1990). Structural equation modeling of travel choice dynamics. Located in Developments in Dynamic and Activity-Based Approaches to Travel Analysis, Aldershot, England; Avebury, copywright 1990; pp. 343-370.

Gordon, Patrice Lynette (1990). An Examination of Wasteful Commuting Using a Discrete Residential Location Choice Model. Ph.D Dissertation, Department of Economics, University of Maryland College Park, Ann Arbor, Michigan.

Greene, W. (1990). Econometric Analysis. Prentice Hall, New Jersey.

Griffiths, W., Hill, R., and Judge, G. (1993). Learning and Practicing Econometrics. John Wiley and Sons, Inc., New York.

Guadagnoli, Edward and Velicer, Wayne (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin* **103(2)**, pp. 265-275.

Hamilton, B. and Roell, A. (1982). Wasteful commuting. *Journal of Political Economy* **90** (5), pp. 1035-1053.

Handy, Susan (1993). Regional versus local accessibility: implications for nonwork travel. *Transportation Research Record* 1400, pp. 58-66.

Handy, Susan (1996a). Understanding the link between urban form and nonwork travel behavior. *Journal of Planning Education and Research* **15** (3), pp. 183-198.

Handy, Susan (1996b). Urban form and pedestrian choices: A study of Austin neighborhoods. *Transportation Research Record* 1552, pp. 135-144.

Handy, Susan (1996c). Methodologies for exploring the link between urban form and travel behavior. *Transportation Research* **1D** (2), pp. 151-165.

Handy, Susan (1997). Travel behavior - land use interactions: An overview and assessment of the research. Resource paper prepared for the International Association of Travel Behavior Research 1997 Triennial Meeting, Austin, Texas.

Harbaugh, Carol and Smith, Theresa (1998). Welfare reform and transportation: There is a connection. *Public Roads* **81** (1), pp. 38-43.

Heckman, J. Dummy endogenous variables in a simultaneous equation system. *Econometrica* **46** (4), 931-959.

Heikkila, E., Gordon, P., Kim, J., Peiser, R., Richardson, H., and Dale-Johnson, D. (1989). What happened to the CBD-distance gradient?: land values in a policentric city. *Environment and Planning* **21A**, pp. 221-232.

Hensher, D. and McLeod, P. (1977). Towards an integrated approach to the identification and evaluation of the transport determinants of travel choices. *Transportation Research* **11**, pp. 77-93.

Hensher, D. and Johnson, Lester (1981). Applied Discrete-Choice Modeling. A Halsted Press Book, John Wiley and Sons, New York.

Hensher, D. And Taylor, A. (1983). Intraurban residential relocation choices for students: an empirical inquiry. *Environment and Planning* **15A**, pp. 815-830.

Horowitz, A.D. (1978). A cognitive dissonance approach to attitudinal modeling in travel behavior research. Proceedings of the annual transportation research board meeting.

Horowitz, Joel, L. (1995). Modeling choices of residential location and mode of travel to work. In The Geography of Urban Transportation, Guilford Press, New York, pp. 219-239.

Houghton, A. G. (1971). The simulation and evaluation of housing location. *Environment and Planning* **3**, pp.383-394.

Hunt, J.D., McMillan, J.D.P., and Abraham, J.E. (1994). Stated preference investigation of influences on attractiveness of residential locations. *Transportation Research Record* 1466, pp. 79-87.

Johnson, R., and Wichern, D. (1992). Applied Multivariate Statistical Analysis. Prentice Hall, New Jersey.

Joseph, A., Smit, B. and McIlravey, G. (1989). Consumer preferences for rural residences: A conjoint analysis in Ontario, Canada. *Environment and Planning* **21B**, pp. 47-63.

Kain, John Forrest (1961). The Journey to Work as a Determinant of Residential Location. Ph.D. Dissertation, Department of Economics, University of California at Berkeley.

Kain, J.F. (1962a). A Contribution to the Urban Transportation Debate; An Econometric Model of Urban Residential and Travel Behavior. The Library of the University of California Los Angeles, Rand Corporation Paper, P-2667.

Kain, J.F. (1962b). A Multiple Equation Model of Household Locational and Tripmaking Behavior. Memorandum, RM-3086-FF of the Rand Corporation, Santa Monica, California.

Kain, J.F. and Quigley, J.M. (1970). Evaluating the quality of the residential environment. *Environment and Planning* **2**, pp. 23-32.

Kain, J.F. and Fauth, G.R. (1976). The Effects of Urban Structure on Household Auto Ownership Decisions and Journey to Work Mode Choice. Department of City and Regional Planning, Harvard University, Cambridge, Massachusetts. Research Report R76-1.

Kain, J.F. and Apgar Jr., W.C. (1985). Housing and Neighborhood Dynamics - A Simulation Study. Harvard University Press.

Kitamura, Ryuichi, Nishii, Kazuo, and Goulias, Konstadinos (1990). Trip chaining behavior by central city commuters: A causal analysis of time-space constraints. Located in Developments in Dynamic and Activity-Based Approaches to Travel Analysis, Aldershot, England; Avebury, copywright 1990; pp.145-170.

Kitamura, Ryuichi, Laidet, Laura, Mokhtarian, Patricia, L., Buckinger, Carol, and Gianelli, Fred (1994). Land Use and Travel Behavior. University of California Davis Institute of Transportation Studies Research Report Number UCD-ITS-RR-94-27.

Kitamura, Ryuichi, Mokhtarian, Patricia, L., and Laidet, Laura (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation* **24** (2), pp. 125-158.

Knoke, D. and Burke, P. (1980). Log-Linear Models. Sage University Papers. Series: Quantitative Applications in the Social Sciences. Number 07-020.

Kroes, E. and Sheldon, R. (1988). Stated preference methods: An introduction. *Journal of Transport Economics and Policy* **22**, pp.11-25.

Lansing, John B. and Marans, Robert W. (1969). Evaluation of neighborhood quality. *American Institute of Planners* **35**, pp. 195-199.

Lerman, Steven R. (1975). A Disaggregate Behavioral Model of Urban Mobility Decisions. Ph.D. Dissertation, Department of Civil Engineering, MIT, Cambridge, Massachusetts.

Lierop, Wal van (1986). Spatial Interaction Modeling and Residential Choice Analysis. Gower Publishing Company Limited, England.

Lindstrom, Bonnie (1997). A sense of place: Housing selection on Chicago's north shore. *The Sociological Quarterly* **38** (1), pp. 19-39.

Little, Roderick and Rubin, Donald (1990). The Analysis of Social Science Data with Missing Values. *Sociological Methods and Research* **18** (3), pp. 292-326.

Long, J.S. (1983). Confirmatory Factor Analysis. Sage University Papers. Series: Quantitative Applications in the Social Sciences. Number 07-033.

Long, J.S. (1983). Covariance Structure Models - An Introduction to LISREL. Sage University Papers. Series: Quantitative Applications in the Social Sciences. Number 07-034.

Los, M. and Nguyen, S. (1983). Solution algorithms for a combined residential location and transportation model. *Environment and Planning* **15A**, pp. 515-524.

Louviere, Jordan (1979). Modeling individual residential preferences: A totally disaggregate approach. *Transportation Research* **13A** (6), pp. 373-384.

Louviere, Jordan and Timmermans, Harry (1990). Hierarchical information integration applied to residential choice behavior. *Geographical Analysis* **22** (2), pp. 127-144.

Lu, M (1998). Analyzing migration decisionmaking: relationships between residential satisfaction, mobility intentions, and moving behavior. *Environment and Planning* **30A**, pp. 1473-1495.

Madanipour, Ali (1996). Urban design and dilemmas of space. *Environment and Planning* **14D**, pp. 331-355.

McFadden, Daniel (1978). Modeling the choice of residential location. *Transportation Research Record* 673, pp. 72-77.

Maddala, G.S. (1983). Limited-Dependent and Qualitative Variables in Econometrics. Econometric Society Monographs No. 3. Cambridge University Press.

Madden, J.F. (1981). Why women work closer to home. *Urban Studies* **18**, pp. 181-194.

Manski, C. and McFadden, D. (1981). Structural Analysis of Discrete Data with Econometric Applications. The MIT Press.

Menchik, M. (1972). Residential environmental preferences and choice: empirically validating preference measures. *Environment and Planning* **4**, pp. 445-458.

Middleton, D. (1997). Cohousing development: Creating genuine communities. *Urban Land* **57** (7), pp. 32-35.

Mokhtarian, Patricia L. and Salomon, Ilan (1997). Modeling the desire to telecommute: The importance of attitudinal factors in behavioral models. *Transportation Research* **31A** (1), pp. 35-50.

Mulaik, Stanley A. (1987). Toward a conception of causality applicable to experimentation and causal modeling. *Child Development* **58**, pp. 18-32.

Mulaik, Stanley A. (1993). Objectivity and multivariate statistics. *Multivariate Behavioral Research* **28** (2), pp. 171-203.

Mueller, Ralph O. (1996). Basic Principles of Structural Equation Modeling - An Introduction to LISREL and EQS. Springer-Verlag New York, Inc.

Muth, R. (1969). Cities and Housing. University of Chicago Press, Chicago.

Newman, P. And Kenworthy, J. (1989). Gasoline consumption and cities: A comparison of U.S. cities with a global survey. *Journal of the American Planning Association* **55** (1), pp. 24-37.

Nijkamp, P., Wissen, L.V., and Rima, A. (1993). A household life cycle model for residential relocation behavior. *Socio-Economic Planning Sciences* **27** (1), pp. 35-53.

Norusis, Marija (1990). SPSS/PC + Advanced Statistics 4.0. Chicago: SPSS Inc.

Onaka, J. (1983). A multiple-attribute housing disequilibrium model of residential mobility. *Environment and Planning* **15A**, pp. 751-765.

Parsons Brinckerhoff Quade and Douglas (1993). The pedestrian environment: Making the land use, transportation, air quality connection. LUTRAQ Project, volume 4A, Portland, Oregon.

Pendyala, Ram (1993). Causal Modeling of Travel Behavior Using Simultaneous Equations Systems: A Critical Examination. Ph.D. Dissertation, Department of Civil Engineering, University of California at Davis. University of California Davis Institute of Transportation Studies Research Report Number UCD-ITS-RR-92-19.

Polak, John and Han, Xiao-Liang (1997). Iterative imputation based methods for unit and item non-response in travel diary surveys. Paper prepared for the International Association of Travel Behavior Research 1997 Triennial Meeting, Austin, Texas.

Porell, Frank W. (1982). Models of intraurban residential relocation. Studies in applied regional science. Kluwer-Nijhoff publishing, Boston.

Pratt, Andy (1996). Coordinating employment, transport and housing in cities: An institutional perspective. *Urban Studies* **33** (8), pp. 1357-1375.

Prevedouros, Panos (1992). Associations of personality characteristics with transport behavior and residence location decisions. *Transportation Research* **26A**, pp. 381-391.

Putman, Stephen H. (1979). Urban Residential Location Models. Studies in applied regional science, Vol. 13. Martinus Nijhoff Publishing.

Quigley, John (1985). Consumer choice of dwelling, neighborhood, and public services. *Regional Science and Urban Economics* **15**, pp. 41-63.

Recker, Wilfred W. and Stevens, Richard F. (1976). Attitudinal models of modal choice: The multinomial case for selected nonwork trips. *Transportation* **5**, pp. 355-375.

Richards, Martin (1982). Disaggregate demand models-promises and prospects. *Transportation Research* **16A** (5), pp. 339-344.

Rummel, R. (1970). Applied Factor Analysis. Evanston, Northwestern University Press.

Ryan, S. and McNally, M. (1995). Accessibility of neotraditional neighborhoods: A review of design concepts, policies, and recent literature. *Transportation Research* **29A**, pp. 87-105.

Rutherford, G. Scott, McCormack, Edward, and Wilkinson, Martina (1996). Travel impacts of urban form: implications from an analysis of two Seattle area travel diaries. Presented at the Travel Model Improvement Program Conference on Urban Design, Telecommuting, and Travel

Behavior. Williamsburg, Virginia. October 27-30.

Salomon, I., and Ben-Akiva, M. (1983). The use of the life-style concept in travel demand models. *Environment and Planning* **15A**, pp. 623-638.
Sawicki, David and Flynn, Patrice (1996). Neighborhood indicators: A review of the literature and an assessment of conceptual and methodological issues. *Journal of the American Planning Association* **62** (2), pp. 165-181.

Scheffler, Olivia S., Schwartz, Seymour I., and Tardiff, Timothy J. (1979). Energy Conservation Attitudes and Behavior in Small Cities: Davis and Woodland, California. Environmental Quality Series No. 31, Institute of Governmental Affairs and Institute of Ecology, University of California, Davis; June.

Shin, Jung-Chuel (1985). Residential Location Choice: A Nested Logit Model with Transportation and Housing Characteristics. Ph.D. Dissertation, Department of Civil Engineering, Stanford University, California.

Sichelman, Lew (1997). Demographics favor apartments ... and bigger houses. *Urban Land* **57** (7), pp. 18.

Simpson, Wayne (1987). Workplace location, residential location, and urban commuting. *Urban Studies* **24**, pp. 119-128.

Skaburskis, Andrejs (1997). Gender differences in housing demand. *Urban Studies* **34** (2), pp. 275-320.

Southworth, Michael (1997). Walkable suburbs? An evaluation of neotraditional communities at the urban edge. *Journal of the American Planning Association* **63** (1), pp. 28-44.

Stanek, David (1995). Modeling perceptions and preference of home-based and center-based telecommuting. Master's Thesis, Department of Civil and Environmental Engineering, University of California at Davis. University of California Davis Institute of Transportation Studies Research Report Number UCD-ITS-RR-95-12.

Steiner, Ruth L. (1994). Residential density and travel patterns: Review of the literature. *Transportation Research Record* 1466, pp. 37-43.

Steptoe, R. and Thornton, C. (1986). Differential influence of an interstate highway on the growth and development of low-income minority communities. *Transportation Research Record* 1074, pp. 60-68.

Stopher, Peter R. And Ergun, Gokmen (1982). The effect of location on demand for urban recreation trips. *Transportation Research* **16A** (1), pp. 25-34.

Tardiff, Timothy J. (1977). Causal inferences involving transportation attitudes and behavior. *Transportation Research* **11**, pp. 397-404.
Tardiff, Timothy J., and Scheffler, Olivia S. (1979). Travel Behavior in Small Urban Areas of Yolo County, California. Institute of Governmental Affairs and Institute of Ecology, University of California, Davis; report 37.

Timmermans, H., Borgers, A., and Oppewal, H. (1992). Residential choice behavior of dual earner households: a decompositional joint choice model. *Environment and Planning* **24A**, pp. 517-533.

To, Minh Chau, Lapointe, Alain, and Kryzanowski, Lawrence (1983). Externalities, preferences, and urban residential location: Some empirical evidence. *Journal of Urban Economics* **14**, pp. 338-354.

Tong, C.O. and Wong, S.C. (1997). The advantages of a high density, mixed land use, linear urban development. *Transportation* **24**, pp. 295-307.

Tu, Yong and Goldfinch, Judy (1996). A two-stage housing forecasting model. *Urban Studies* **33** (3), pp. 517-537.

Uslaner, Eric M. (1976). Editor introduction to Causal Modeling. Sage University Papers. Series: Quantitative Applications in the Social Sciences.

Veldhuisen, K. and Timmermans, H. (1984). Specification of individual residential utility functions: a comparative analysis of three measurement procedures. *Environment and Planning* **16A**, pp. 1573-1582.

Velicer, Wayne and Jackson, Douglas (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research* **25** (1), pp. 1-28.

Verster, A. C. P. (1985). Commuting costs and the residential mobility of job changers. *Transportation Planning and Technology* **10**, pp. 193-207.

Waddell, Paul (1993). Exogenous workplace choice in residential location models: Is the assumption valid? *Geographical Analysis* **25** (1), pp. 65-82.

Webber, M. (1983). Life-cycle stages, mobility, and metropolitan change: 1. Theoretical issues. *Environment and Planning* **15A**, pp. 293-306.

Weisbrod, G., Lerman, S., and Ben-Akiva, M. (1980). Tradeoffs in residential location decisions: transportation versus other factors. *Transport Policy and Decision Making* **1**, pp. 13-26.

West, Jim and Lowe, Allen (1997). Integration of transportation and land use planning through residential street design. *ITE Journal* **67** (8), pp. 48-51.

White, Michelle J. (1977). A model of residential choice and commuting by men and women workers. *Journal of Regional Science* **17** (1), pp. 41-52.

Young, William (1984). A non-tradeoff decision making model of residential location choice. *Transportation Research* **18A** (1), pp. 1-11.

Table 16: Sociodemographic Characteristics of the Raw Employed Sample[1] (N_3 = 615)

Variable	**5 Neighborhoods (N_3=615): NSF(N=121) SSF(N=119) CON(N=105) PH(N=140) SJ (N=130)**
Occupation[2]: number, percent cases (missing cases)	
Manager/administration	115, 19.7% (32)
Professional/technical	262, 44.9% (32)
Administrative support	90, 15.4% (32)
Retired	Not Applicable
Household composition: mean, standard deviation (missing cases)	
Household size	2.33, 1.14 (0)
No. people 16 or over	1.94, 0.76 (8)
No. people under 16	0.41, 0.77 (8)
No. full-time workers	1.22, 0.60 (0)
No. workers (part- and full-time)	1.46, 0.50 (0)
Personal characteristics: mean, standard deviation (missing cases)	
Age	44.6, 10.1 (178)
Education[3]	4.16, 1.28 (8)
Female (=1, Male = 0)	0.52, 0.50 (0)
Household income[3]	6.64, 1.25 (0)
Years lived in Bay Area	24.48, 15.53 (4)

[1] The values given in this table are based on raw data (i.e., no imputed means).
[2] Not all job categories are presented, and thus, percentages do not sum to 100%.
[3] Education and household income were entered as categorical data (e.g., 5 = annual income ranging from $20,001 to $35,000), but here averaged as if they were continuous. Respondents were on average well-educated (a value of 4 represents completion of 4-year degree) with moderate income levels (a value of 6 represents annual income varying from $35,001 to $50,000).

Table 17: Travel and Residential Characteristics of the Raw Employed Sample[1] (N_3 = 615)

Variable	**5 Neighborhoods (N_3=615): NSF(N=121) SSF(N=119) CON(N=105) PH(N=140) SJ (N=130)**
General travel information: mean, standard deviation (missing cases)	
No. of vehicles	1.95, 1.01 (0)
No. of vehicles / driver	1.05, 0.50 (0)
Commute distance (1-way, miles)	12.20, 12.24 (49)
Daily person trips	4.44, 2.34 (63)
Daily vehicle-miles traveled	30.38, 29.47 (65)
Daily transit-miles traveled	4.97, 14.01 (66)
Daily walk/bike-miles traveled	0.56, 2.83 (66)
Residential characteristics: mean, standard deviation (missing cases)	
Home size (square feet)	1496, 639 (47)
No. of bedrooms	2.71, 1.00 (6)
Home value category[2] (for the 435 homeowners)	4.39, 1.15 (0)
Monthly rent category[2] (for the 180 renters)	3.56, 1.05 (2)
Most important reasons for choosing current neighborhood: number, percent	
Housing cost	348, 56.6%
Close to shops and services	135, 21.9%
Close to work	175, 28.5%

Good school	65, 10.6%

[1] The values given in this table are based on raw data (i.e., no imputed means).
[2] Home value and monthly rent were collected as ordinal categorical variables. Reference points for each category include: 4 (home value ranging from $180,001 to $250,000),
6 (home value ranging from $375,001 to $575,000), 3 (monthly rent ranging from $501 to $700), and 4 (monthly rent ranging from $701 to $1,000).

Table 18: Sociodemographic Characteristics of the Raw Reduced Sample[1] (N_4 = 515)

Variable	**5 Neighborhoods (N_4=515): NSF (N=95) SSF (N=100) CON (N=87) PH (N=121) SJ (N=112)**
Occupation[2]: number, percent cases (missing cases)	
Manager/administration	93, 19.1% (27)
Professional/technical	222, 45.6% (27)
Administrative support	76, 15.6% (27)
Retired	Not Applicable
Household composition: mean, standard deviation (missing cases)	
Household size	2.31, 1.12 (0)
No. people 16 or over	1.91, 0.72 (8)
No. people under 16	0.42, 0.77 (8)
No. full-time workers	0.80, 0.40 (0)
No. workers (part- and full-time)	1.45, 0.50 (0)
Personal characteristics: mean, standard deviation (missing cases)	
Age	44.79, 11.16 (151)
Education[3]	4.16, 1.28 (8)
Female (=1, Male =0)	0.54, 0.50 (4)
Household income[3]	6.63, 1.20 (0)
Years lived in Bay Area	24.68, 15.61 (4)

[1] The values given in this table are based on raw data (i.e., no imputed means).
[2] Not all job categories are presented, and thus, percentages do not sum to 100%.
[3] Education and household income were entered as categorical data (e.g., 5 = annual income ranging from $20,001 to $35,000), but here averaged as if they were continuous. Respondents were on average well-educated (a value of 4 represents completion of 4-year degree) with moderate income levels (a value of 6 represents annual income varying from $35,001 to $50,000).

Table 19: Travel and Residential Characteristics of the Raw Reduced Sample[1] (N_4 = 515)

Variable	**5 Neighborhoods (N_4=515): NSF(N=95) SSF(N=100) CON(N=87) PH(N=121) SJ (N=112)**
General travel information: mean, standard deviation (missing cases)	
No. of vehicles	1.94, 0.93 (0)
No. of vehicles / driver	1.05, 0.44 (0)
Commute distance (1-way, miles)	10.81, 10.02 (40)
Daily person trips	4.29, 1.91 (56)
Daily vehicle-miles traveled	28.48, 26.09 (58)
Daily transit-miles traveled	4.20, 12.37 (57)
Daily walk/bike-miles traveled	0.18, 0.53 (57)
Residential characteristics: mean, standard deviation (missing cases)	
Home size (square feet)	1488, 602 (37)
No. of bedrooms	2.71, 1.00 (6)
Home value category[2] (for the 368 homeowners)	4.38, 1.11 (0)
Monthly rent category[2] (for the 147 renters)	3.61, 1.01 (0)

Most important reasons for choosing current neighborhood: number, percent	
Housing cost	288, 55.9%
Close to shops and services	112, 21.7%
Close to work	152, 29.5%
Good school	54, 10.5%

[1] The values given in this table are based on raw data (i.e., no imputed means).
[2] Home value and monthly rent were collected as ordinal categorical variables. Reference points for each category include: 4 (home value ranging from $180,001 to $250,000),
6 (home value ranging from $375,001 to $575,000), 3 (monthly rent ranging from $501 to $700), and 4 (monthly rent ranging from $701 to $1,000).

Table 20: Regression Imputation Model for Variable: Commute Distance

Variable Name and Type[1]	Coefficient	t-statistic (p-value)
Age (S)	-0.10	-2.01 (0.05)
Manager (S)	4.29	2.71 (0.01)
Professional (S)	3.88	3.05 (0.00)
Pro-transit (A)	1.33	2.40 (0.02)
Adventurer (L)	1.28	2.32 (0.02)
North San Francisco (D)	-8.98	-6.21 (0.00)
South San Francisco (D)	-7.44	-5.03 (0.00)
Constant	16.90	6.36 (0.00)
Number of observations	379	
R^2	0.18	
R^2-Adjusted	0.17	
F (significance)	12.00 (0.00)	

[1] **S = Sociodemographic, A = Attitude Factor, L = Lifestyle Factor, D = Dummy Variable**

Prediction Equation:

Commute Distance(i) =
(-0.10)*Age(i) + (4.29)*Manager(i) + (3.88)*Professional(i) +
(1.33)*Pro-transit(i) + (1.28)*Adventurer(i) +
(-8.98)*North_San_Francisco(i) + (-7.44)*South_San_Francisco(i) + 16.9

Table 21: Regression Imputation Model for Variable: Daily-Transit Miles

Variable Name and Type[1]	**Coefficient**	**t-statistic (p-value)**
Commute Distance (S)	0.40	8.77 (0.00)
Number of vehicles (S)	-1.99	-3.42 (0.00)
Pro-transit (A)	3.30	5.87 (0.00)
Hobbyist (L)	-1.18	-2.04 (0.04)
San Jose (D)	-5.64	-3.98 (0.00)
Constant	5.23	3.90 (0.00)
Number of observations	507	
R^2	0.24	
R^2-Adjusted	0.23	
F (significance)	31.70 (0.00)	

[1] **S = Sociodemographic, A = Attitude Factor, L = Lifestyle Factor, D = Dummy Variable**

Prediction Equation:

Transit-Mileage Rate(i) =
(0.40)*Commute_Distance(i) + (-1.99)*Number_of_vehicles(i) +
(3.30)*Pro-transit(i) + (-1.18)*Hobbyist(i) + (-5.64)*San_Jose(i) + 5.23

Table 21: Regression Imputation Model for Variable: Daily-Transit Miles

Variable Name and Type [1]	**Coefficient**	**t-statistic (p-value)**
Commute Distance (S)	0.40	8.77 (0.00)
Number of vehicles (S)	-1.99	-3.42 (0.00)
Pro-transit (A)	3.30	5.87 (0.00)
Hobbyist (L)	-1.18	-2.04 (0.04)
San Jose (D)	-5.64	-3.98 (0.00)
Constant	5.23	3.90 (0.00)
Number of observations	507	
R^2	0.24	
R^2-Adjusted	0.23	
F (significance)	31.70 (0.00)	

[1] **S = Sociodemographic, A = Attitude Factor, L = Lifestyle Factor, D = Dummy Variable**

Prediction Equation:

Transit-Mileage Rate(i) =
(0.40)*Commute_Distance(i) + (-1.99)*Number_of_vehicles(i) +
(3.30)*Pro-transit(i) + (-1.18)*Hobbyist(i) + (-5.64)*San_Jose(i) + 5.23

Table 22: Regression Imputation Model for Variable: Daily-Vehicle Miles

Variable Name and Type[1]	Coefficient	t-statistic (p-value)
Commute Distance (S)	0.94	10.43 (0.00)
Female (S)	-6.00	-2.49 (0.01)
Number of vehicles (S)	6.01	5.31 (0.00)
Pro-driving (A)	3.00	2.56 (0.01)
Pro-transit (A)	-2.43	-2.20 (0.03)
Adventurer (L)	3.07	2.75 (0.01)
Traveler (L)	2.34	2.13 (0.03)
Constant	9.66	3.08 (0.00)
Number of observations	507	
R^2	0.28	
R^2-Adjusted	0.27	
F (significance)	28.31 (0.00)	

[1] **S = Sociodemographic, A = Attitude Factor, L = Lifestyle Factor, D = Dummy Variable**

Prediction Equation:

Vehicle-Mileage Rate(i) =
(0.94)*Commute_Distance(i) + (-6.00)*Female(i) +
(6.01)*Number_of_vehicles(i) + (3.00)*Pro-driving(i) +
(-2.43)*Pro-transit(i) + (3.07)*Adventurer(i) + (2.34)*Traveler(i) + 9.66

Table 23: Strongest Pattern Matrix Loadings for Attitudinal Factor Scores[1]

Statement	Loading
Factor 1: Pro-High Density	
I need to have space between me and my neighbors	-0.75
I would only live in a multiple family unit, (apartment, condo, etc.) as a last resort	-0.69
It's important for children to have a large backyard for playing	-0.67
High-density residential development should be encouraged	0.55
Factor 2: Pro-Environment	
Environmental protection costs too much	-0.78
Environmentalism hurts minority and small businesses	-0.75
People and jobs are more important than the environment	-0.73
Environmental protection is good for California's economy	0.73
Stricter vehicle smog control laws should be introduced and enforced	0.47
Factor 3: Pro-Pricing	
I would be willing to pay a toll to drive on an uncongested road	0.76
We should raise the price of gasoline to reduce congestion and air pollution	0.38
Traffic congestion will take care of itself because people will make adjustments	-0.25
Factor 4: Pro-Alternatives	
Having shops and services within walking distance of my home would be important to me	0.50
Vehicle emissions increase the need for health care	0.49
I use public transportation when I cannot afford to drive	0.44
We should provide incentives to people who use electric or other clean-fuel vehicles	0.42
More lanes should be set aside for carpools and buses	0.39
Factor 5: Pro-Driving	
Driving allows me to get more done	0.74
Driving allows me freedom	0.71
I would rather use a clean-fuel car than give up driving	0.66

[1] Some lower and secondary factor loadings are presented when they help improve interpretation of the factors.

Table 23: Strongest Pattern Matrix Loadings for Attitudinal Factor Scores[1] - (Continued)

Statement	Loading
Factor 6: Pro-Drive Alone	
I like someone else to do the driving	-0.70
I am not comfortable riding with strangers	0.62
Ridesharing saves money	-0.49
Factor 7: Pro-Growth	
We need to build more roads to help decrease congestion	0.59
Too many people drive alone	-0.48
Too much agricultural land is consumed for housing	-0.44
Getting stuck in traffic doesn't bother me too much	0.39
Factor 8: Work-Driven	
I like to spend most of my time working	0.73
When things are busy at work, I get more done by cutting back on personal time	0.71
Factor 9: Time-Satisfied	
I would like to have more time for leisure	-0.74
I feel that I am wasting my time when I have to wait	-0.69
Getting stuck in traffic doesn't bother me too much	0.48
Factor 10: Pro-Transit	
Public transportation is unreliable	-0.70
It costs more to use public transportation than it does to drive a car	-0.59
Buses and trains are pleasant to travel in	0.57
I can read and do other things when I use public transportation	0.43

[1] Some lower and secondary factor loadings are presented when they help improve interpretation of the factors.

Table 24: Strongest Pattern Matrix Loadings for Lifestyle Factor Scores[1]

Activity Description[2]	Loading
Factor 1: Culture-Lover	
Attended a concert or symphony	0.49
Attended the ballet	0.46
Read material on art or architecture	0.44
Attended the theater	0.39
Factor 2: Altruist	
Read material on religion	0.61
Spent last weekend participating in religious activities	0.58
Volunteered to help the community	0.53
Spent last weekend doing volunteer work	0.52
Participated in community events	0.43
Factor 3: Nest-Builder	
Read material on home improvement	0.65
Read material on gardening	0.57
Made house improvements myself	0.57
Put in a flower or vegetable garden	0.55
Spent last weekend doing yard work	0.53
Factor 4: Relaxer	
Spent last weekend reading	0.56
Spent last weekend at home relaxing	0.55
Spent last weekend shopping	0.48
Spent last weekend doing chores	0.47
Factor 5: Traveler	
Traveled to another country	0.47
Took a cruise	0.43
Visited another state	0.37
Visited a wildlife refuge	0.35

Table 24: Strongest Pattern Matrix Loadings for Lifestyle Factor Scores[1] -(Continued)

Activity Description[2]	**Loading**
Factor 6: Adventurer	
Went hunting	0.53
Used an off-road vehicle	0.51
Went to a shooting range	0.47
Participated in a motor cross	0.41
Factor 7: Fun-Seeker	
Went to a zoo	0.54
Read children's stories	0.51
Visited an aquarium	0.45
Visited an amusement park	0.36
Factor 8: Homebody	
Read materials on women's issues	0.65
Read material on fashion	0.60
Sewed (made clothes, quilts, etc.)	0.58
Read material on cooking or recipes	0.56
Did needlework or embroidery	0.54
Read material on decorating	0.50
Factor 9: Outdoor Enthusiast	
Visited a national park or historic site	0.64
Visited a state park or historic site	0.60
Visited a local park or historic site	0.56
Went hiking or backpacking or camping	0.51
Visited a beach	0.49

Table 24: Strongest Pattern Matrix Loadings for Lifestyle Factor Scores[1] -(Continued)

Activity Description[2]	**Loading**
Factor 10: Athlete	
Participated in a sports event	0.64
Played tennis or golf	0.59
Attended a professional sports event	0.57
Read material on sports or exercise or health	0.55
Spent last weekend outdoors participating in sports	0.48
Factor 11: Hobbyist	
Read material on science or nature	0.56
Read material on the environment	0.50
Read material on the outdoors	0.45
Read material on history	0.44
Read material on photography	0.41
Read material on humor	0.39
Read material on pets	0.32
Spent last weekend doing hobbies	0.29

[1] Some lower and secondary factor loadings are presented when they help improve interpretation of the factors.

[2] The time frame for these activities is as follows: "Read material on..." within the past month; all other activities occurred within the past 12 months except where noted to have taken place the past weekend.

Table 25: Relative Effects of Sociodemographic Characteristics and Attitude/Lifestyle Factors in a Binary Logit Residential Choice Model (Dependent Variable: Suburb = 1 and Traditional Neighborhood = 0)

Variable Name and Type **S = Sociodemographic** **A = Attitude Factor** **L = Lifestyle Factor**	**Full Model**		**Attitude and Lifestyle Factors Excluded**		**Sociodemographic Characteristics Excluded**	
	b	**t (p-val.)**	b	**t (p-val.)**	b	**t (p-val.)**
Constant	-1.36	-3.49 (0.00)	-2.52	-7.66 (0.00)	1.02	7.24 (0.00)
Number of people under 16 (S)	0.83	3.50 (0.00)	0.91	4.77 (0.00)		
Number of vehicles available (S)	0.77	4.43 (0.00)	1.01	6.83 (0.00)		
Years lived in Bay Area (S)	0.03	3.38 (0.00)	0.05	6.59 (0.00)		
Pro-pricing (A)	-0.58	-3.34 (0.00)			-0.63	-4.10 (0.00)
Pro-environment (A)	-0.29	-1.87 (0.06)			-0.35	-2.40 (0.02)
Pro-high density (A)	-0.84	-4.86 (0.00)			-0.94	-6.00 (0.00)
Pro-alternatives (A)	-0.63	-3.70 (0.00)			-0.60	-3.85 (0.00)
Altruist (L)	0.32	2.23 (0.03)			0.28	2.17 (0.03)
Culture-lover (L)	-0.76	-4.91 (0.00)			-0.86	-5.95 (0.00)
Nest-builder (L)	0.57	3.83 (0.00)			0.69	5.14 (0.00)
Number of observations	492		492		492	
Initial log-likelihood	-341.03		-341.03		-341.03	
Log-likelihood at convergence	-151.95		-219.04		-181.46	
ρ^2	0.55		0.36		0.47	
Adjusted ρ^2	0.52		0.35		0.44	
χ^2	378.16		243.98		319.16	

Table 26: Contributions of Attitude/Lifestyle Factors and Sociodemographic Variables to the Explanatory Power of the Residential Choice Model

Model Description[1]	r^2	**% of Total Information Explained by Block**[2]	**% of Full Model Information Explained by Block**[3]
Market share model (1 variable)	0.101	10.1	18.2
Model with constant and 3 sociodemographic variables only	0.358	**8.6** to *25.7*	**15.5** to *46.4*
Model with constant and 7 attitudinal and lifestyle variables only	0.468	**19.6** to *36.7*	**35.4** to *66.2*
Full model (11 variables)	0.554	55.4	100.0

[1] ρ^2 rather than adjusted ρ^2 should be used for an analysis of this type since it is ρ^2 that has the interpretation of percent information in the data explained by the model, and allows for the decomposition of the "perfect information" scenario into incremental contributions of blocks of variables (see Hauser, 1978).
[2] Bold numbers are obtained by: ρ^2(Full model) - ρ^2(Model w/sociodemographic only or attitudinal and lifestyle variables only) x 100%. Italic numbers are obtained by: ρ^2(Model w/sociodemographic only or attitudinal and lifestyle variables only) - ρ^2(market share model) x 100%. For example, the percent contribution of sociodemographic variables ranges from [0.554 - 0.468] * 100 = 0.086 * 100 = **8.6** to [0.358 - 0.101] * 100 = 0.257 * 100 = *25.7*.
[3] Values in this column are equal to [% of total information contributed by block /Full model ρ^2].

Table 53: Direct Effects (t-stat.) of Nontransformed Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Constant	1.512 (8.06)	-0.733 (-3.12)	0.145 (3.03)	-1.041 (-5.47)	-0.010 (-0.26)	28.668 (6.96)	5.241 (2.22)	0.608 (5.60)	1.798 (1.02)
Sociodemographic									
Age	-0.015 (-4.63)	0.008 (2.44)							
Female				0.550 (5.64)		-9.322 (-3.91)			
Household Size	-0.128 (-3.63)								
Number of People Under Age 16			-0.225 (-4.69)						
Number of Vehicles	-0.247 (-6.47)					5.121 (4.92)	-2.19 (-4.95)		
Years Lived in Bay Area				0.007 (3.36)					
Lifestyle									
Adventurer						4.16 (4.30)			
Culture-Lover		-0.158 (-4.41)	0.251 (6.83)						
Hobbyist		-0.100 (-2.96)							
Homebody		0.175 (4.81)							
Nest-Builder	-0.141 (-3.79)		-0.182 (-4.90)						

Outdoor Enthusiast			0.091 (2.57)						

Table 53: Direct Effects (t-stat.) of Nontransformed Res. Choice SEM - continued

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Attitudes									
Pro-Alternatives		-0.105 (-2.73)						0.275 (2.59)	
Pro-Drive Alone					0.161 (4.16)				
Pro-Environment					0.167 (4.25)				
Pro-Growth							-1.082 (-2.86)		
Pro-Pricing	0.109 (3.42)	-0.123 (-3.58)							
Time-Satisfied				-0.119 (-3.28)					
Work-Driven			0.119 (3.63)						
Pro-High[1] Density	0.498 (5.21)								
Pro-Driving[1]			-0.643 (-4.40)					-0.566 (-1.60)	
Pro-Transit[1]							3.192 (2.25)		
Residential Choice/Preference									
Suburb[1]									2.692 (2.89)
Travel Demand									
Vehicle Miles[1]				0.020 (4.06)					0.223 (5.33)

Transit Miles[1]									0.712 (4.89)
Walk/Bike[1] Miles						-7.441 (-2.02)			
Job Location									
Commute[1] Distance		0.030 (1.97)					0.341 (1.94)		

[1] Endogenous variable.

Table 54: Total Effects of Nontransformed Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Constant	0.070	-0.007	0.062	0.042	0.001	30.134	5.055	0.561	12.084
Sociodemographic									
Age	-0.015	0.009					0.011		0.033
Female	-0.128	-0.076	-0.257	0.400		-7.637	-0.857	-0.226	-2.514
Household Size	-0.128								
Number of People Under Age 16	-0.112		-0.225						
Number of Vehicles	-0.282	-0.014	-0.070	0.110		5.583	-2.348	-0.062	-0.468
Years Lived in Bay Area	-0.003		-0.005	0.008		0.034	0.004	-0.005	0.011
Lifestyle									
Adventurer	-0.028	0.045	-0.057	0.089		4.536	0.509	-0.050	1.493

Culture-Lover	0.125	-0.177	0.251				-0.215		-0.630
Hobbyist		-0.112					-0.136		-0.399
Homebody		0.196					0.238		0.698
Nest-Builder	-0.232		-0.182						
Outdoor Enthusiast	0.045		0.091						

Table 54: Total Effects of Nontransformed Residential Choice SEM - continued

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Attitudes									
Pro-Alternatives	0.014	-0.140	0.028	-0.044		-2.229	-0.393	0.300	-1.151
Pro-Drive Alone		0.016			0.161		0.698		0.541
Pro-Environment		0.017			0.167		0.724		0.561
Pro-Growth		-0.034					-1.471		-1.140
Pro-Pricing	0.109	-0.138					-0.167		-0.490
Time-Satisfied	0.042	-0.005	0.084	-0.130		-0.547	-0.061	0.073	-0.180
Work-Driven	0.059		0.119						
Pro-High[1] Density	0.498								

Pro-Driving[1]	-0.349	0.046	-0.701	0.090		4.589	0.515	-0.617	1.511
Pro-Transit[1]		0.101					4.339		3.362
Residential Choice/Preference									
Suburb[1]		0.120					1.358		3.982
Travel Demand									
Vehicle Miles[1]	-0.007	0.011	-0.014	0.021		0.090	0.122	-0.012	0.359
Transit Miles[1]		0.032					0.359		1.053
Walk/Bike[1] Miles	0.051	-0.080	0.102	-0.159		-8.111	-0.911	0.090	-2.670
Job Location									
Commute[1] Distance		0.045					0.505		0.479

[1] Endogenous variable.

Table 55: Direct Effects (t-stat.) of Transformed[1] Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles[1]	Transit Miles[1]	Walk/ Bike Miles[1]	Commute Distance
Constant	1.822 (8.3)	-1.034 (-3.8)	0.204 (3.9)	-1.407 (-4.7)	-0.046 (-1.1)	3.838 (4.3)	0.823 (3.1)	0.160 (11.6)	-3.313 (-1.2)
Sociodemographic									
Age	-0.018 (-5.06)	0.009 (2.61)							
Female				0.543 (6.03)		-0.812 (3.63)			
Household	-0.081 (-1.92)								

Size									
Number of People Under Age 16[1]			-0.328 (-4.86)						
Number of Vehicles[1]	-0.761 (-5.44)					1.652 (5.17)	-0.870 (-6.30)		
Years Lived in Bay Area				0.004 (2.21)					
Lifestyle									
Adventurer[1]						0.554 (2.40)			
Culture-Lover		-0.178 (-4.48)	0.230 (5.80)						
Hobbyist		-0.099 (-2.75)							
Homebody		0.217 (5.50)							
Nest-Builder	-0.143 (-3.53)		-0.140 (-3.61)						
Outdoor Enthusiast			0.106 (2.87)						
Relaxer			0.042 (1.63)						

[1] These variables were transformed to reduce their univariate kurtosis values – all variables had a natural log transformation except number of vehicles, which was a square-root transformation.

Table 55: Direct Effects (t-stat.) of Transformed[1] Res. Choice SEM - continued

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles[1]	Transit Miles[1]	Walk/ Bike Miles[1]	Commute Distance
Attitudes									
Pro-Alternatives		-0.070 (-1.57)						0.026 (2.13)	

Pro-Drive Alone					0.178 (4.21)				
Pro-Environment					0.161 (3.73)				
Pro-Growth							-0.058 (-2.07)		
Pro-Pricing	0.108 (3.11)	-0.074 (-2.23)							
Time-Satisfied				-0.109 (-3.19)					
Work-Driven			0.092 (2.55)						
Pro-High[2] Density	0.591 (5.00)								
Pro-Driving[2]			-0.702 (-4.41)					-0.082 (-2.00)	
Pro-Transit[2]							0.286 (2.56)		
Residential Choice/Preference									
Suburb[2]									2.149 (2.84)
Travel Demand									
Vehicle Miles[2]				0.226 (4.07)					2.213 (4.89)
Transit Miles[2]									5.732 (4.96)
Walk/Bike[2] Miles						-7.201 (-2.04)			
Job Location									
Commute[2] Distance		0.059 (2.62)					0.065 (3.24)		

[1] These variables were transformed to reduce their univariate kurtosis values – all variables had a natural log transformation except number of vehicles, which was a square-root transformation.

[2] Endogenous variable.

Table 56: Total Effects of Transformed[1] Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory	Trad-	Sub-	Pro-	Pro-	Pro-	Vehicle	Transit	Walk/	Commute

Variable ®	itional	urb	HD	Driving	Transit	Miles[1]	Miles[1]	Bike Miles[1]	Distance
Constant	0.052	0.015	0.053	0.101	-0.037	4.802	0.616	0.149	10.88
Sociodemographic									
Age	-0.018	0.011					0.002		0.038
Female	-0.172	-0.148	-0.292	0.416		-0.563	-0.161	-0.035	-2.489
Household Size	-0.081								
Number of People Under Age 16[1]	-0.194		-0.328						
Number of Vehicles[1]	-0.940	-0.091	-0.303	0.432		1.910	-0.969	-0.036	-1.523
Years Lived in Bay Area	-0.002		-0.003	0.005		0.003	0.001	-0.005	0.013
Lifestyle									
Adventurer[1]	-0.060	0.045	-0.102	0.145		0.641	0.184	-0.012	2.832
Culture-Lover	0.136	-0.224	0.230				-0.050		-0.765
Hobbyist		-0.124					-0.028		-0.425
Homebody		0.272					0.060		0.930
Nest-Builder	-0.226		-0.140						
Outdoor Enthusiast	0.063		0.106						
Relaxer	-0.020	0.008	-0.034	0.048		0.029	0.008	-0.004	0.127

[1] These variables were transformed to reduce their univariate kurtosis values – all variables had a natural log transformation except number of vehicles, which was a square-root transformation.

Table 56: Total Effects of Transformed[1] Res. Choice SEM - continued

Endogenous	Residential		Travel	Job

Variable ®	Choice/ Preference		Attitudes			Demand			Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles[1]	Transit Miles[1]	Walk/ Bike Miles[1]	Commute Distance
Attitudes									
Pro-Alternatives	0.021	-0.146	0.035	-0.050		-0.220	-0.082	0.031	-1.272
Pro-Drive Alone		0.035			0.178		0.088		0.581
Pro-Environment		0.031			0.161		0.080		0.528
Pro-Growth		-0.040					-0.101		-0.664
Pro-Pricing	0.108	-0.093					-0.021		-0.318
Time-Satisfied	0.052	-0.020	0.089	-0.126		-0.076	-0.022	0.010	-0.334
Work-Driven	0.054		0.092						
Pro-High[2] Density	0.591								
Pro-Driving[2]	-0.479	0.182	-0.811	0.156		0.692	0.198	-0.096	3.056
Pro-Transit[2]		0.195					0.498		3.271
Residential Choice/Preference									
Suburb[2]		0.255					0.278		4.293
Travel Demand									
Vehicle Miles[2]	-0.108	0.304	-0.183	0.262		0.156	0.331	-0.022	5.111
Transit Miles[2]		0.681					0.742		11.450
Walk/Bike[2] Miles	0.780	-2.189	1.321	-1.883		-8.327	-2.385	0.156	-36.804
Job Location									
Commute[2] Distance		0.119					0.129		0.997

[1] These variables were transformed to reduce their univariate kurtosis values – all variables had a natural log transformation except number of vehicles, which was a square-root transformation.

[2] Endogenous variable.

Table 59: Direct Effects (t-stat.) of Standardized N.T. Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Sociodemographic									
Age	-0.153 (-4.63)	0.084 (2.44)							
Female				0.275 (5.64)		-0.165 (-3.90)			
Household Size	-0.146 (-3.64)								
Number of People Under Age 16			-0.170 (-4.69)						
Number of Vehicles	-0.249 (-6.46)					0.183 (4.93)	-0.166 (-4.95)		
Years Lived in Bay Area				0.114 (3.36)					
Lifestyle									
Adventurer						0.147 (4.31)			
Culture-Lover		-0.158 (-4.41)	0.251 (6.83)						
Hobbyist		-0.100 (-2.96)							

Homebody		0.175 (4.81)							
Nest-Builder	-0.141 (-3.78)		-0.182 (-4.90)						
Outdoor Enthusiast			0.091 (2.57)						

Table 59: Direct Effects (t-stat.) of Standardized N.T. Res. Choice SEM - continued

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Attitudes									
Pro-Alternatives		-0.105 (-2.73)						0.103 (2.59)	
Pro-Drive Alone					0.161 (4.16)				
Pro-Environment					0.167 (4.25)				
Pro-Growth							-0.081 (-2.86)		
Pro-Pricing	0.109 (3.42)	-0.123 (-3.58)							
Time-Satisfied				-0.119 (-3.28)					
Work-Driven			0.119 (3.63)						
Pro-High[1] Density	0.498 (5.22)								
Pro-Driving[1]			-0.643 (-4.40)					-0.211 (-1.59)	
Pro-Transit[1]							0.239 (2.25)		
Residential Choice/Preference									

Suburb[1]									0.228 (2.89)
Travel Demand									
Vehicle Miles[1]				0.554 (4.06)					0.532 (5.32)
Transit Miles[1]									0.805 (4.89)
Walk/Bike[1] Miles						-0.704 (-2.02)			
Job Location									
Commute[1] Distance		0.356 (1.97)					0.301 (1.94)		

Table 60: Total Effects of Standardized N.T. Residential Choice SEM

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Sociodemographic									
Age	-0.153	0.094					0.008		0.028
Female	-0.064	-0.038	-0.129	0.200		-0.135	-0.032	-0.042	-0.106
Household Size	-0.146								
Number of People Under Age 16	-0.085		-0.170						
Number of Vehicles	-0.285	-0.014	-0.071	0.111		0.200	-0.178	-0.023	-0.040
Years Lived in Bay Area	-0.040	0.005	-0.080	0.124		0.018	0.004	-0.026	0.015

Lifestyle									
Adventurer	-0.028	0.045	-0.057	0.089		0.161	0.038	-0.019	0.126
Culture-Lover	0.125	-0.177	0.251				-0.016		-0.053
Hobbyist		-0.112					-0.010		-0.034
Homebody		0.196					0.018		0.059
Nest-Builder	-0.232		-0.182						
Outdoor Enthusiast	0.045		0.091						

Table 60: Total Effects of Standardized N.T. Residential Choice SEM - continued

Endogenous Variable ®	Residential Choice/ Preference		Attitudes			Travel Demand			Job Location
Explanatory Variable ®	Trad-itional	Sub-urb	Pro-HD	Pro-Driving	Pro-Transit	Vehicle Miles	Transit Miles	Walk/ Bike Miles	Commute Distance
Attitudes									
Pro-Alternatives	0.014	-0.140	0.028	-0.044		-0.079	-0.029	0.112	-0.097
Pro-Drive Alone		0.016			0.161		0.052		0.046
Pro-Environment		0.017			0.167		0.054		0.048
Pro-Growth		-0.034					-0.110		-0.097
Pro-Pricing	0.109	-0.138					-0.012		-0.041
Time-	0.042	-0.005	0.084	-0.130		-0.019	-0.005	0.027	-0.015

Satisfied									
Work-Driven	0.059		0.119						
Pro-High[1] Density	0.498								
Pro-Driving[1]	-0.349	0.046	-0.701	0.090		0.162	0.039	-0.231	0.128
Pro-Transit[1]		0.102					0.325		0.285
Residential Choice/Preference									
Suburb[1]		0.120					0.102		0.337
Travel Demand									
Vehicle Miles[1]	-0.193	0.306	-0.389	0.604		0.090	0.259	-0.128	0.858
Transit Miles[1]		0.424					0.359		1.191
Walk/Bike[1] Miles	0.136	-0.215	0.274	-0.426		-0.768	-0.182	0.090	-0.604
Job Location									
Commute[1] Distance		0.527					0.446		0.479

[1] Endogenous variable.

Figure 7: Standardized Means (by Neighborhood) of the Nine Highest-Loading Traits on the Aggregate Traditionalness Factor

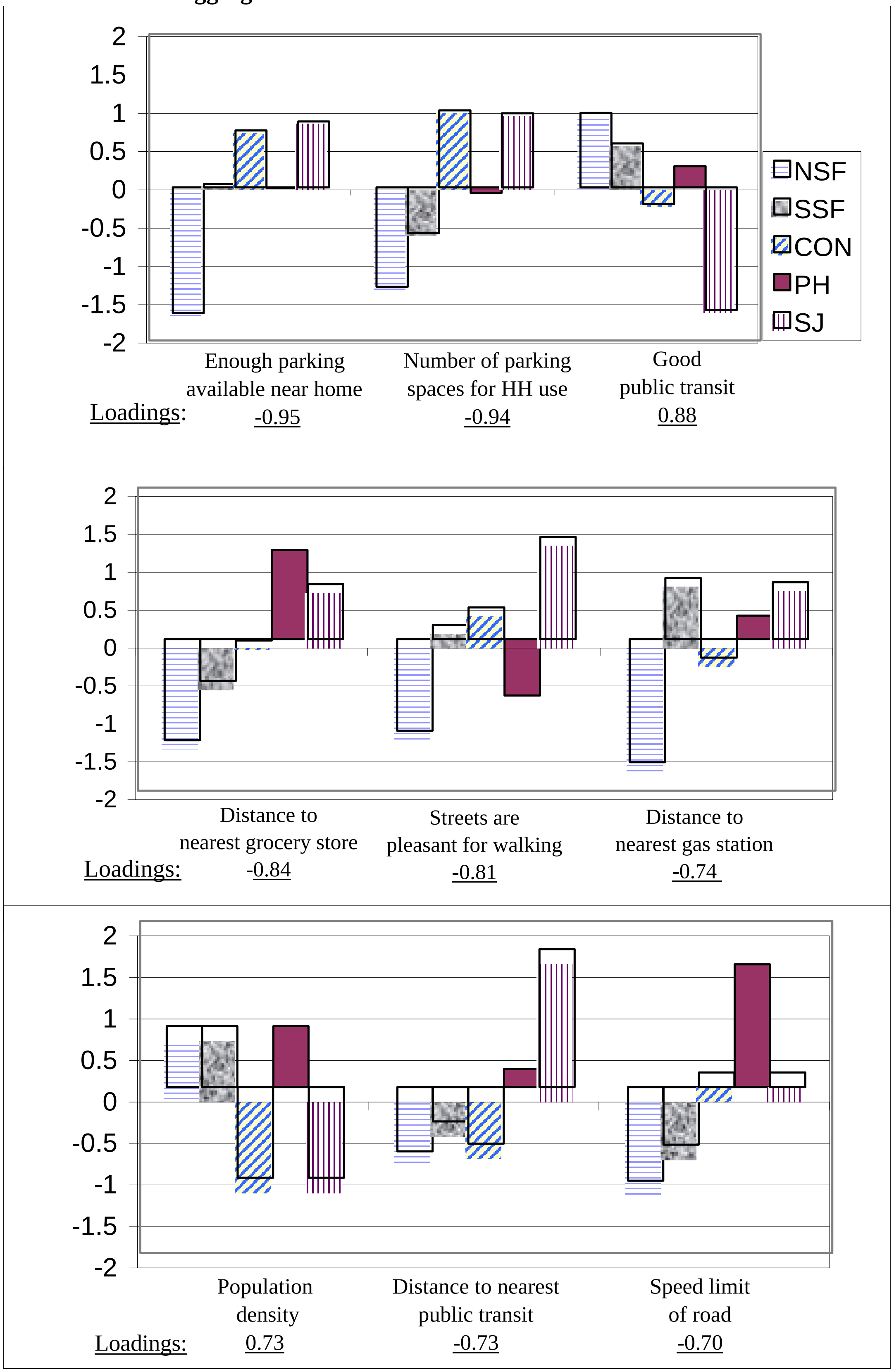

Figure 8: Standardized Means (by Neighborhood) of the Nine Highest-Loading Traits on the Single Disaggregate Traditional Factor

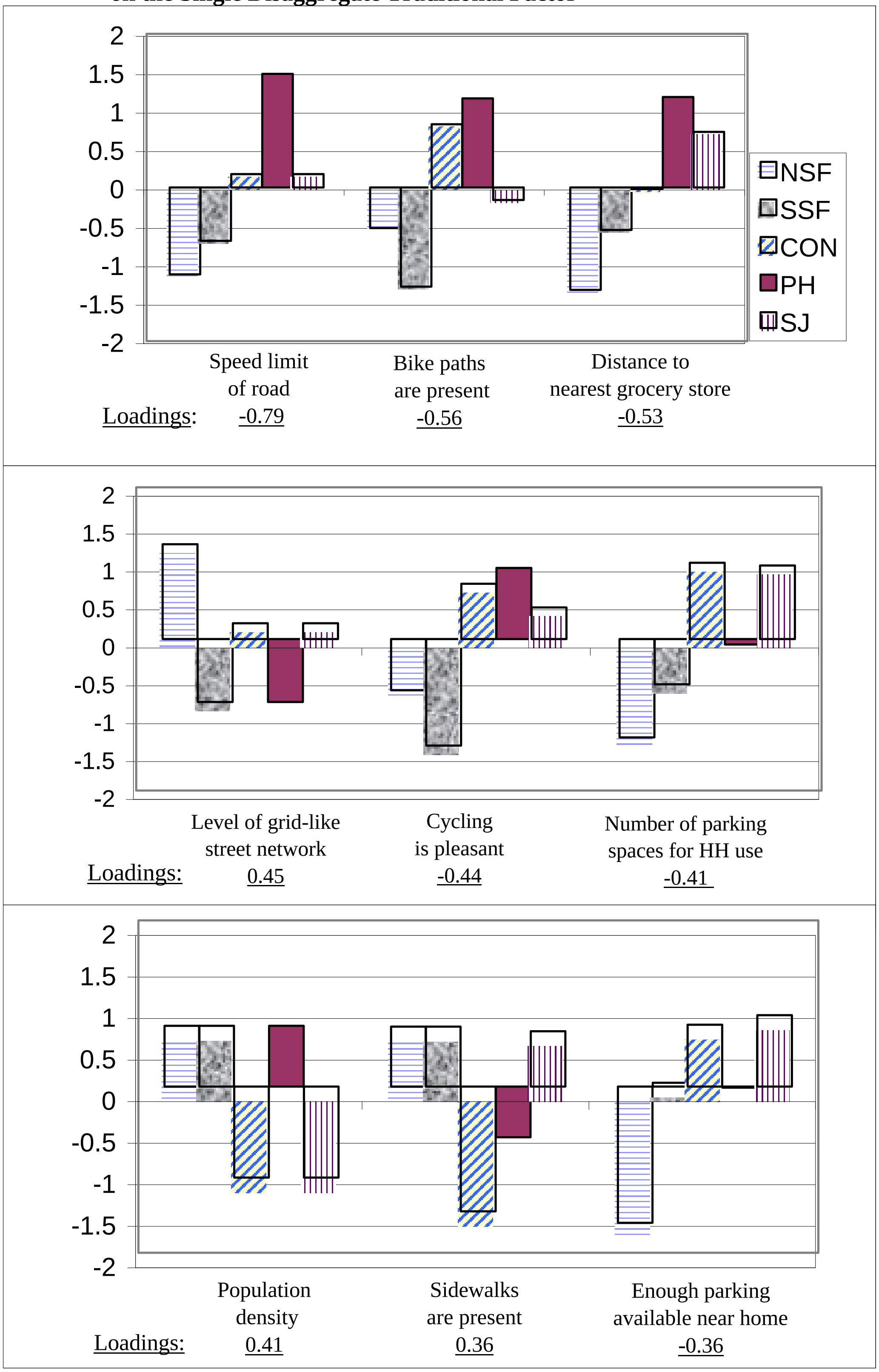

Figure 9: Standardized Means (by Neighborhood) of the Nine Highest-Loading Traits on the Suburban Factor of the Two-Factor Solution

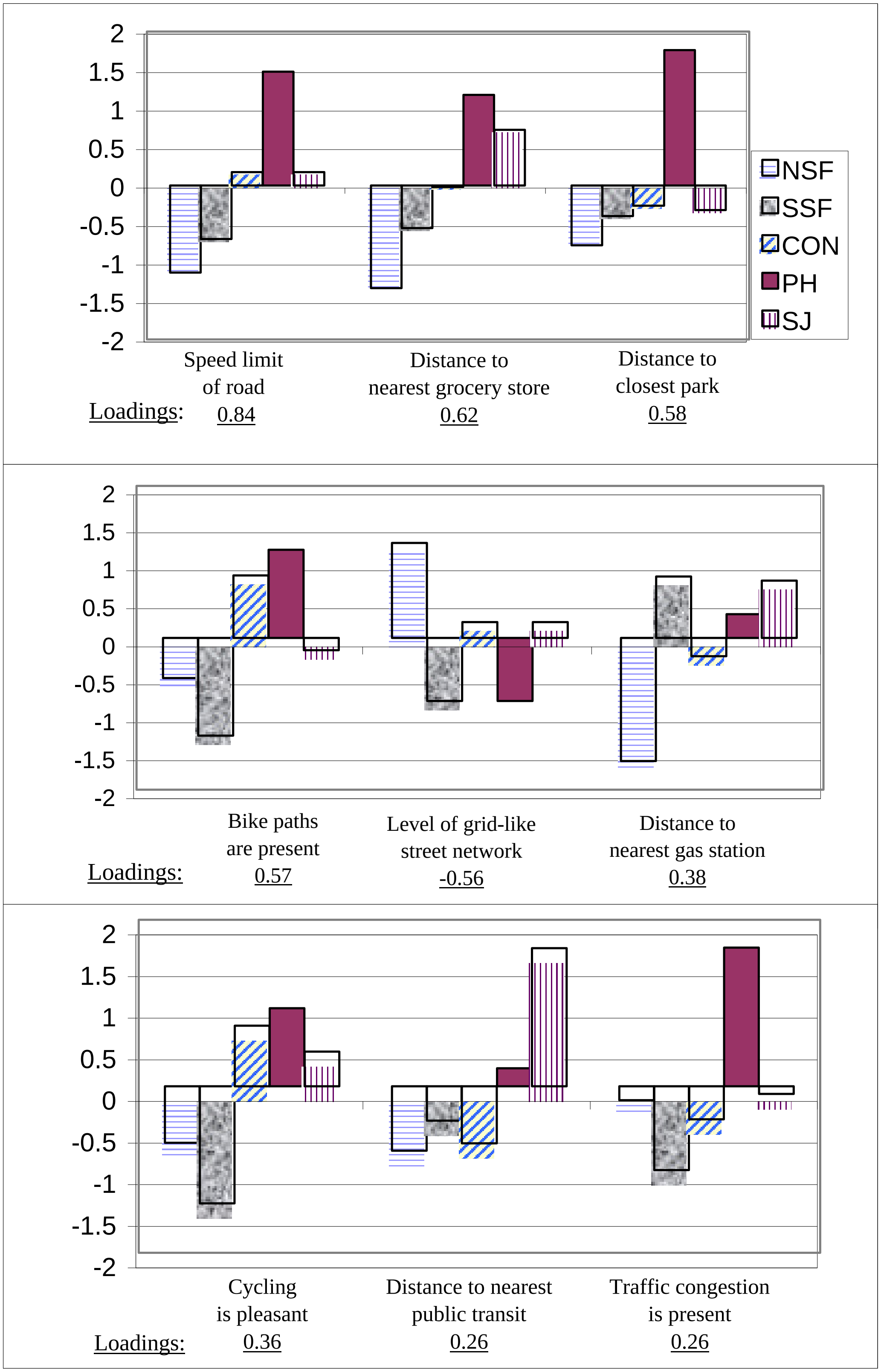

Figure 10: Standardized Means (by Neighborhood) of the Nine Highest-Loading Trai on the Traditional Factor of the Two-Factor Solution