Title
Cars and the City: An Investigation of Transportation and Residential Location Choices in New York City

Author
Salon, Deborah

Publication Date
2006

Cars and the City: An Investigation of Transportation and Residential Location Choices in New York City

By

DEBORAH SALON
B.A. (Carleton College) 1994

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

AGRICULTURAL AND RESOURCE ECONOMICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA,

DAVIS

Approved:

Committee in Charge

2006

Contents

List of Figures

List of Tables

Abstract

This dissertation is an exploration of the relationship between the transportation-land use system in New York City and the transportation and residential location choices made by New Yorkers. The focus is on understanding these location and travel choices made by urbanites. Specifically, this research uses discrete choice models to identify and quantify the effects of the variables that factor into New Yorkers' decisions about where to live, whether to own a car, and how to get around in their daily lives. These models, along with GIS technology, are used to answer the following questions:

1. How far off are the results of models that do not take all three of these decisions as endogenous?

2. In a densely populated urban environment, what are the policy-sensitive factors that determine whether households own cars and how often walking is the mode of choice?

3. How does the relative importance of these factors change across different neighborhoods within the city?

4. How much of the relationship between land use patterns and travel behavior is due to the indirect effects of neighborhood and car ownership choice, and how large is the direct effect of land use patterns on travel behavior?

Results indicate that the choices of residential location and commute mode are closely related; models of only commute mode choice produce biased results. Models that do not take the choice of car ownership as endogenous in New York do not appear to be severely biased.

Full models of the three choices indicate that the most important policy-sensitive factors influencing car ownership and mode choice are commute cost, commute time, and population density. A set of spatial scenario analyses illustrate that the importance of these factors does indeed vary across neighborhoods within the city.

Finally, a methodology is developed to separate the direct effect of land use patterns on travel behavior from the indirect effects. The example used here identifies the direct and indirect effects of population density on the propensity to walk, finding that approximately half of the total effect is direct.

Acknowledgements

First and foremost, I would like to thank my dissertation committee members Jim Wilen, Susan Handy, and Dan Sperling for their support, questions, and advice throughout this dissertation project. Dan has been the best supporter I could want throughout my graduate school career – he brought me to Davis, helped to ensure I always had funding for my work, and asked the "big questions" that every researcher needs to hear periodically (whether she wants to or not!). Susan became part of my committee after the project design phase, and has been a tremendous help in working with me and the devil (in the details). Susan's input helped me stay on-track in my attempt to make my research relevant to both economists and urban planners. Jim worked with me intensively through the project design phase, continually encouraging me to think more "like an economist" and to ask clear questions in my research. Jim's persistence on these points (although admittedly frustrating at the time) taught me how to be a more effective researcher, improving the quality of not only this project but hopefully also of all my future research projects. Although she is not officially on my committee, I'd like to acknowledge Pat Mokhtarian who was kind enough to work with me on the analysis presented in Chapter 5, providing advice.

I would also like to thank a number of the participants in the Pan-American Studies Institute on Transportation that took place in July and August of 2005. In particular, Francisco (Pancho) Martinez and Marcela Munizaga, both faculty at the Universidad Catolica in Santiago, Chile, were extremely generous with their time. It is not an exaggeration to say that the analysis in this research project could have been seriously flawed without their input.

In addition to the faculty who supported my work, there are a number of my graduate student peers who have materially contributed to this project. Julie Witcover, Nathan Parker, Steve Newbold, Jim Barrett, Yuko Onozaka, and Gorm Kipperberg have all been extraordinarily helpful sounding boards for many of the ideas and methodologies that are incorporated here. Many others have provided much-needed moral support throughout the years that this project has spanned. These include my parents, Joel and Carolyn Salon, my sister, Heather Salon, my grandfather, Leo Salon, and my local support system of friends who listened to me both when things were going well and through my frustrations.

I would also like to acknowledge Alan Gershowitz at the New York Metropolitan Tranportation Council, who helped me tremendously by convincing New York City Transit to grant me access to a full digital map of the public transportation system for the New York metropolitan area.

And, last but definitely not least, this work was financially supported in part by a dissertation grant from the University of California Transportation Center, and in part by a dissertation grant from the Robert Wood Johnson Foundation's Active Living Research program.

Of course, any and all errors are my own.

Chapter 1

Introduction: Urban Transportation Choices, Land Use Patterns, and The Environment

> *Most Americans, including most New Yorkers, think of New York City as an ecological nightmare, a wasteland of concrete and garbage and diesel fumes and traffic jams, but in comparison with the rest of America it's a model of environmental responsibility. By the most significant measures, New York is the greenest community in the United States, and one of the greenest cities in the world. (Owen, 2004)*

This dissertation is an exploration of the relationship between the transportation-land use system in New York City and the transportation and residential location choices made by New Yorkers. The focus is on understanding these location and travel choices made by urbanites. Specifically, this research attempts to identify and quantify the effects of the variables that factor into New Yorkers' decisions about where to live, whether to own a car, and how to get around in their daily lives.

The motivation for this line of inquiry is a joint concern for the natural environment, for public health, and for urban communities. Natural ecosystems are negatively impacted as suburban sprawl moves outward through direct habitat reduction for native plant and animal species. Use of motorized transport produces emissions that lower urban air quality, directly affecting city-dwellers through increased incidence of asthma and other respiratory problems. Use of motorized transport also results in emissions of greenhouse gases that contribute to global climate change.

The public health community has identified obesity as a growing (no pun intended) public health concern in most of the developed world. One of the possible contributing factors is a reduction in physical activity levels. This observation has renewed interest in fostering non-motorized modes of transport in the United States as a way to combat the obesity epidemic.

Interaction between its members is vital for the continued health of any community. In urban neighborhoods, much of this interaction naturally occurs on the street. Many observers blame the private car for negatively impacting urban communities through reducing the need to physically be on the street in a way that fosters community-building interaction.

Some researchers have maintained that as societies become wealthier, decreased population density and increased use of private cars for transport are inevitable. Others point out that while this is certainly the prevailing trend, there is high variation in all of these metrics among the wealthy urban areas of the world. It is important to aim for a better understanding of what makes some wealthy cities able to prosper while keeping their negative impact on the natural environment in check

The current research focuses on New York because it is a good example of a relatively wealthy city in which many residents choose a car-free lifestyle, walk for many of their trips, and many neighborhoods are vibrant, healthy communities. If we understand what is behind these lifestyle choices in New York, perhaps we can help city planners in other areas to create similar choice environments.

In this dissertation, data from New York City form the basis for a statistical model of New Yorkers' choices of residential neighborhood, car ownership status, and transport mode for both work and non-work trips. The main contributions to the literature from this dissertation are in answering the following questions:

1. How related are these three decisions, and how far off are the results of models that do not take all of them as endogenous?

2. In a densely populated urban environment, what are the policy-sensitive factors that determine whether households own cars and how often walking is the mode

of choice?

3. How does the relative importance of these factors change across different neighborhoods within the city?

4. How much of the relationship between land use patterns and travel behavior is due to the indirect effects of neighborhood and car ownership choice, and how large is the direct effect of land use patterns on travel behavior?

1.1 The urban transportation-land use connection

One of the unique aspects of this dissertation work is that it jointly models both the choice of location and the transportation choices of car ownership and transport mode. This joint modeling approach is theoretically appealing due to the fact that urban transportation systems and land use are inextricably linked.

Location is the link between land use and transportation. Land use is defined as the pattern of built structures and activities that land is used for, and every piece of land is endowed with a location. Through its impact on accessibility, the transport system helps to define the desirability of locations for any sort of economic or social activity. When households choose their residential location, then, they are also choosing the level of access they will enjoy from their home location to every other location they might want to travel to. "Level of access" here means the time and money cost of getting from home to various destinations by all available modes of transport. One of the main determinants of residential location choice is the level of access that the location provides to the destinations to which the household wants the best access.

The household influences this level of access through its choice of car ownership status. Owning a car greatly increases access to locations not served or poorly served by public transportation. That level of access comes at a set of fixed costs that may not be acceptable if access is available through other means. When residential location decisions are made, a change in car ownership status may provide for the best possible living arrangement for the household. Getting rid of a car may free up resources to live near work or buying a car may allow the household to move to a less expensive home in a less dense area. It is in this way that the choices of residential location and car ownership status are linked decisions, and it is important to model them jointly.

Given the joint nature of these choices, it is interesting to note that researchers focusing on these decisions in the United States often do not explicitly model the choice of car ownership. This may actually be a fine approach for many cities. Judging by the high levels of car ownership in most parts of the United States, households apparently consider the level of access to important destinations to be inadequate without a car, and car ownership becomes a necessity rather than a choice. However, when studying cities such as New York, where alternatives to the private car provide adequate access to important destinations, it is important to use a joint choice modeling approach.

To use myself as an example, I am planning to move to New York City in a few months. In deciding which neighborhood to live in, I will be heavily considering proximity to my job, proximity to public transport, and proximity to my favorite non-work destinations such as restaurants and parks. Because my preferred mode of transport is walking, I will be looking for a place that is quite close to these destinations. Because I know that in the neighborhood of my future work, owning and parking a private car is prohibitively expensive, I plan to live a car-free lifestyle.

1.2 What's so special about the "Big Apple"?

New Yorkers live in an urban environment that is uniquely dense, transit-rich, and car-poor in the United States, and New Yorkers make different transportation and residential location choices than other urban Americans. According to the 2000 US Census, New York City is the densest US city, with approximately 26,500 people per square mile (see Figure 1.1). The second ranking US city in population density is San Francisco, with approximately 16,500 people per square mile. The Census

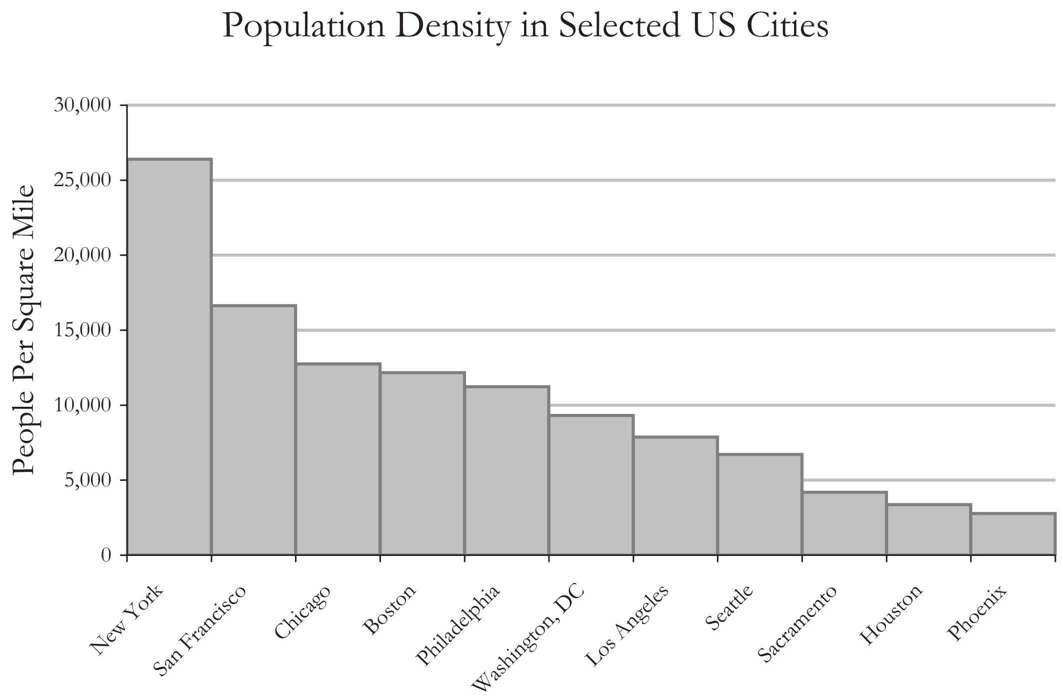

Figure 1.1: New York is by far the densest city in the United States.

defines city boundaries as the boundaries of the city proper, and does not include the entire metropolitan area of each city. Older cities such as New York are highly centralized, with approximately half of its metropolitan area population living within the city limits. Newer cities such as Los Angeles have a more even spatial population distribution, with less than one third of its metropolitan area population living within the city limits.

Transit provision is difficult to compare across cities, so I look at transit use. For commute trips, which is the trip purpose for which people everywhere use transit the most, more than half of New York commuters use transit (see Figure 1.2). The next highest cities in transit use for commuting are Washington, DC, Boston, and San Francisco, where approximately a third of commuters use transit. The number of New Yorkers reporting to have no vehicle access in the Census is also extraordinary for the US, at more than 55 percent (see Figure 1.3). The next highest level of car-free living is in Washington, DC, where 37 percent of households have no vehicle access.

This means that survey data gathered in New York City will have much higher

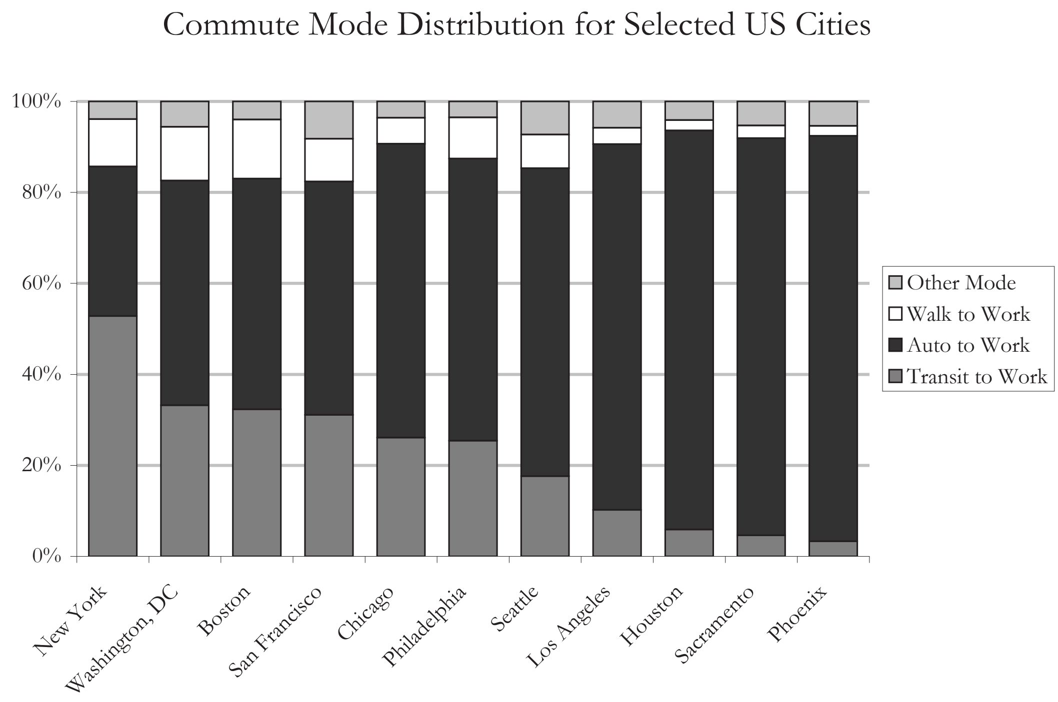

Figure 1.2: New York has the largest transit share for the commute mode of all US cities.

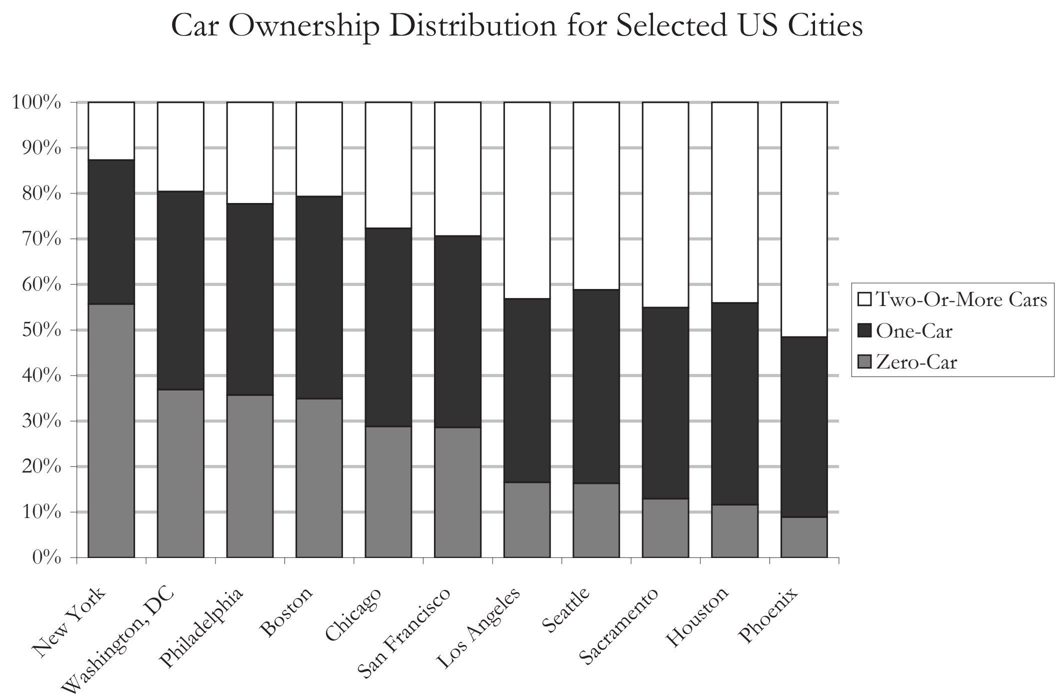

Figure 1.3: Note that New York City has the highest percent of car-free households in the United States.

variation in both the dependent and most of the key independent variables that are relevant to the modeling effort in this dissertation, allowing for more robust parameter estimation. Many previous studies of environmentally-friendly urban transportation behavior (not owning cars and using transit or walking for most trips) have suffered from the fact that a relatively small percentage of the people surveyed actually made these choices. Using classical statistical techniques, it is impossible to accurately estimate the determinants of choosing an alternative that hardly anyone in the sample actually chose.

Fortunately, the uniqueness of New York does not mean that New Yorkers have fundamentally different underlying preferences regarding the characteristics of the transportation and residential location options that they face. This dissertation presents evidence that in fact New Yorkers make their transport and residential choices almost identically to people in other cities - the difference in their behavior stems from the difference in the alternatives they face.

1.3 Conceptual background for the models

This dissertation uses discrete choice models to gain insight into how New Yorkers make decisions about where to live and how to travel. The estimated models include variables that are thought to affect New Yorkers' decisions about which neighborhood to live in, whether to own a car, and how to travel around the city. Economic theory guides the choice of variables to include in the models to explain each of these choices. This section provides an overview of the rationale for inclusion of variables in the models. More details will be provided with each model estimated in this dissertation.

Discrete choice models are essentially models of demand for the available alternatives. As in all models of demand, economic theory dictates that the variables included should represent the income of the individual making the choice, the prices of each alternative, and the prices of the substitutes and complements for each alternative. These prices need not be confined to be only in terms of money. Time prices can and should be included as well. Economic theory tells us that higher prices for an alternative or for its complements should decrease the demand for that alternative. Higher prices for an alternative's substitutes should increase demand for the alternative. Higher income should increase the demand for all alternatives that are considered to be normal goods. The magnitude of the effect of both income and the money prices of the alternatives is dependent on the original income of the decision maker. In general, the effect on choice of both income and money prices decreases as income increases. This means that for those who are already wealthy, increases in income or price have a smaller effect on choice than an income or price increase would have for a not-so-wealthy decision maker. In the case of differentiated goods, higher quality of a good should increase demand for that good.

To model the choice of transportation mode for a particular trip, then, variables to use in the estimation would include the time and money prices of traveling by each mode, the income of the traveler, and perhaps some measure of the relative qualities of service offered by each mode alternative. In modeling the choice of car ownership status, important variables include the price of owning the car, the income of the potential car owner, the availability of alternative transport modes in his or her residential location, and the access that is afforded by these alternative modes to important destinations. Because the models here do not aim to explain the choice between different vehicle types, but rather the choice of whether or not to own a car, the quality of the car does not enter into the analysis.

To model the choice of residential location, key variables include the price of a dwelling both in that location and in alternative locations, the income of the potential resident, the access afforded by this and alternative locations to important destinations, and measures of the quality of the neighborhood in other ways. These measures of neighborhood quality could include qualities of the people who already live there such as their race, income, and whether they own their homes. Most people prefer to live in neighborhoods where they "fit in", and fitting in means matching the racial makeup of the area and having a similar economic status as the neighbors. Researchers have hypothesized that neighborhoods with higher percentages of homeowning households are better cared-for, and therefore make for higher quality residential locations.

This dissertation is primarily concerned with gaining an understanding of the influence policy makers can have on people's transportation and residential location choices in New York. This places the focus on the variables that can be changed using policy levers. These include things like the prices of transport modes, the availability of transit, and the cost of parking. They do not include variables such as household income, household size, and the racial makeup of neighborhoods. It remains important for the estimation, however, to include these variables in the models because they serve an important role as control variables.

1.4 Previous literature

This dissertation is a continuation of the literature of behavioral models of transportation and residential location choice. Although transport mode choice models are commonly estimated, both residential location choice and car ownership choice models are less common. Joint choice models that endogenize more than one of these choices are even scarcer. Here, a brief review of the joint choice models that do exist in the literature is presented. These papers and books are the background for this dissertation's entire modeling effort. Literature that pertains to the specific applications used will be reviewed in later chapters.

Train (1980) uses survey data from the Bay Area to estimate a nested logit model of the choices of car ownership and transport mode to work. The sample used here is different from Train's sample in that a large fraction of the current sample is composed of non-car owners, allowing for more robust estimation of the coefficients on the variables relating to the car ownership choice. In choosing the variables related both to car ownership choice and commute mode choice for the models presented in this dissertation, I take direction from Train's 1980 work.

Modeling the choice of residential location presents the problem that the number of alternatives in the choice set is large. McFadden (1978) shows that consistent estimates of the coefficients can be obtained by estimating the logit model using a sample of the full set of alternatives, as long as the sampling rule satisfies certain properties. This estimation methodology is used in most studies of residential location choice, including Anas (1982), Quigley (1985), and this dissertation.

Anas (1982) estimates a nested logit model of the choices of housing, residential location, and transport mode for the work trip using US Census data from the Chicago area. Anas focuses on identifying the theoretical and empirical implications of using Census-style frequency data rather than fully disaggregate household observations.

Quigley (1985) focuses on the choice of residential location, using data from Pennsylvania to estimate a nested logit model of the choice of dwelling, neighborhood, and town. One of his main findings is that workplace accessibility is more important to household choice of residential location than other models have shown.

Lerman (1977) is the only existing paper known to me that estimates a joint logit model of the choices of housing, residential location, car ownership, and transport mode to work. Lerman used data from the Washington, D.C. metropolitan area in his analysis. This dissertation is in some senses an extension of Lerman's work using a different data set. In contrast to Lerman's work, this dissertation:

- is based on a data set from New York City that includes many observations of car-free households,

- draws on the additional 25 years of available literature (and computing advances) to improve and test the estimation techniques,

- uses the data to determine whether a nested or a joint logit model structure is more appropriate,

- uses GIS technology both to create spatial data and to illustrate the spatial heterogeneity of the results,and

- includes a chapter that focuses on the choice of walking level.

1.5 Outline of the dissertation

The dissertation is organized as follows. Chapter 1 provides the motivation for the dissertation, introduces both the overarching questions to be addressed and the main methodological approach, and provides context for the study site of New York City.

Chapter 2 details the common data and methodology used for the entire work. A review is provided of discrete choice theory, and detailed descriptions of how the data set was put together are given. Although this dissertation relies entirely on existing data sources, a substantial amount of effort was expended to obtain these sources and then to merge them into a single geographically-referenced data set that can be used for econometric estimation and spatially-explicit scenario analysis. Geographic information system technology was used to create some variables, and the statistical techniques of factor analysis, cluster analysis, and regression were used to create others. Summary statistics for the entire data set can be found in Chapter 2.

Chapter 3 begins the modeling effort, focusing on the effect of variables that can be affected by public policies on the choices of car ownership and car use for commuting. The empirical basis for the chapter is a multinomial logit model of the joint choices of residential location, car ownership status, and commute mode choice. The estimated coefficients in this model are interpreted, and using this model, the elasticities of car ownership and car use for commuting are calculated with respect to policy-relevant variables. This model is then contrasted with models that take one or more of the choices as given, identifying cases where results of these simpler models are substantially different from the results of the full model.

Chapter 4 is a spatial exploration of the results from Chapter 3. One of the unique aspects of the models developed in this dissertation is the fact that they are spatially-referenced. This means that it is possible to display the results from these models on a map of New York City, gaining additional insights into how changes in the explanatory variables of the models would produce spatial patterns of behavior change across space. This insight is crucial for evaluation and implementation of any policy alternatives that are implemented differentially across space.

Chapter 5 looks at the question of residential self-selection through a model of the joint choice of walking level, car ownership status, and residential location. It proposes a methodology for quantifying the extent of residential self-selection through estimation of elasticities that are based on probability results from the full choice model, conditional on residential location choice.

Chapter 6 concludes the dissertation, tying together and evaluating the analyses presented in Chapters 3 through 5. In the final section, this chapter identifies directions for future research.

Chapter 2

Methodology and Data

This chapter accomplishes two goals. First, it describes in detail the theory behind the statistical models used in the next three chapters of this dissertation. Second, it describes the data sources and details the process to make the dataset ready for use with these statistical models. This includes use of the additional statistical tools of regression analysis, factor analysis, and cluster analysis, as well as use of geographic information systems for spatial data organization.

2.1 Discrete Choice Model Methodology

Discrete choice models are distinct from continuous choice models in that the dependent variable can take only discrete values. In the models estimated in this dissertation, all of the dependent variables are categorical.

Similar to other types of economic models, discrete choice models assume that individuals will choose the alternative that yields the highest payoff. The difference in the case of discrete choice is that the alternatives are not available in every possible combination of continuously variable attributes. This means that it is unlikely that the absolute highest payoff can be reached, and the individual must settle for the alternative that has the highest payoff of those available. In the case of the models estimated here, this payoff is in the form of utility. If the utility function can be specified and estimated for each alternative, then choosing the alternative yielding the highest utility is a simple task.

2.1.1 Random utility theory and the multinomial logit model

Although the theoretical basis for discrete choice models differs little from that of continuous models, estimation of the two types of models differs substantially. Random utility theory forms the basis for the type of discrete choice model estimated in the present dissertation (see Train, 2002 or Ben-Akiva and Lerman, 1985 for further details on logit model theory). Under this model, a utility function based on the attributes $\mathbf{x}_{nj}$ of J alternatives as well as the characteristics $\mathbf{s}_n$ of N individuals is postulated to have both deterministic and stochastic parts.

$U_{nj} = V_{nj} + \epsilon_{nj}$ where: $U_{nj} = U(\mathbf{x}_{nj}, \mathbf{s}_n)$ and $V_{nj} = V(\mathbf{x}_{nj}, \mathbf{s}_n)$

The individual n chooses alternative i if and only if $U_{ni} > U_{nj}$ for all $i \neq j$. The ϵ_{nj} represent the portion of the utility that is not observable by the researcher. The probability that individual n chooses alternative i is then dependent on the distribution of the ϵ_{nj}, and is equal to:

$$P_{ni} = Prob(U_{ni} > U_{nj} \forall j \neq i)$$

$$P_{ni} = Prob(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \forall j \neq i)$$

$$P_{ni} = Prob(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \forall j \neq i)$$

$$P_{ni} = \int_{\epsilon} I(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \forall j \neq i) f(\epsilon_n) d\epsilon_n$$

where $I(\cdot)$ equals one when the term in parentheses is true and zero otherwise, and $f(\epsilon_n)$ is the joint density of the unobserved portion of utility over the alternatives. The logit model results when the ϵ_{ni} are assumed to be independently and identically distributed (i.i.d.) extreme value for all i. This is a convenient specification for the analyst because it results in an easily solved integral for the P_{ni}, making the choice probabilities equal to the following expression.

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j \in C_n} e^{V_{nj}}}$$

where C_n represents the choice set for individual n.

Estimation of discrete choice models is done using the method of maximum likelihood. This method begins with the assumption that the sample is the most likely sample to have been drawn from the population. A likelihood function is defined, consisting of the joint probability of drawing the sample observations, which is the product of the P_{ni}. Once the choice probabilities are specified, it is a relatively simple matter to generate the likelihood function. The method of maximum likelihood finds the set of coefficients for the independent variables in the V_{ni} that maximizes this function.

In a continuous regression model, the outputs from the estimation are the estimated coefficients with their associated standard errors and a goodness-of-fit measure for the model as a whole. From this information, elasticities can also be derived. From estimation of a discrete choice model, the outputs are somewhat different. Again, coefficients are estimated, but if the discrete choice model is based in random utility theory, these coefficients actually represent the parameters of the indirect utility function. Recall that the units of this function are utils, and are an ordinal rather than a cardinal numerical concept. This means that, while the signs of the estimated coefficients mean something in terms of the direction of the effect of that attribute on representative utility, the coefficient magnitudes are meaningful only in relationship to each other.

2.1.2 Assuming the Independence of Irrelevant Alternatives

A major limitation of the multinomial logit model is that the construction necessitates that alternatives do not violate the assumption of the independence of irrelevant alternatives (IIA). The IIA assumption is described well in an illustrative example by Train (2002).

> Suppose the set of alternatives available to a worker for his commute to work consists of driving an auto alone, carpooling, taking the bus, and taking rail. If any alternative were removed, the probabilities of the other alternatives would increase (e.g., if the worker's car was being repaired such that he could not drive to work by himself, then the probability of carpool, bus, and rail). The relevant question is: by what proportion would each probability increase when an alternative is removed? (p. 89)

If the probabilities of the remaining alternatives would increase by the same proportion, then the IIA property holds. If, however, there is some difference in the proportional increase in probabilities, the IIA property is violated. If the rail alternative were removed, for instance, one might expect that a disproportionate percent of the probability of choosing rail for a given individual might be allocated to bus because these two alternatives are both transit. This would be a violation of the IIA assumption, and it occurs because the unobserved utility (the model's error term) is correlated between the alternatives of rail and bus. As in this example, violation of the IIA assumption can happen even in a single choice model, but it is more likely in a joint choice situation.

In the present model, individuals make not one, but three discrete choices. They choose where to live, whether to own a car, and their transport mode. Since these three choices are interrelated, assuming that they are not by estimating separate multinomial choice models for each of them could lead to bias in the estimation due to incorrect model specification. Using a multinomial logit model for the joint choices is a better approach, since the joint choice model endogenizes all three sub-choices in the same model.

However, the relatively simple joint choice model could still have estimation bias due to the fact that interdependence between choices can actually cause the IIA property to be violated. An example will illustrate this idea best. Imagine a model of the joint choice of residential location and commute mode choice. There are two location options - urban and rural - and two mode options - car and transit. There are therefore four alternatives, with the choice set defined as follows {urban & car, urban & transit, rural & car, rural & transit}. The joint choice model assumes that all four of these alternatives satisfy the IIA property. However, it is easy to see that there might be a correlation between the sets of alternatives that include the same residential location. If the alternative "urban & car" were removed, it might be that the likelihood that someone previously in the "urban & car" alternative group would choose "urban & transit" is disproportionately high. This would be a violation of the IIA property, and the joint choice model would be biased.

Using this example, it is hard to imagine a situation where a compound choice situation would be correctly modeled using a joint choice model. However, it is important to remember that the joint choice model is only biased if the correlation between alternatives is in the error term of the model, and not accounted for by the explanatory variables. After a rigorous model selection process, the selected "best" models in this dissertation are all actually joint choice models. This topic will be discussed further where the model selection process is described in Chapter 3.

The assumption of the independence of irrelevant alternatives is an artifact of the simplicity of the multinomial logit mathematics; the multinomial logit probability ratios between any two alternatives do not depend on other alternatives. Every model estimated in this dissertation has been tested for violations of the IIA property by relaxing the assumption and comparing the results as follows.

2.1.3 Using the nested logit model to relax IIA

The estimation method of nested logit is used in this dissertation to relax the IIA assumption as a test of the joint multinomial logit specification. The nested logit allows for structured correlation across the unobserved utility of a subset of the alternatives in the choice set. These subsets of the alternatives are the "nests". Within each nest, the alternatives are assumed to be closer substitutes for each other than they are for the alternatives outside of the nest. Although some analysts let the data dictate the nesting structure, the nests are usually chosen according to what the analyst believes to be theoretically defensible groupings of the alternatives into close substitutes.

In a two-level nested logit model with K nests, the probability that individual n chooses alternative i that is in nest k is as follows:

$$P_{ni} = \frac{e^{\frac{V_{ni}}{\lambda_k}} \times (\sum_{j \in J_{nk}} e^{\frac{V_{nj}}{\lambda_k}})^{(\lambda_k - 1)}}{\sum_{l=1}^{K} (\sum_{j \in J_{nl}} e^{\frac{V_{nl}}{\lambda_l}})^{\lambda_l}}$$

where: λ_k is the inclusive value for nest k (the inclusive value is related to the correlation between alternatives in the nest),
J_{nk} is the set of alternatives for individual n in nest k, and
K is the total number of nests.

Within each nest, the IIA property must still hold. Between nests, however, probability ratios can depend on other alternatives.

It is the estimated coefficients on the inclusive values that provide estimates of the extent to which the choices are interdependent. These coefficients should range from 0 to 1 to be consistent with utility theory (although real-world models often return coefficients greater than 1). A coefficient of 1 indicates that the unobserved portions of utility for the alternatives within a nest are not correlated, and therefore a joint multinomial logit is an appropriate model specification. A coefficient that is between 0 and 1 and significantly less than 1 indicates positive correlation between the unobserved portions of utility for alternatives within the nest, and therefore the IIA assumption is not valid and the joint multinomial logit specification would be incorrect. Greater detail on this topic is provided in the model selection section of Chapter 3.

2.1.4 Elasticities

As described above, discrete choice models parameterize an indirect utility function. This means that the estimated parameters indicate how the variables affect the utility of the alternatives. From this, the probability for each individual of choosing each alternative in the choice set can be determined as shown in the equations of the previous sections of this chapter. This information, while interesting in its own right, fails to provide a clear indication of what effect policy might have on this choice behavior.

One way to extract policy-relevant information from a discrete choice model is to calculate elasticities from it. In a continuous model, an elasticity is defined as the percentage by which the dependent variable changes when a particular independent variable is increased by one percent. This definition becomes problematic when the dependent variable is discrete, however, because it cannot change by small amounts.

Elasticities estimated from discrete choice model results, therefore, have a slightly different interpretation. An elasticity in a discrete choice model is defined as the percent change in the probability of choosing a particular alternative when a particular independent variable is increased by one percent. This sounds as though it is quite far removed from being a policy-relevant measure, but in fact it is not. Since the estimated probabilities for the alternatives can be viewed as estimates of the market shares of the alternatives, the elasticities can be interpreted as the change in the market shares of the alternatives that arise from a one percent change in an independent variable. Calculation of appropriate elasticities from discrete choice models is not, however, a trivial matter.

All of the elasticities presented in this dissertation are calculated in the following way. First, the coefficients that parameterize the model are estimated based on the actual data, and the predicted probabilities[1] for each alternative for each individual are calculated. These predicted probabilities are represented by $wtp0_{nj}$ in the equations that follow. Second, the independent variable for which the elasticity is being calculated is increased by one percent. For example, to calculate the elasticity of car use for commuting with respect to the cost of commuting by car, the cost of commuting by car is increased in the data by one percent. Third, the predicted probabilities are recalculated for each alternative for each individual. These predicted probabilities are represented by $wtp1_{nj}$ in the equations that follow. Note that the model is not re-estimated, rather the existing model estimates are used to predict new probabilities based on the altered underlying data. Fourth, both the original and the new predicted probabilities are summed over the alternatives that contain the relevant sub-choice. For my example calculation of the elasticity of car use for commuting, this would mean summing all of the probabilities of choosing alternatives that included the sub-choice of commuting by car. Fifth, the percent change in the probability of choosing the relevant alternative is calculated for each individual n. This is represented as the individual elasticity estimates ϵ_n in the equations below. Finally, the individual elasticity estimates are averaged across the sample, weighted by the original probability for each individual of choosing the alternative. The final

[1]These probabilities are weighted by the chosen residential neighborhood of the household, as described in the "Neighborhood Weights" section later in this chapter.

elasticities are given by ϵ. This weighting is necessary because, for example, a change from a one percent probability to a two percent probability will appear as a 100 percent increase in the probability, but actually represents almost zero change in the likelihood of choosing that alternative.

In equation form, the elasticity estimates in this dissertation can be represented as follows:

$$\epsilon = \sum_n \frac{\sum\limits_{j \in J} wtp0_{nj}}{\sum\limits_n \sum\limits_{j \in J} wtp0_{nj}} \epsilon_n$$

where: $\epsilon_n = \frac{\sum\limits_{j \in J} wtp0_{nj} - \sum\limits_{j \in J} wtp1_{nj}}{\sum\limits_{j \in J} wtp0_{nj}}$,

n indexes individuals,
j indexes alternatives,
J is the set of alternatives that contain the relevant sub-choice,
$wtp0_{nj}$ is the neighborhood-weighted probability that individual n chooses alternative j in the base model, and
$wtp1_{nj}$ is the neighborhood-weighted probability that individual n chooses alternative j in the model with the altered underlying data.

2.1.5 Weighting the observations

In the case where a random sample has been collected, there is no need to weight the observations. On the other hand, in the case where a certain sub-population has been oversampled, using this data set as if it were a random sample of the population can bias the results. Oversampling of a sub-population can occur either because the researcher collecting the data used the method of choice-based sampling, or because a systematic pattern exists in the choices of survey respondents versus those who refused to participate in the survey. To obtain unbiased results from a non-random sample, it is necessary to weight the observations so that they are more representative of the underlying population.

The weighting scheme used in this dissertation follows Manski and Lerman (1977). The weight for each neighborhood is the percent of the population in that neighborhood (according to the 2000 Census) divided by the percent of the sample in the neighborhood, as in the following equation.

$$\frac{Neighborhood\ Population/NYC\ Population}{Number\ of\ Sampled\ Individuals\ in\ Neighborhood/Total\ Sample}$$

If a neighborhood is represented in the sample exactly how it is represented in the population, the weight will be 1. If the neighborhood is underrepresented (overrepresented) in the sample, the weight will be greater than (less than) 1.

These weights are used in the estimation by multiplying each of the probabilities P_{ni} by the neighborhood weight for that individual, and using these weighted probabilities to create the joint probability function to be maximized.

2.1.6 Sampling from the alternatives to reduce the choice set size

Most of the multinomial logit models estimated in this dissertation include the choice of residential location as defined by a census tract. It turns out that there are 2216 census tracts in New York City. This presents a computational estimation challenge, even for today's computer systems. In the case of the choice models that endogenize more than just the choice of residential location, the number of alternatives in the choice set is this number of residential location alternatives *multiplied by* the number of car ownership status alternatives and/or mode choice alternatives. In this dissertation, the model with the largest number of alternatives in its choice set has more than 40,000 alternatives.

To reduce these choice sets such that they are computationally manageable, I follow McFadden (1978) by taking a sample of the alternatives to be the choice set in my statistical models. McFadden proved that estimating a model using a random sample of the alternatives as the choice set is asymptotically equivalent to estimating the model using the full choice set. My sampling methodology is as follows. First, the chosen residential location was set aside for each individual to ensure that every individual's choice set included their actual choice. Then, 10 additional census tracts were randomly sampled for each individual, making 11 location alternatives in each individual's choice set. In the case of the models that endogenize more than just the choice of residential location, the compound choice set was created for each individual that included all feasible mode-car ownership combinations and 11 possible census tract locations, making modeled choice sets of compound alternatives that are computationally feasible.

2.2 Data

Incorporated into the present dissertation are data extracted from eight separate data sources. The main data source is the Regional Travel - Household Interview Survey (RT-HIS), conducted in the fall of 1997 and the spring of 1998 by NuStats International and jointly commissioned by the New York Metropolitan Transportation Council and the North Jersey Transportation Planning Authority. The survey was completed by 11,264 households in the 28 counties that comprise the New York-New Jersey-Connecticut metropolitan area. Households completed both a 24-hour travel diary on an assigned day and a lengthy telephone interview that collected information about their socioeconomic situation, their residential location choice, and their travel habits. The data was condensed into four basic data files: household-level information, person-level information, trip-level information, and vehicle-level information. The present analysis relies heavily on the first three of these.

The RT-HIS used a complex sampling plan to insure that all of the counties were adequately represented and that all of the important transportation mode-residential density combinations were adequately represented in the sample. Note that the sampling plan did *not* insure that all mode-residential density combinations were adequately represented within each county. To accomplish this, each census tract was classified based on two dimensions: residential density and a quality that the survey labels “mode leadership”. The surveyors identified seventeen possible Mode Leadership-Density categories. “Mode leadership” relates to the available transport modes in the area. In the models estimated in this dissertation, weights (described above) were used to correct for the bias that could result from this sampling stratification.

As the present analysis focuses on New York City proper, it uses only the 3,397 households that completed the survey in the five boroughs of the city. In these urban households live 7,505 individuals who made a total of 23,115 trips on their assigned travel diary days. This data set provides the individuals doing the choosing in the model - the dependent variables - as well as most of the independent variables used to explain the travel mode sub-choice and some of the independent variables used to explain the car ownership and residential location sub-choices.

The rest of the independent variables come from a variety of other data sources including the US Decennial Censuses of 1990 and 2000, the 1997 Business Patterns Census, the New York State Insurance Department, the National Climatic Data Center, the New York City Department of City Planning, and New York City Transit.

In all of the estimations in this dissertation, households were weighted using the methodology described above by their residential neighborhood[2] choice. These weights are used to make the geographic distribution of the estimation sample closer to the actual geographic distribution of New York City's population.

2.2.1 Use of geographic information systems (GIS) to merge data sets

Each of these data sources provides its information at a different level of spatial aggregation. The 1990 and 2000 Censuses and the Regional Travel - Household Interview Survey provide information at the level of the census tract, the 1997 US Business Patterns Census provides information by zip code, and a number of other data sources provide disaggregate data with exact geographic location information.

In order to transform these data sources so that they are at the same level of spatial aggregation, a number of steps are involved. First, GIS was used to geographically match larger geographic areas such as zip codes and insurance areas to census tracts. Unfortunately, the boundaries of census tracts and these larger geographic areas do not always exactly line up. The approach used here was to first find the

[2]Determination of what constitutes a "neighborhood" for this purpose is detailed in a later section of this chapter.

center of each census tract, and then identify which of the larger geographic areas contain these census tract centers. Of course, this approximation does result in some error because some households living at the edge of a census tract whose center is in one zip code area might actually live in another zip code area. However, since the residential locations of the households are not geographically identified at the sub-census-tract level in my data, this cannot be helped.

2.3 Variables used in the discrete choice models

The joint choice of residential location, car ownership status, and transport mode is dependent on many variables. Some of the variables help to explain all three sub-choices, while others explain only one or two of the sub-choices. This section aims to introduce the explanatory variables used in the choice models in the next three chapters, providing summary statistics for the sample. It will also detail which data source they came from and, in some cases, how they were derived from the raw data.

The sample consists of half men and half women, 25% black, 43% white, 17% Hispanic, and the remainder some "other" race. Thirty-one percent of the people in the sample are not licensed to drive a vehicle. Thirty-six percent of them have children under the age of 18 at home. Thirty-six percent of them are homeowners.These sample summary statistics are geographically weighted sample statistics from the sample used in Chapters 3 and 4 of this dissertation, which includes only commuters. The sample used in Chapter 5 includes all adults sampled in New York City, and is therefore larger. See Table 2.1 for a comparison between the unweighted sample composition, the weighted sample composition, and the 2000 Census information.

2.3.1 Dependent variables

The data that underlies the dependent variables in this dissertation are taken from the RT-HIS data set. Figures 2.6 through 2.8 are pie charts of the dependent variables in the data set. These figures represent the portion of the full RT-HIS dataset that includes commuters and was used in the estimation and analysis of Chapters 3

Table 2.1: Sample Composition vs. Census Information

	2000 Census	**Unweighted Sample**	**Weighted Sample**
Percent black	27%	15%	25%
Percent white	45%	64%	57%
Percent hispanic	27%	10%	17%
Percent with kids (HH)	34%	29%	37%
Percent homeowners (HH)	30%	38%	34%

and 4. The estimation and analysis in Chapter 5 was done on an expanded dataset that includes all adults in the sample, whether they are employed or not, so the summary statistics are slightly different. For the sub-choice of residential location, the dependent variable is simply the census tract that the individual resides in. For the sub-choice of car ownership status, the dependent variable is the number of household vehicles, grouped into 0-car, 1-car, and 2-or-more-car categories. Both of the dependent variables that make up these sub-choices are common throughout the dissertation.

The dependent variable that represents the transport mode sub-choice is a bit more complicated. The original data contains 26 possible transport mode alternatives. Many of these alternatives, however, are seldom actually chosen (e.g. wheelchair). Some of them are not actually available to the residents of New York City, and are included in the model because they are available to residents of New York suburbs. For the analyses in this dissertation, the 22 of these 26 modes that are available to those who reside in New York City have been aggregated into seven modes as illustrated in Figure 2.1. In early model runs, aggregation into 12 modes, 8 modes, and 5 modes were tested. The models with seven mode alternatives performed the best - enough individuals chose each of the seven modes to allow for robust model estimation, but the aggregation did not seem to obscure important results.

The transport mode sub-choice in Chapters 3 and 5 of this dissertation are based on different parts of the RT-HIS data. Chapter 3 focuses on the commute trip for the mode sub-choice of the model, while Chapter 5 focuses on the percent of trips on the travel day that were walk-only trips. One of the questions survey

Walk (only) Bicycle Skates Wheelchair Ferry	**Walk**
Yellow-Medallion Taxi For Hire Van-Jitney Car Service-Black Gypsy Cab	**Taxi**
Auto Passenger	**Auto Passenger**
Auto Driver	**Auto Driver**
Express Bus Airport Service Charter Bus Local Bus Contract Bus Shuttle-Commute Van	**Bus**
Subway (NY) + Walk PATH + Walk	**Subway with Walk Access**
Subway (NY) + Non-Walk PATH + Non-Walk Commuter Rail	**Subway with Non-Walk Access**

Figure 2.1: Mode Aggregation from Original 22 Modes to 7 Modes Used in Model

respondents answered on the "lengthy telephone interview" mentioned above was, "What do you use most often to get to work?" Responses to this question were used as the commute mode part of the dependent variable in the Chapter 3 analysis. In contrast, the analysis in Chapter 5 relies on the travel day mode choices to determine walking level.

2.3.2 Independent variables

Figures 2.9 through 2.29 display additional attributes of some key independent variables for the sample used in this dissertation. Some of these figures are pie charts that display the proportions of characteristics in the weighted sample. Others are histograms of the distributions of variable values in the weighted sample.

The origin of the travel time data used in this dissertation deserves further explanation. Travel time is an included self-reported variable in the RT-HIS data set. Another variable included in the RT-HIS data set is trip distance. The trip distances are calculated rather than self-reported, based on origin and destination addresses given by survey respondents. The calculated distances are the shortest path on the street network and on the rail network, and therefore are trustworthy for walking, rail, and auto trips, but less so for bus trips. The reason they are less trustworthy for bus trips is that bus routes are often not the most direct path from origin to destination.

Dividing the trip distance by the travel time yields the trip speed. For many trips, these calculated speeds are not within a reasonable range for the reported transportation mode. The problem is twofold. First, there are the inaccuracies in the trip distances for bus and train trips as noted above. Second, there are problems with self-reported travel time data. Some survey respondents will not report travel times accurately due to simple recollection difficulty, but even respondents with good memories are highly likely to round travel times to the nearest 5, 10, or even 30 minutes. Because of these issues, the travel time variables used in the models in this dissertation are not those from the RT-HIS data set.

Instead, the travel time variables in the models in this dissertation are calculated based on the RT-HIS trip distance variable and speeds that were estimated by ordinary least squares regressions for each mode alternative. The estimated regressions are given in Tables 2.2 through 2.5.

The dependent variable in each of these regressions is actual average speed data taken from the RT-HIS for each mode alternative, period of the day, and borough origin-destination pair. Even though the discrete choice models in this dissertation that use the travel time variables are the choice of commute mode, the average speed regressions include all of the trip purposes in the data set. Even using all of the trips in the data set, there were many mode-time-origin-destination combinations for which there were zero or very few observations. It is for this reason that regression analysis was used rather than simple average speeds for each category.

Using average speed data rather than individual trip speed data reduces the effect of the problems identified above with the raw trip distance and travel time information in the RT-HIS. Each mode-time-origin-destination category that had at least 30 observations was used as a data point for the speed regressions. There are two ways to calculate average speeds. The first is to calculate the speed for each trip, and to average the speeds. This leads to overrepresentation of the speeds of shorter trips in the average. The second is to average trip speeds weighted by the length of the trip. This leads to overrepresentation of the speeds of longer trips in the average. The solution used here was to calculate both types of average speed, and then to use the mean of these two measurements as the dependent variable in the speed regressions.

A second variable that deserves further explanation is the car insurance price variable. Car insurance prices vary depending on the household's driving history, the age and gender of the drivers in the household, the type of car or cars owned by the household, the home location of the household, and how much driving is done each year by the household drivers. The RT-HIS did not include any actual household-specific information about car insurance prices. Fortunately, the New York State Insurance Department does collect car insurance quotes from 25 major insurance

Table 2.2: Speed Regression for Auto Mode

Dependent variable = Average Auto Speed R-squared = 0.840 Number of observations = 7466		
Variable	Coefficient	Standard Error
Avg of Trip Distance	0.482***	0.009
Same Borough	-2.505***	0.097
Staten Island (O or D)	2.605***	0.040
Origin Manhattan	-1.222***	0.061
Destination Manhattan	-1.553***	0.061
12am-6am	0.603***	0.111
6am-10am	-0.544***	0.059
10am-4pm	-1.068***	0.056
4pm-8pm	-1.327***	0.057
Constant	12.129***	0.128

Table 2.3: Speed Regression for Bus Mode

Dependent variable = Average Bus Speed R-squared = 0.846 Number of observations = 1962		
Variable	Coefficient	Standard Error
Avg of Trip Distance	0.478***	0.019
Same Borough	-3.407***	0.213
Staten Island (O or D)	0.821***	0.093
Origin Manhattan	-0.582***	0.102
Destination Manhattan	-0.030	0.100
12am-6am	0.243	0.304
6am-10am	-0.084	0.153
10am-4pm	-0.044	0.152
4pm-8pm	-0.726***	0.158
Constant	7.578***	0.299

Table 2.4: Speed Regression for Subway Mode

Dependent variable = Average Subway Speed
R-squared = 0.744
Number of observations = 3055

Variable	Coefficient	Standard Error
Avg of Trip Distance	0.841***	0.016
Same Borough	0.965***	0.091
Staten Island (O or D)	1.304***	0.165
Origin Manhattan	0.576***	0.048
Destination Manhattan	0.725***	0.048
12am-6am	-2.023***	0.166
6am-10am	0.117	0.079
10am-4pm	0.285***	0.080
4pm-8pm	-0.140*	0.079
Constant	1.455***	0.167

Table 2.5: Speed Regression for Commuter Rail Mode

Dependent variable = Average Commuter Rail Speed
Adjusted R-squared = 0.764
Number of observations = 460

Variable	Coefficient	Standard Error
Avg of Trip Distance	0.249***	0.033
Same Borough	-3.371***	0.388
Staten Island (O or D)	-1.596***	0.331
Origin Manhattan	0.736***	0.189
Destination Manhattan	-0.528***	0.192
12am-6am	-0.702	0.472
6am-10am	1.143***	0.362
10am-4pm	0.245	0.339
4pm-8pm	-0.657**	0.333
Constant	8.977***	0.504

companies, and publishes a report each year with these prices by home location region and gender at three ages. The car insurance price variable in the models in this dissertation is derived from this data source. Specifically, the median car insurance price was chosen for each home location region (of which there are 7 in New York City) and each age and gender of driver. Then, prices were estimated via direct linear interpolation for the ages not represented. As indicated in the data from the New York State Insurance Department, car insurance prices in this dataset drop with age until age 30, stay constant until retirement, and then drop further. These estimated prices were then matched with the age and gender of the head of each household to represent the car insurance price in each possible residential location.

2.3.3 Use of GIS in data creation

A number of key pieces of information used to create some of the independent variables in the models in this dissertation were created using GIS. These include the distance from home to work for all of the residential location alternatives that were not chosen, the distance from home to the nearest subway station for all residential location alternatives, the distance from home to midtown Manhattan for all residential location alternatives, and the number of subway lines available within a half-mile radius of home and work for all residential and employment locations. The first two of these were not used directly as variables in the models, but instead were used along with the mode-origin-destination-time-specific travel speeds (as described in the "variables" section earlier in this chapter) to obtain the estimation variables Ride Time and Walk Time.

The distances from home to work and from home to midtown Manhattan were calculated using the same GIS methodology. For these two variables, the centers of census tracts were used as origin-destination pairs, and the shortest path between them along the street network was calculated using the Network Analyst extension in ArcView 3.3. An example is shown in Figure 2.2.

Calculation of the distance from home to the nearest subway station was a bit trickier. For this, the geographic aggregation of the neighborhood was used (see next section for details of neighborhood determination), and the neighborhood average distance between a possible home location and its closest subway station was calculated. To do this, I created a grid of points spaced evenly throughout the neighborhood area, except for where there were parks or bodies of water. These points represent possible home locations. Then, the distance between each of these points and its closest subway station was calculated (including subway stations outside of the neighborhood as well) along the street network using the Network Analyst extension of ArcView 3.3. Finally, the average of these distances was used in the creation of the variable Walk Time for the Subway, Walk Access mode alternative. Figure 2.3 illustrates this methodology for a neighborhood in downtown Manhattan.

Identifying the number of subway lines available (note that this differs from subway stations) within a half-mile radius of home and work locations involved yet another type of GIS application. Assuming that the population in a census tract all live at the center point of the tract and work locations within a census tract are likewise at the tract's center, circles were created with centers at these points and radii of a half-mile. Then ArcView 9.0 was used to make a list of the subway stations within each of these circles along with the subway lines that were available at each station. Finally, database software made it possible to count the number of subway lines represented by these stations.

2.3.4 Missing values and the imputation of income

In addition to the spatial aggregation issues in putting together this dataset, there is also the problem that approximately 25 percent of the RT-HIS households did not report their incomes. There are a number of possible solutions to the missing data problem (and a substantial bit of statistical theory to accompany them). These include everything from listwise deletion (throwing away households with missing data) to auxiliary regression to using existing data to generate a distribution for the missing values, and then drawing from this distribution.

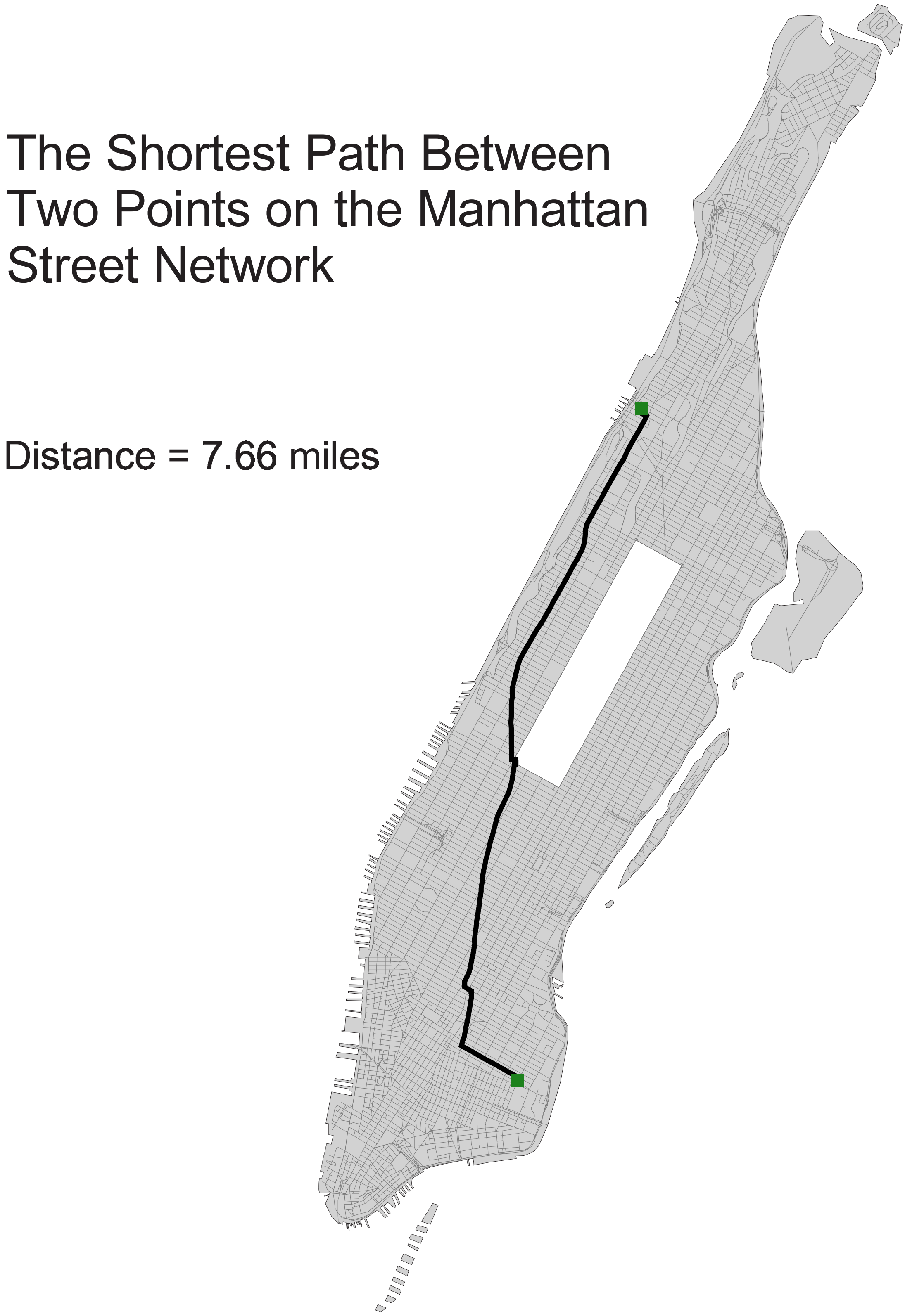

Figure 2.2: Shortest Path Between Two Points Illustration

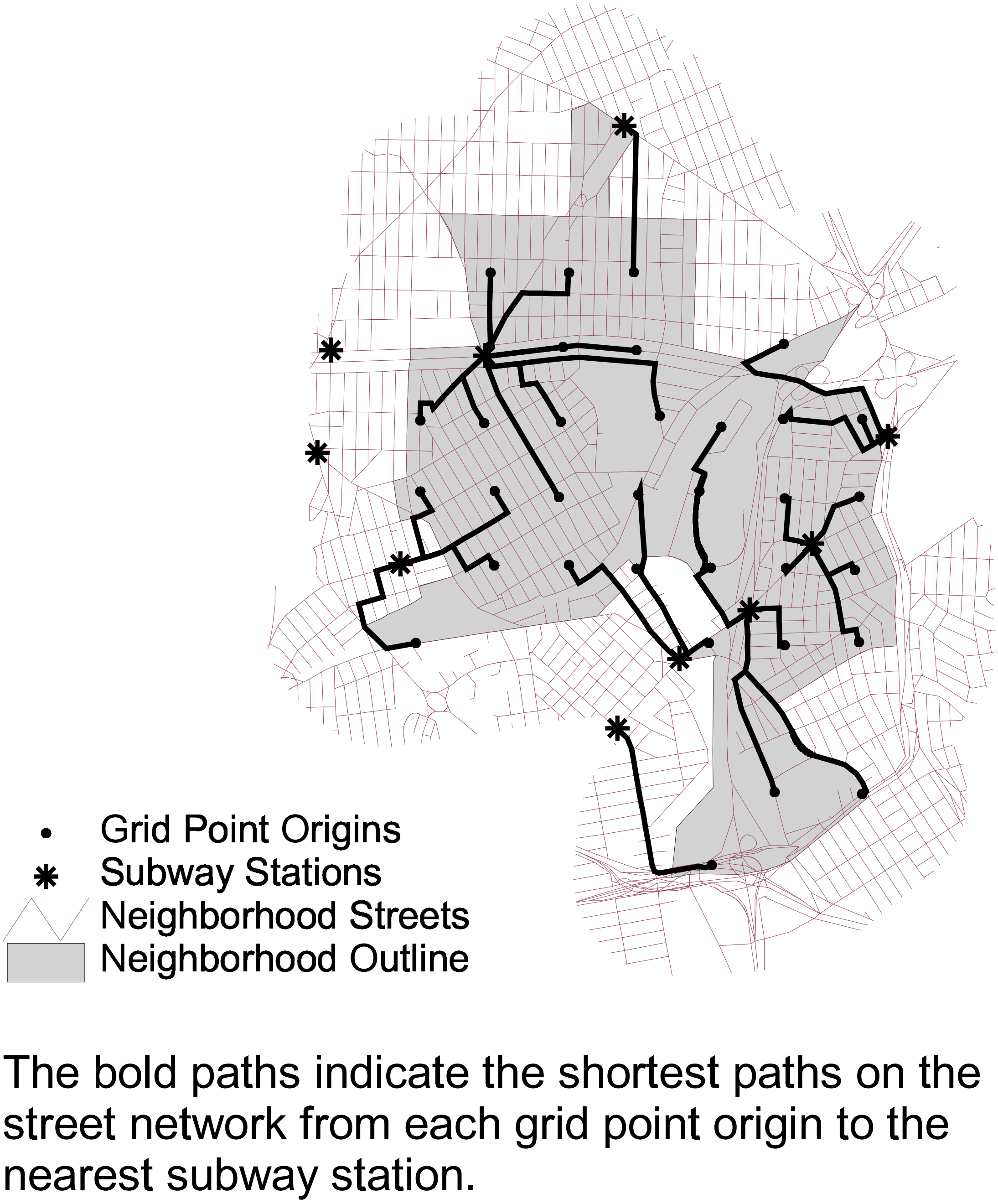

Figure 2.3: Distance from Home to Nearest Subway Station Illustration

I have chosen to use one of the simpler imputation procedures - the auxiliary regression. The reason for this choice is that it does not throw away information and it appears to produce results similar to more advanced methods, particularly for relatively large datasets (as mine is) and in the case where the data appear to be missing-at-random[3] (Onozaka, 2002). The results of the auxiliary regression are reported in Table 2.6.

For other variables that have missing values, the method of listwise deletion was used. This method is deemed acceptable in cases where less than 5 percent of observations have missing data (Schafer, 1997), which is the case for all variables in this model except for income.

In addition, there were a few cases where the information in the data file simply did not seem possible, and these households/people were removed from the estimation sample. Examples of this include two households who claimed to reside in Central Park, one single-person "household" that contained only a 6-year-old, and a few individuals who claimed to walk more than 5 miles (one-way) for their commute.

2.3.5 Data that this dissertation does not include

Additional variables that would have been desirable to have included are neighborhood-specific crime statistics, local school quality information, and home parking costs. Attempts were made to obtain all of these variables, but none were fully successful. The New York Police Department crime statistics are publicly available at the precinct level. This data was originally incorporated into the analyses in

[3]Missing-at-random (MAR) means that the pattern of missingness in the data is dependent on the values of the other variables, but not on the value of the missing variable. For instance, if older people tend not to report their incomes, the pattern of missingness is MAR. If, on the other hand, people with high incomes tend not to report their incomes, the pattern of missingness is not MAR. To assess the pattern of missingness in this dataset, a set of simple linear regressions were performed. From this analysis, it appears that households that did not report their incomes in this dataset are not systematically high-income or low-income households. However, it does appear that there is a pattern to the missingness that is dependent on the values of other independent variables such as age. Therefore, my data is MAR.

Table 2.6: Auxiliary Regression For Income Imputation

Dependent variable = log of income		
R-squared = 0.494		
Number of observations = 2640		
Variable	Coefficient	Standard Error
Log of median income	0.282***	0.044
Number of HH phone lines	0.210***	0.049
Number of HH workers	0.342***	0.027
Manhattan HH	0.196***	0.039
Bronx HH	0.127**	0.051
Hispanic HH	-0.300***	0.058
Black HH	-0.170***	0.057
Other ethnicity HH	-0.116**	0.054
Homeowner	0.294***	0.047
Live in house	-0.056	0.047
HH Head under 30	-0.125**	0.047
HH Head over 60	-0.118**	0.057
Male HH head	0.122***	0.038
HH Head is a manager	0.219***	0.053
HH Head is a professional	0.080*	0.047
HH Head is a clerical	-0.194**	0.072
HH Head is a laborer	-0.405***	0.142
HH Head in finance	0.211***	0.053
HH Head in health	0.123**	0.055
HH Head works on weekends	-0.134***	0.044
HH Head at job more than 5 years	0.173***	0.042
Constant	6.795***	0.463

this dissertation. Unfortunately, precincts are too large a geographic unit for this information to be statistically significant as a determinant of either residential location or travel choices, and therefore the crime variables were dropped from the models.

The New York City Department of Education provides detailed information about every public school in the city, including a variety of quality measures, its street address, and its catchment area. Unfortunately, even though the Department of Education has full digital maps of all of these attributes of their system, these maps are not publicly available. A decision was made not to recreate these maps manually.

The reason for this decision is that local school quality is less important as a determinant of residential location choice in New York City than it is in most other cities and towns in the United States. In New York, a much larger percent of children than in most other areas of the US have effective school choice. Even though there was no large-scale, officially-sanctioned school choice program in New York City in 1997-98, children did not necessarily attend the public school that was assigned to them by virtue of their home location. For instance, parents in New York will often send their smaller children to the public school near their job location rather than near their home location. Similarly, parents will often petition to send their children to a neighboring school if that school is on their way to and from work. Older children often ride public transit to school, and therefore have greater flexibility in which school they attend. On top of this public school effective choice, New York parents who have enough money will pay to send their children to private school. Total New York City public school enrollment was 1,083,943 in 1998, while non-public school enrollment was 268,913 for that same year (New York City Department of City Planning, 2002). This means that 20 percent of all school children in New York City attended non-public schools.

Unfortunately, the RT-HIS did not collect information on home parking costs. The survey did ask questions about work parking costs, but few respondents answered them and in New York, parking rates in the same place for residents will be different than those for commuters. This is because residents and commuters have different parking needs in terms of both hours of the day when parking is needed and length of parking contract. For these reasons, the parking cost data that was collected by the RT-HIS is not useful for estimating home parking costs.

Data from other sources that accurately represents the cost to park a vehicle in New York City are not readily available. Part of the reason for this is that the parking decision is complex in New York City, particularly so in Manhattan. There are three basic categories of parking services in Manhattan. The most convenient, secure, and expensive service is a parking garage near one's home. This type of parking service is expensive (between \$300 and \$700 per month, depending on the home location), and the people who choose it are likely to be wealthier and to have more expensive vehicles. The second type of service is not as convenient, but is secure and less expensive. This option is a parking lot or garage farther from one's home in a less-expensive area of the city - or even outside of the city limits. This is a popular option for middle income people who own cars but use them infrequently. The final parking service option is on-street parking. This option is the least expensive and it is not secure or convenient. People who park cars on the street in Manhattan are required to move them each day due to street cleaning regulations, and finding an on-street parking space each day takes a significant amount of time. This time cost is the entire cost of parking for most people who choose this option - incredibly, on-street parking in most parts of Manhattan remains free, particularly at night.

In the models in this dissertation, a number of home parking assumptions and cost estimates were tested. None of them were found to be acceptable, and home parking cost remains a conspicuously missing variable.

2.4 Neighborhood Determination

In this dissertation, the more-than-2000 census tracts that comprise New York City are aggregated into 51 neighborhoods. These neighborhoods are used primarily in the weighting of observations as described earlier in this chapter. The neighborhoods are also used in the creation of some of the variables that were created using GIS. Chapter 4 of this dissertation uses the neighborhoods heavily to represent spatial differentiation in model results on maps.

The problem of neighborhood determination is more challenging than it appears at first glance. Census tracts in New York City are geographically small; there are 2216 census tracts covering the 308.9 square miles of New York City. In Manhattan alone, 296 census tracts cover its approximately 23 square miles. For example, the "West Village" (a recognizable neighborhood in lower Manhattan), for example, is composed of 10 census tracts.

For the purposes of the present modeling exercise, these neighborhoods should have three key properties:

- They should be groups of census tracts so that census data is easy to use,
- The census tracts should be grouped such that within each neighborhood, they are similar, but different neighborhoods have distinct characteristics, and
- They should be geographically contiguous.

There is no existing set of defined neighborhoods in New York that fits these criteria.

This dissertation makes use of statistical cluster analysis to group New York's census tracts into "neighborhoods". Cluster analysis is a multivariate statistical toolset for grouping data that fulfills the first two of the three criteria listed above. In a broad sense, cluster analysis techniques aim to minimize within-group variance while maximizing between-group variance. Cluster analysis techniques do not easily allow for geographic contiguity constraints in groups, and this is an issue that will be discussed further later.

The present analysis uses two types of cluster analysis. Optimization cluster analysis determines which census tracts should be grouped into neighborhood clusters given the number of clusters to create. Hierarchical cluster analysis is used to provide guidance as to how many natural clusters there are in a dataset.

Before detailing these types of cluster analysis that are used here, this section describes some properties of cluster analysis in general. Both types of cluster analysis used here are distinct from many other multivariate statistical procedures (such as regression analysis) in a number of important ways. This should not be surprising, since the objective of cluster analysis is completely different from that of regression. However, it is easy to fall into the habit of assuming that statistical tools generally have properties similar to regression, and these next few paragraphs emphasize that this is not true.

In cluster analysis, every included variable is weighted equally. This has a number of implications. It means that although the analyst could manually weight the data, the procedure does not assign coefficients (other than 1) to the variables. This is fundamentally different from regression; if a variable in a regression analysis is not statistically significant, it is assigned a coefficient of effectively zero. This property of regression can be viewed as one way that the procedure corrects for analyst error in deciding which variables to include. Cluster analysis is not self-correcting in this way; if a variable in a cluster analysis actually should not be there, the clusters produced will not be what the analyst wanted. Because there are no simple tests for statistical significance of a variable or of goodness of fit of the whole analysis, there is no easy way for the analyst to know if the correct variables are included or if the resulting clusters are "real".

Another implication of equal variable weighting is that collinearity between variables is not removed by the procedure of the analysis (again in contrast to regression, where the collinearity between variables is split between them by the estimated coefficients). This means that if two included variables are highly correlated with each other, it is almost like double counting the variable. This may be acceptable if the analyst wants the collinearity to be counted twice, but it is important for the analyst to be aware that this is what is happening.

Everitt (1993) writes, "the use of cluster analysis in practice does not involve simply the application of one particular technique to the data under investigation, but rather necessitates a series of steps each of which may be dependent on the results of the preceding one… There is no optimal strategy for either applying clustering or evaluating results." (p. 141) This being the case, the following is a description of the

Table 2.7: Summary Statistics for Census Tract Variables

Variable	Mean	Stand.Dev.	Minimum	Maximum
Median Income	$ 41,442	$ 19,605	$ 2,499	$ 188,697
Median House Value	$ 236,425	$ 127,512	$ 9,999	$ 1,000,001
Rent Per Room[a]	$ 177	$ 90	$ 35	$ 903
Percent Homeowners	35%	24%	0%	100%
Percent White	36%	33%	0%	100%
Subway Lines[b]	1.4	1.4	0	7
Cars Per Housing Unit	0.72	0.41	0	2.85
Population Density	48,443	35,858	4	227,122
Employment Density	14,739	42,993	117	527,687
Service Density[c]	592	1,194	12	9,202

a. This is actually median rent divided by median number of rooms.
b. This is the number of subway lines available within 0.5 miles of the center of the census tract.
c. This is the number of establishments in the census tract that are classified as service or retail according to the NAIC code system.

key elements of the series of steps that produced the neighborhood clusters used in this dissertation.

In cluster analysis, the two main decisions that are made by the analyst are the number of clusters to create and which variables to include in the analysis. The approach taken here to neighborhood determination is to perform a few cluster analyses using different combinations of variables and numbers of clusters, and choose the one that best represents real neighborhoods in the areas of New York City that I know well.

The variables used in various combinations to create the neighborhoods come from the US Decennial Census 2000 and from the 1997 US Business Patterns Census. Summary statistics for these variables are given in Table 2.7, and a table of correlations between these variables is given in Table 2.8.

Before the cluster analysis procedure, a principal components analysis to prepare the data is conducted. This is a technique to create variables that embody the essence of the information in your original variables, but remove the collinearity between them. Principal components analysis is easy to picture graphically. Consider the two variable case. The correlation between these variables can be summarized in

Table 2.8: Correlations Between Census Tract Variables

	inc	house	rent	own	white	sub	cars	pop	emp	serv
inc	1.00	.	.	.	.	.	.	.	.	.
house	0.46	1.00	.	.	.	.	.	.	.	.
rent	0.42	0.33	1.00	.	.	.	.	.	.	.
own	0.62	0.05	-0.17	1.00	.	.	.	.	.	.
white	0.53	0.40	0.32	0.33	1.00	.	.	.	.	.
sub	-0.04	0.23	0.43	-0.44	0.00	1.00	.	.	.	.
cars	0.48	-0.02	-0.27	0.85	0.31	-0.56	1.00	.	.	.
pop	-0.22	0.06	0.25	-0.57	-0.15	0.36	-0.62	1.00	.	.
emp	0.29	0.33	0.66	-0.13	0.21	0.40	-0.28	0.15	1.00	.
serv	0.37	0.44	0.71	-0.14	0.28	0.45	-0.34	0.28	0.92	1.00

Note: The abbreviated variable names in this table are in the same order as the more complete variable descriptions in Table 2.7.

a regression line that best fits the data. The first principal component represents the information in this regression line. There will still be some residual variation around this line. The second principal component is found by fitting a regression line to this residual variation. Extrapolating to the multivariate case, this continues until all of the original variation has been represented by principal component variables.

For each geographic area of New York City, the software package Stata 8.1 is used to perform separate principal component analyses. The eigenvalue of 1.0 is used as the cutoff for using the principal components. This cutoff value is chosen such that every principal component contains at least as much information as one of the original variables. With this criterion and using the 10 variables listed in Table 2.7, Manhattan is represented by 3 principal components that contain 74% of the information in the original variables, the Bronx by 2 containing 58% of the original information, Brooklyn and Queens by 2 containing 62% of the original information, and Staten Island by 3 containing 73% of the original information. These principal components are rotated using a varimax rotation before using them in the cluster analysis.

The last step to prepare the data for cluster analysis is to standardize all the variables. All cluster analysis techniques use some form of a multivariate distance measure to form the clusters. In both types of cluster analysis employed in this dissertation, this distance measure is the Euclidean distance between items in the dataset.

$$\sum_{k=1}^{p}(x_{ik} - x_{jk})^2 = Euclidean\ Distance$$

where p represents the number of variables.

Euclidean distance formulas are scale-sensitive - results of a cluster analysis will be different depending on whether a distance variable is measured in feet or miles! For this reason, it is critical that all variables are standardized before the cluster analysis is performed.

To decide on the number of clusters to create, a hierarchical cluster analysis is used along with my own knowledge of New York City. Hierarchical cluster analysis techniques produce a series of partitions of the data, where the largest partition includes all of the observations and the smallest has only one observation in group. The specific technique used here is called complete linkage cluster analysis, and is an agglomerative method. In the first partition, each observation has its own group. The observations are then lumped together iteratively according to the complete linkage algorithm. Each iteration produces another partition of the data until finally all of the observations are in the same group. For an illustrative example of agglomerative clustering methods, see Everitt (1993), pp. 58-59.

Two tools exist to use this analysis to determine how many natural groups there are in a data set: the dendrogram and the Calinski and Harabasz statistic (or other similar statistics). The dendrogram is a graphical representation of the distance between data partitions. If the analyst is fortunate, the dendrogram pattern will clearly dictate the number of natural clusters in the data. In the case of the data used in this dissertation, however, the dendrogram does not provide great insight. Fortunately, a number of researchers have suggested numerical methods to find the number of natural clusters in the data, and the method of Calinski and Harabasz is used here:

$$CH(k) = \frac{B(k)/(k-1)}{W(k)/(n-k)}$$

where $B(k)$ and $W(k)$ indicate the between and within cluster sums of squares. This value is calculated at different numbers of clusters. The "correct" number of clusters for the data is that which maximizes the Calinski Harabasz statistic.

Once the number of natural clusters in the data is identified, the next step is to try different combinations of variables in an optimization cluster analysis, and then choose the results of one of these to be the neighborhoods. Optimization cluster analysis takes as given the number of clusters to create, and groups data into those groups that minimize within-group variance and maximize between-group variance. Calculation of within-group variance is straightforward, but calculation of between-group variance is less so because the analyst needs to choose a single point to represent the group. K-means optimization cluster analysis is one of the most common forms, using the group mean of each variable to represent the group. It is used here.

Stata 8.1 is used to calculate 4 k-means cluster analyses with different included variables. Since Stata does not have information about which census tracts are neighbors, enforcing spatial contiguity is a challenge. There are two things done here that aim to encourage cluster spatial contiguity.[4] The first is to calculate clusters for each of the four water-separated geographic areas of New York City separately. This means that 16 cluster analyses were actually performed: 4 for Manhattan, 4 for Staten Island, 4 for the Bronx, and 4 for Queens and Brooklyn together (see the map that is Figure 2.4). The second strategy employed here to encourage spatial contiguity is to include the x,y coordinates of each census tract as variables in the cluster analysis. In fact, for each analysis, the x,y coordinates were double-weighted. These strategies together do a reasonably good job of creating spatially contiguous clusters, but the results on this front are not perfect. The final neighborhood clusters that you see in Figure 2.4 have been somewhat manually adjusted so that they are contiguous groups of census tracts.

Two of the cluster analyses are done with only one extra variable (on top of the

[4]Some time was spent searching for a way to enforce rather than simply encourage spatial contiguity. There appears to be a software package that has this capability (CLUSTAN), but I do not currently have access to this tool.

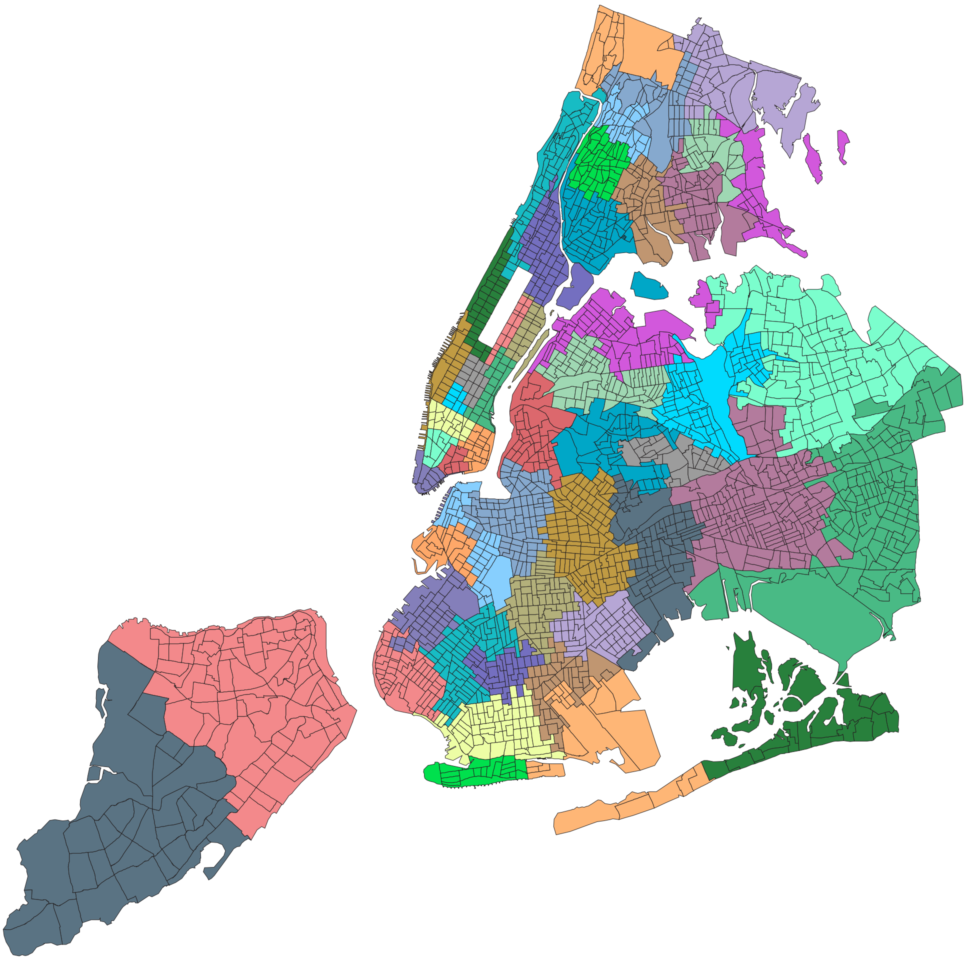

Figure 2.4: Fifty-one New York City Neighborhoods Identified Through Cluster Analysis

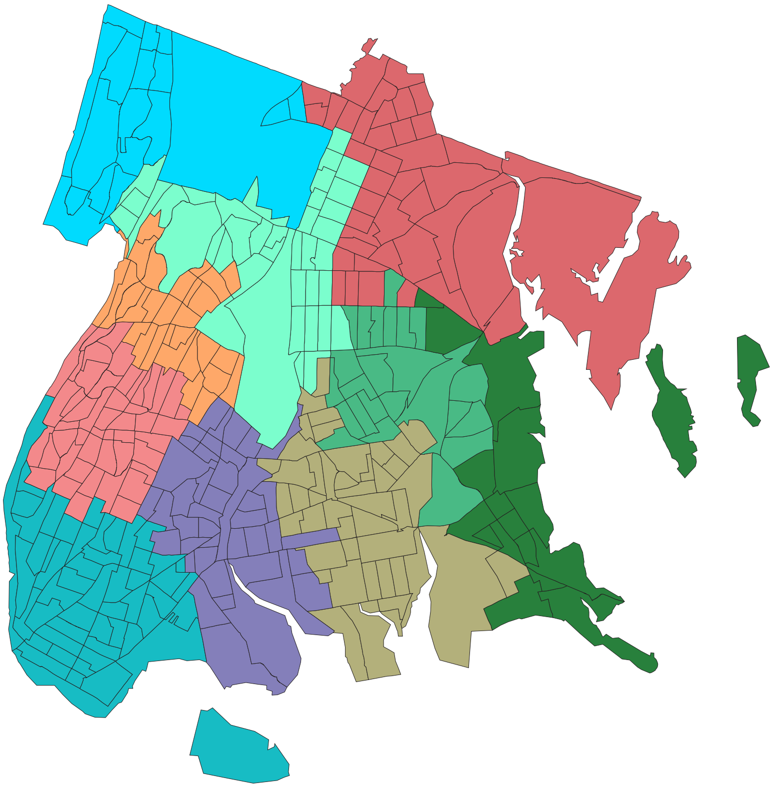

Figure 2.5: Ten Neighborhoods of the Bronx as Identified through Cluster Analysis

x,y coordinates), one is done using principal components, and one is done using the full set of original relevant variables. To represent New York neighborhoods in the discrete choice models in this dissertation, the set of cluster analyses that use principal components as the relevant variables and double weight on the x,y coordinates to encourage cluster spatial contiguity are used. These were chosen because the areas of New York that I know well appear to be represented accurately by these neighborhood clusters.

Focusing on one borough - the Bronx - provides an example of cluster analysis results for a smaller area. The Bronx is made up of 10 neighborhoods, all of which are close to spatially contiguous (see Figure 2.5). Table 2.9 reports both the within-neighborhood and between-neighborhood variation for census tract median income. It turns out that there is one wealthy census tract in the Bronx (in neighborhood 4) that causes high variation within that neighborhood. Other than this outlier, however, the cluster analysis has done its job - the within-neighborhood variations

Table 2.9: Bronx Census Tract Median Income: Within- Versus Between- Neighborhood Variation

Within-Neighborhood Variation				
Neighborhood	Mean	Stand.Dev.	Minimum	Maximum
1	$ 20,857	$ 5,161	$ 12,304	$ 37,950
2	$ 30,384	$ 8,271	$ 2,499	$ 47,647
3	$ 45,150	$ 10,059	$ 20,417	$ 68,542
4	$ 60,018	$ 29,866	$ 9,625	$ 168,061
5	$ 17,016	$ 4,362	$ 9,821	$ 31,767
6	$ 21,774	$ 5,439	$ 10,825	$ 31,832
7	$ 18,421	$ 6,880	$ 7,044	$ 36,316
8	$ 44,588	$ 10,895	$ 30,750	$ 80,488
9	$ 52,309	$ 8,247	$ 33,438	$ 66,250
10	$ 30,854	$ 5,530	$ 13,250	$ 41,147
Between-Neighborhood Variation				
Bronx	$ 34,137	$ 15,364	$ 18,421	$ 60,018

for median income in all of the other Bronx neighborhoods are substantially lower than the between-neighborhood variation in the Bronx.

2.5 Chapter Summary

This chapter has described the modeling methodology used in the present dissertation, identified data sources, and explained how the data has been organized and manipulated to be ready for the next step: statistical modeling to explore the questions posed in this dissertation.

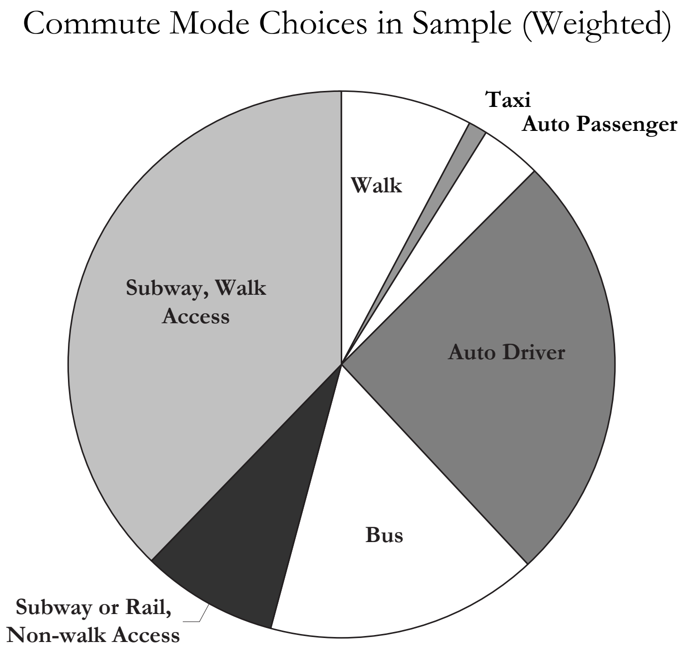

Figure 2.6: Note that for commuting, the transit mode options in the sample are much more popular than the auto modes.

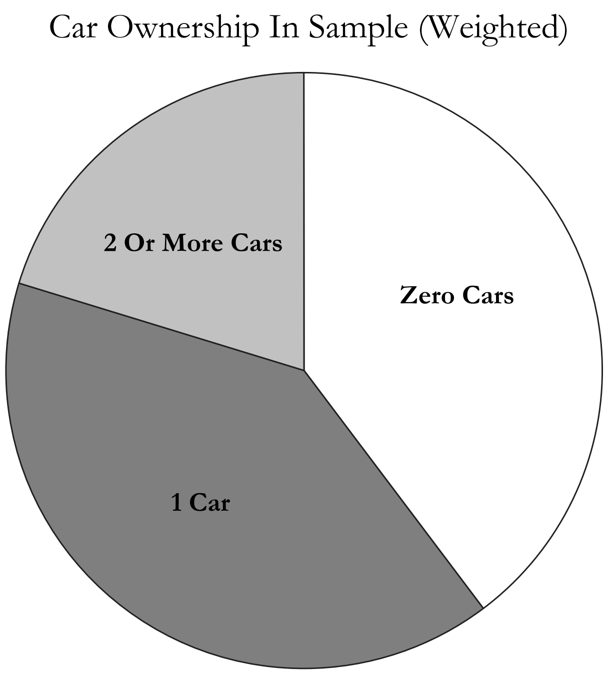

Figure 2.7: Note that car-free households comprise almost half of the commuting sample.

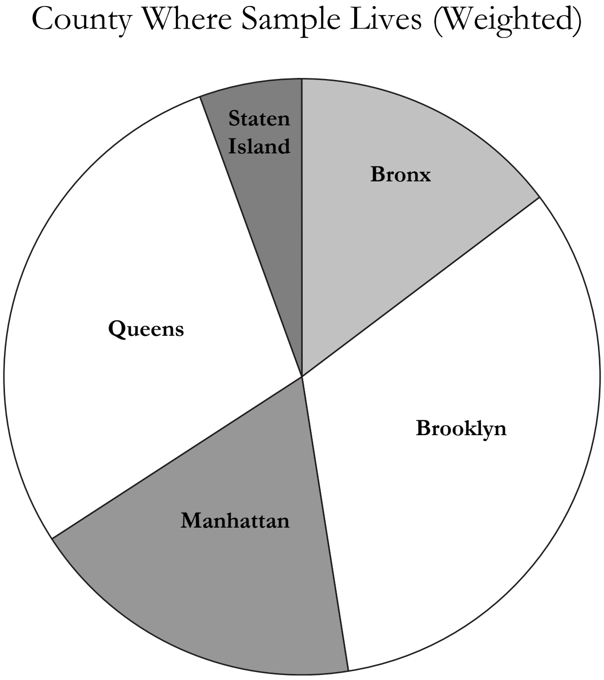

Figure 2.8: Most of the commuting households live in Manhattan.

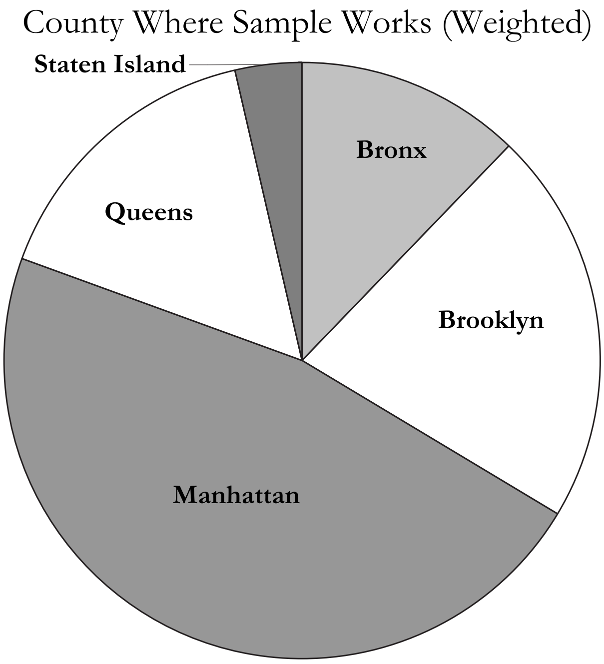

Figure 2.9: Even a higher percentage of commuters in the sample work in Manhattan.

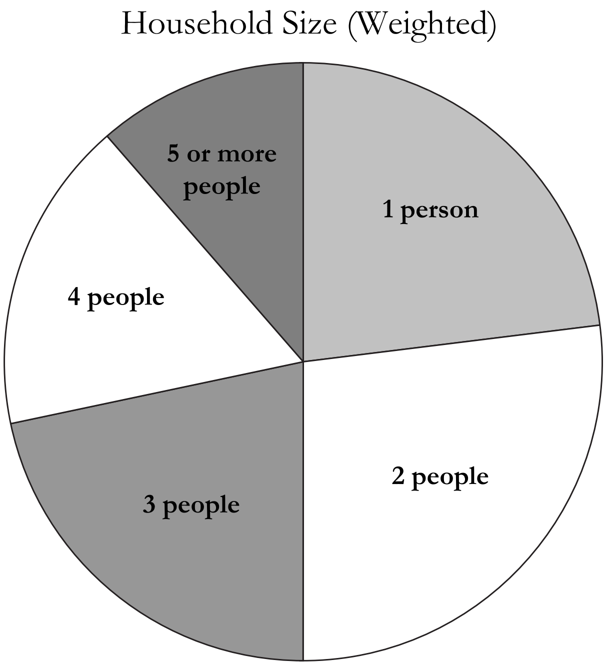

Figure 2.10: Most households in the commuting sample are small, with only 1 or 2 people.

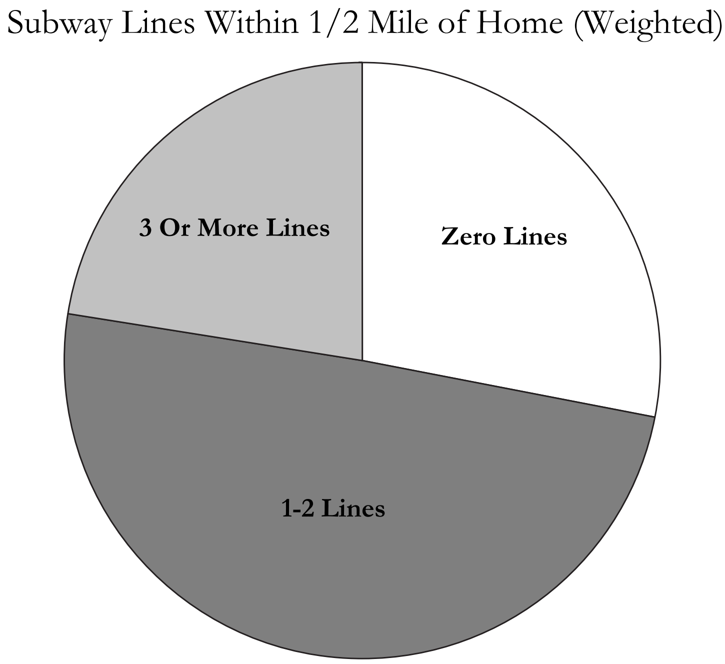

Figure 2.11: Household level of home subway access is evenly split in the commuting sample.

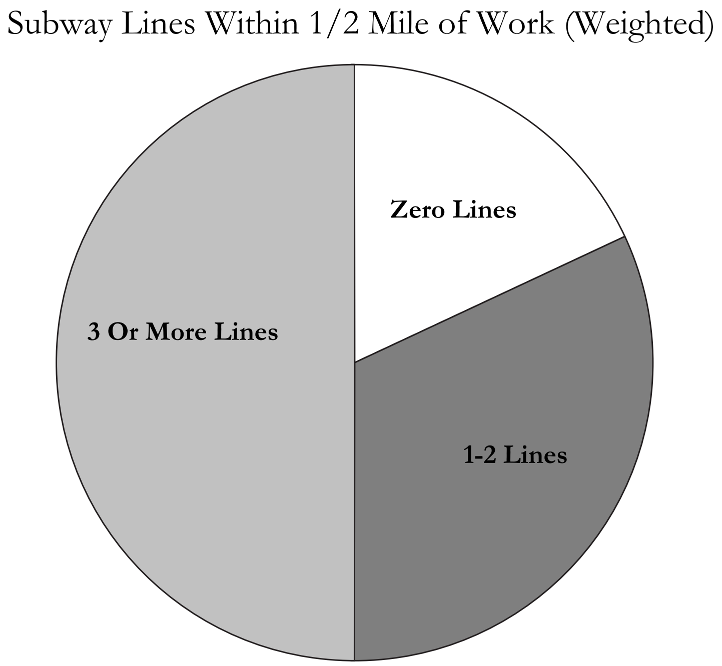

Figure 2.12: Subway access near work locations in the commuting sample is better than it is near residences.

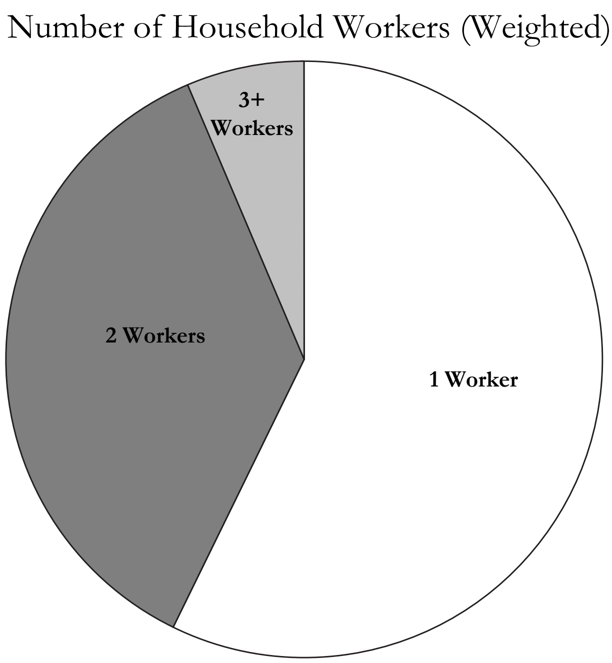

Figure 2.13: More than half of the households have only one worker, but a sizable percent have two.

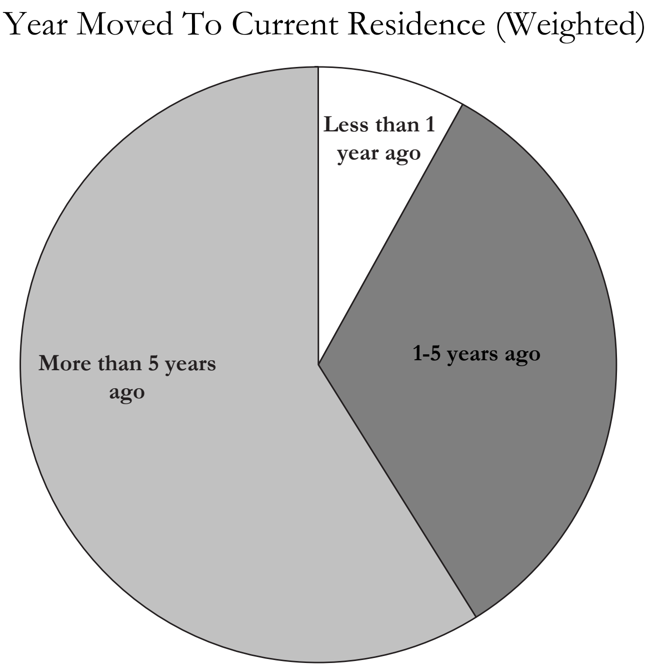

Figure 2.14: Most of the commuting sample has lived in their neighborhood for a number of years.

Figure 2.15: For commuters in the estimation sample who took a taxi to work, this figure shows the distribution of the amount of time spent in the taxi. This distribution is based on a total of 53 commuters with a mean of 18 minutes and a standard deviation of 12 minutes.

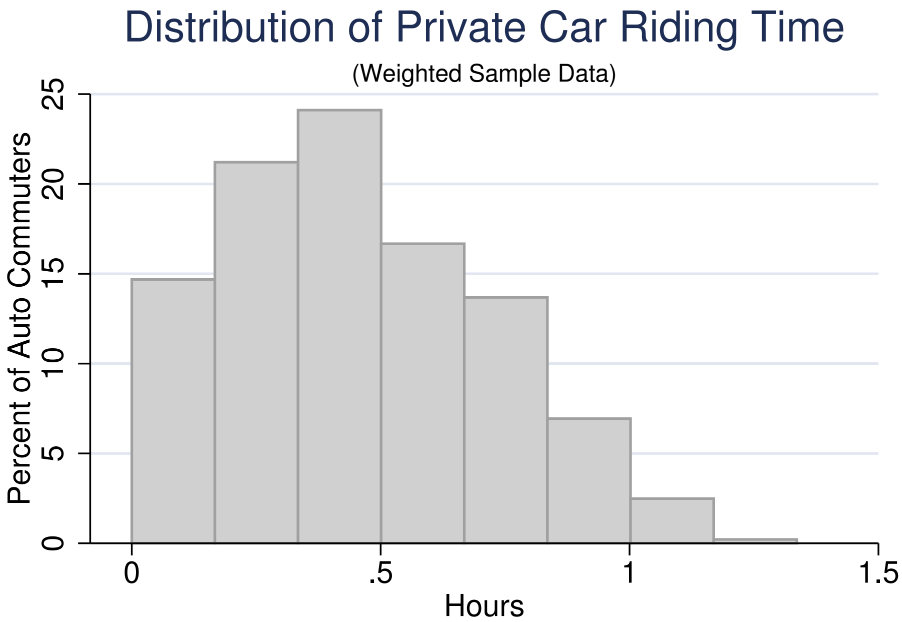

Figure 2.16: For commuters in the estimation sample who drove themselves or got a ride in a private car to work, this figure shows the distribution of their travel time. This distribution is based on a total of 786 commuters, with a mean of 27 minutes and a standard deviation of 16 minutes.

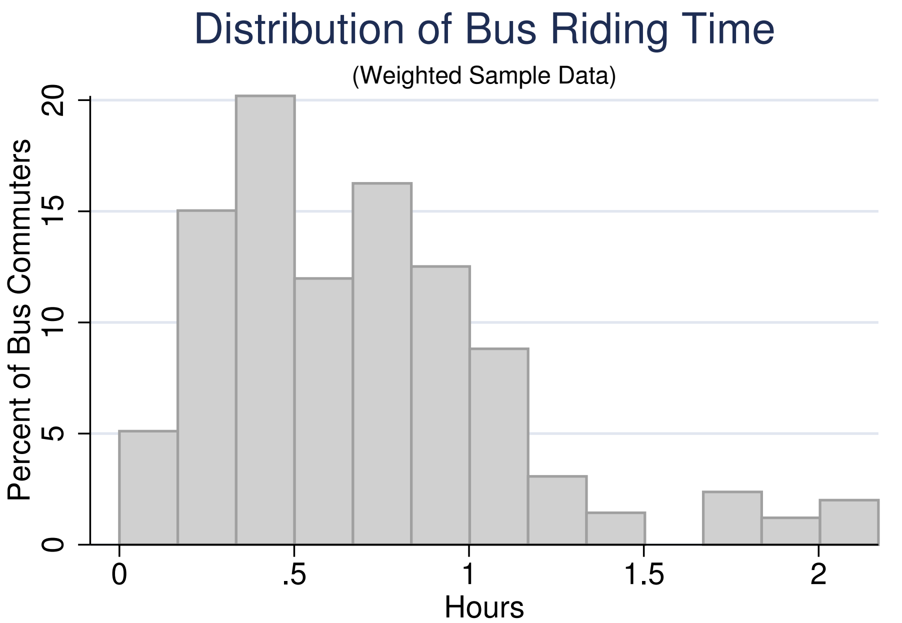

Figure 2.17: For commuters in the estimation sample who took the bus to work, this figure shows the distribution of their time riding the bus. This distribution is based on a total of 451 commuters, with a mean of 41 minutes and a standard deviation of 26s minutes.

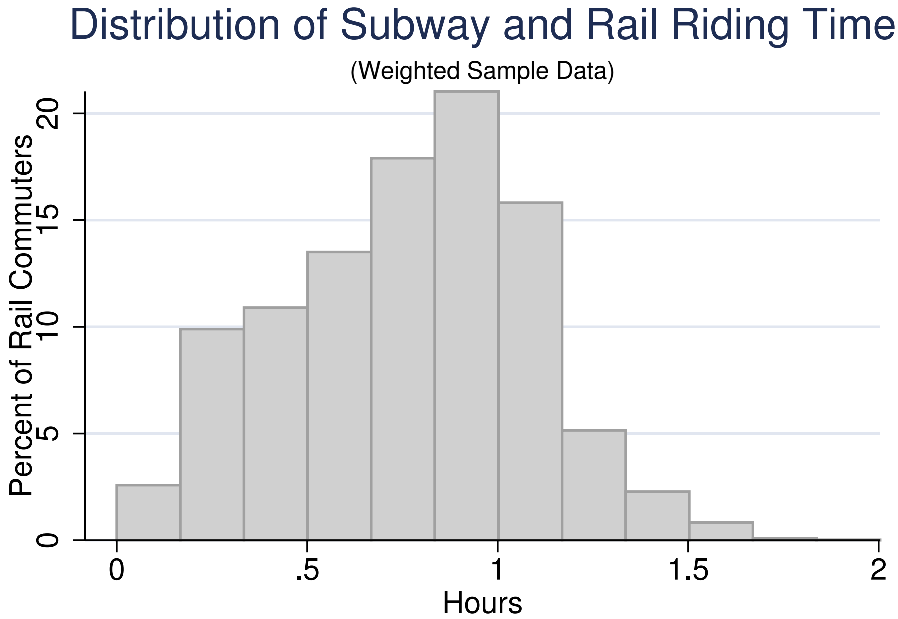

Figure 2.18: For commuters in the estimation sample who took the subway or commuter rail to work, this figure shows the distribution of their time riding the train. This distribution is based on a total of 1132 commuters, with a mean of 46 minutes and a standard deviation of 19 minutes.

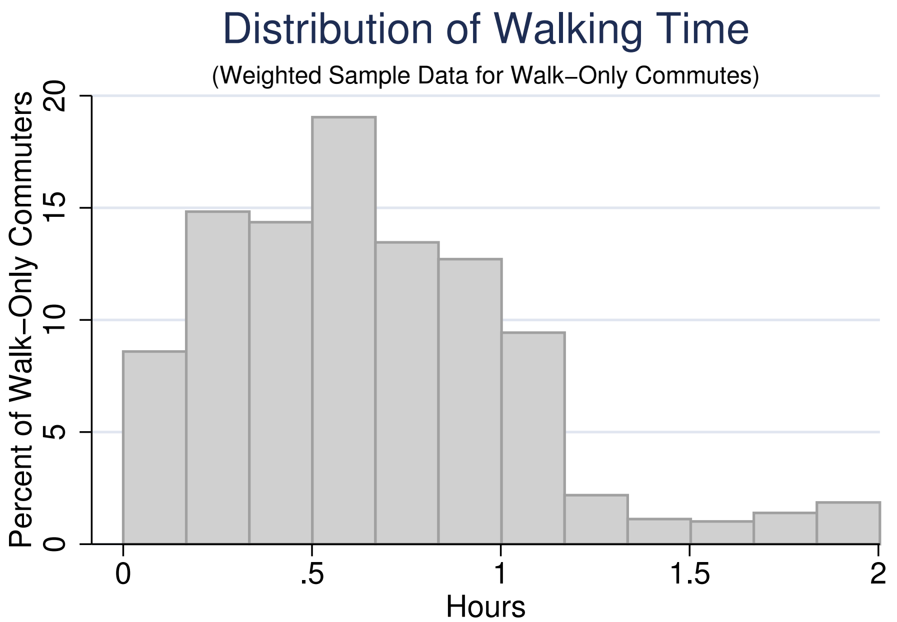

Figure 2.19: For commuters in the estimation sample who walked to work, this figure shows the distribution of their travel time. This distribution is based on a total of 306 commuters, with a mean of 39 minutes and a standard deviation of 24 minutes.

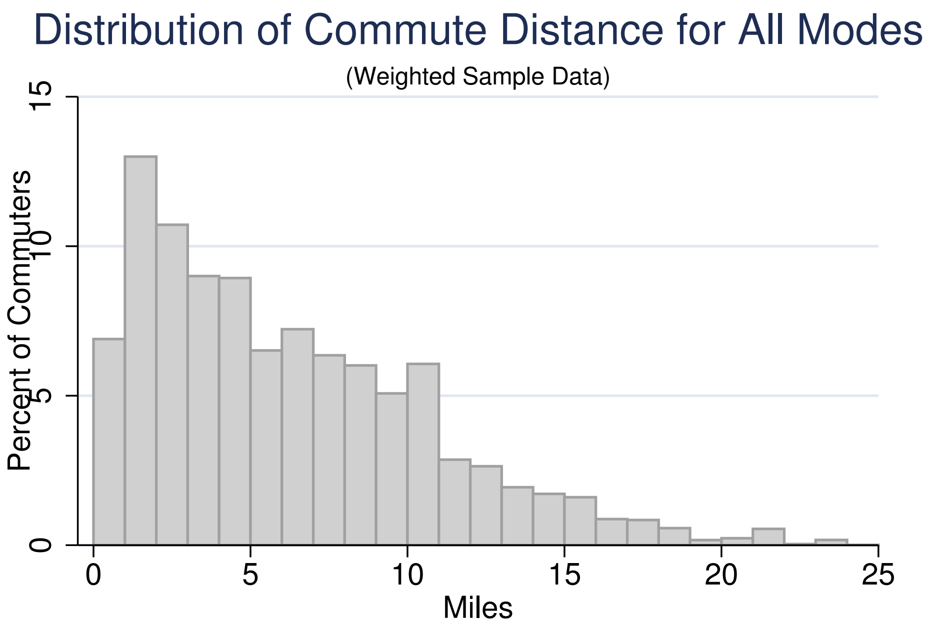

Figure 2.20: This distribution of commute distance in the estimation sample for all modes is based on a total of 2728 commuters, with a mean of 6.2 miles and a standard deviation of 4.5 miles.

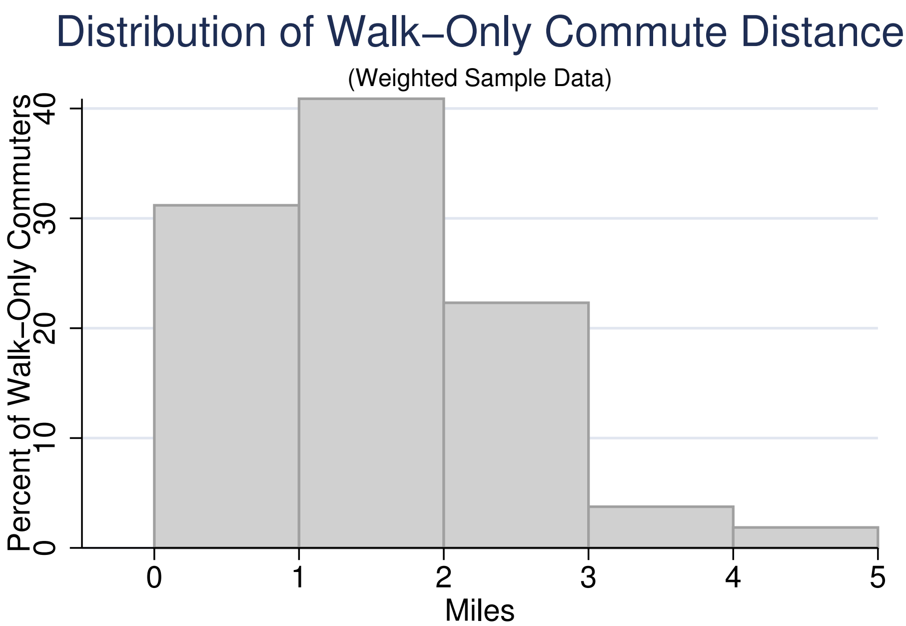

Figure 2.21: This distribution of commute distance for people who walk to work is based on a total of 306 commuters, with a mean of 1.5 miles and a standard deviation of 0.9 miles.

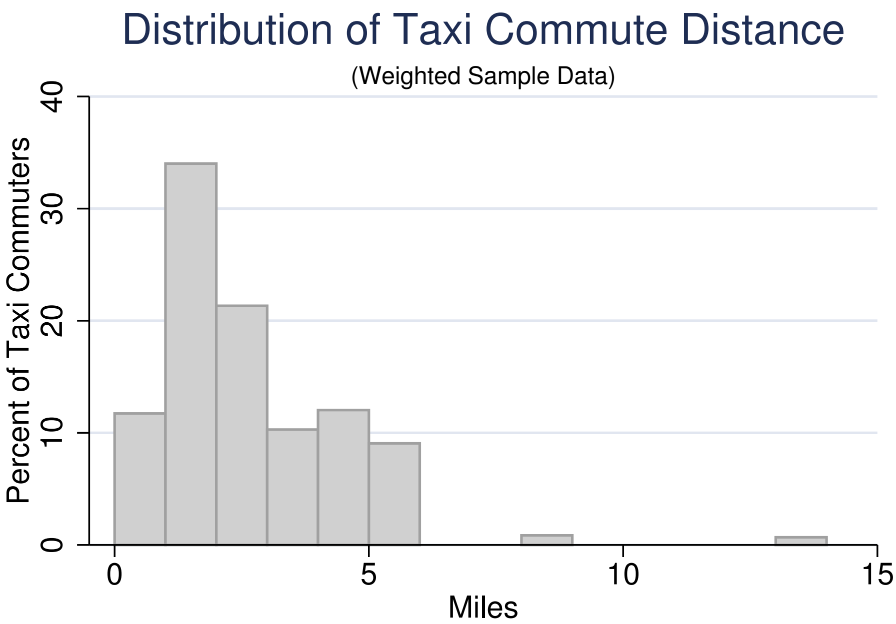

Figure 2.22: This distribution of commute distance by taxi is based on a total of 53 commuters, with a mean of 2.7 miles and a standard deviation of 1.9 miles.

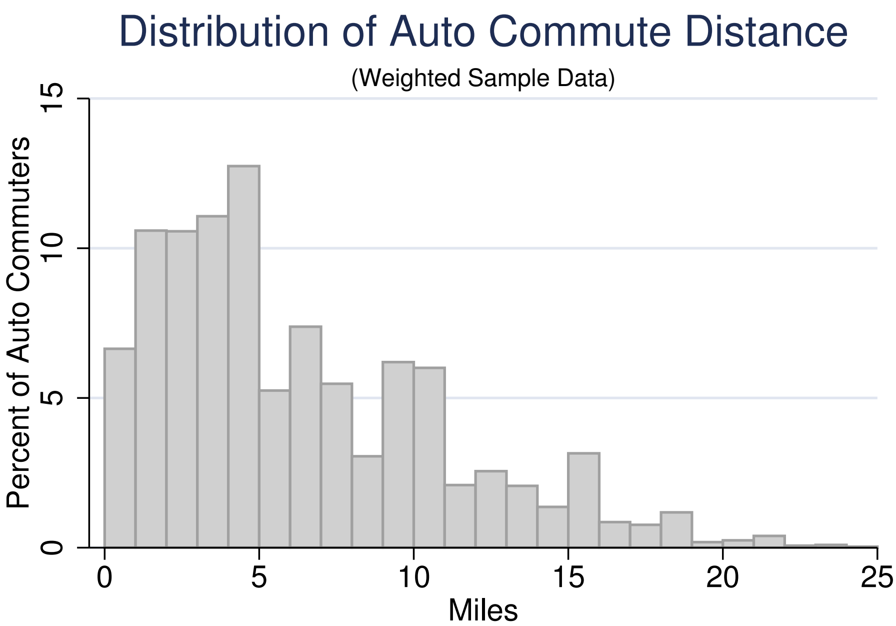

Figure 2.23: This distribution of commute distance by auto is based on a total of 786 commuters, with a mean of 6.3 miles and a standard deviation of 4.6 miles.

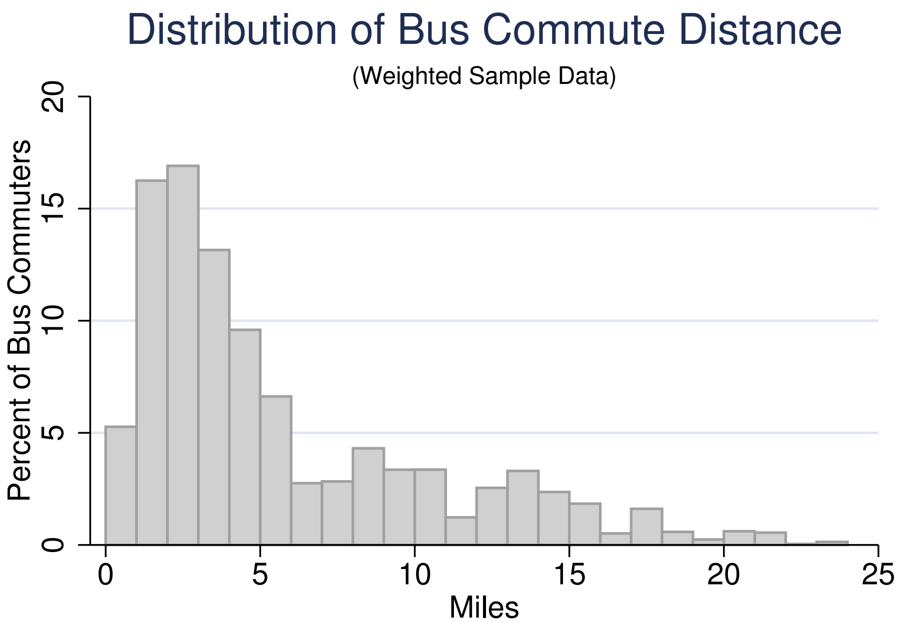

Figure 2.24: This distribution of bus commute distance is based on a total of 451 commuters, with a mean of 5.7 miles and a standard deviation 4.8 miles.

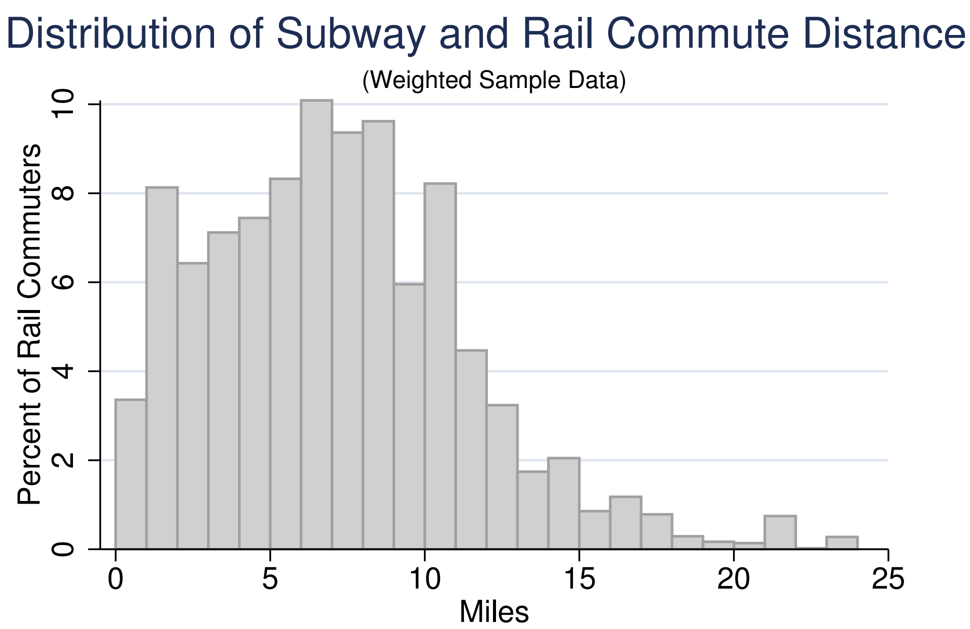

Figure 2.25: This distribution of rail commute distances is based on a total of 1132 commuters, with a mean of 7.2 miles and a standard deviation of 4.2 miles.

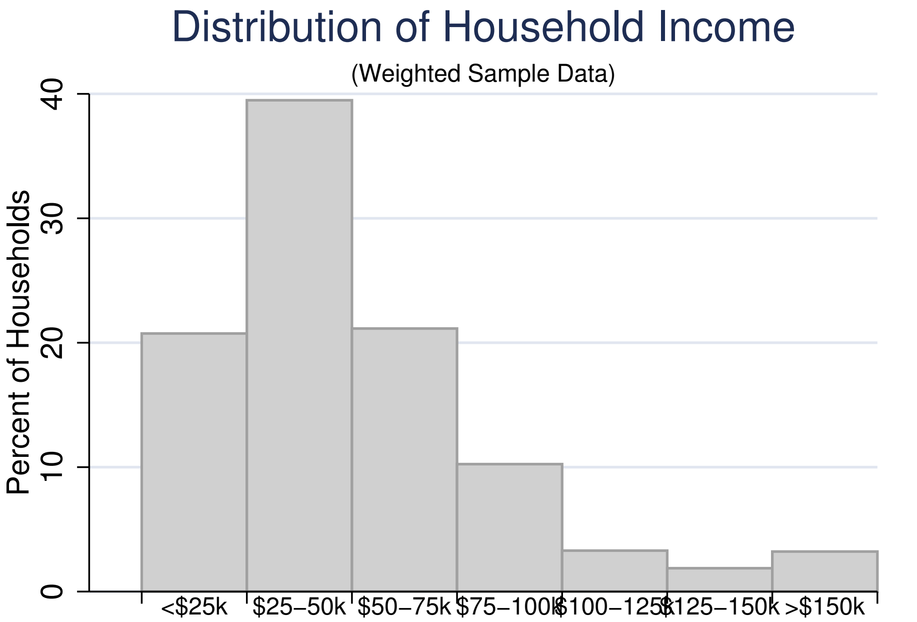

Figure 2.26: This household income distribution is based on the estimation sample of 2013 households with commuters, and has a mean of $51,500 and a standard deviation of $33,500.

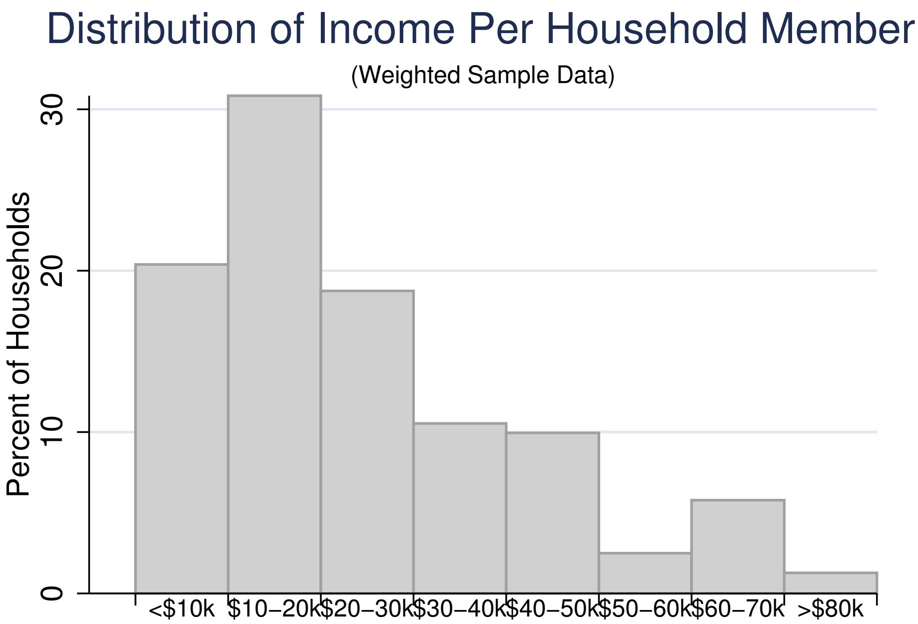

Figure 2.27: This household income distribution is normalized by household size. It is based on the estimation sample of 2013 households with commuters, and has a mean of $25,000 and a standard deviation of $22,500.

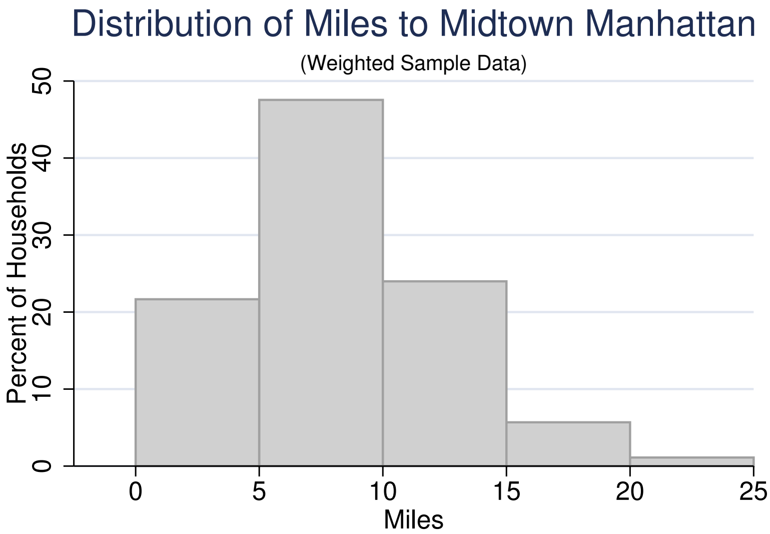

Figure 2.28: This figure shows the distribution of distance on the street network from home to midtown Manhattan for the estimation sample of commuter households. Since I do not have exact address data for the sample households, this variable was estimated by calculating the distance from the center of each neighborhood to midtown Manhattan. This distribution is based on a sample of 2013 households, and has a mean of 8.4 miles and a standard deviation of 4.2 miles.

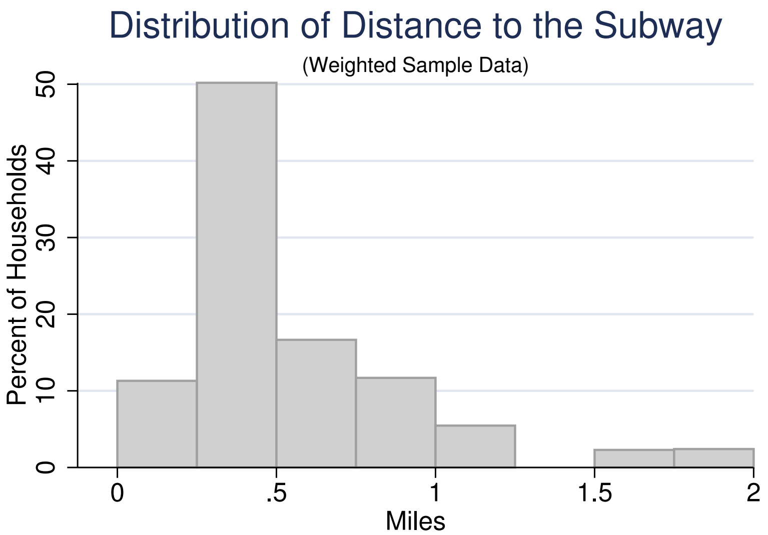

Figure 2.29: This figure shows the distribution of the distance along the street network from home to the nearest subway station. Since I do not have exact address data for the sample households, this variable was estimated by taking the average of the distances from a representative set of locations in each neighborhood. This distribution is based on a sample of 2013 households, and has a mean of 0.7 miles and a standard deviation of 0.7 miles.

Chapter 3

Cars and the City: A Model of the Determinants of Auto Ownership and Use For Commuting in New York City with Endogenous Choice of Residential Location

Heavy reliance on the private automobile for urban transportation causes substantial externalities, the most prominent being traffic congestion, air pollution, and, many would argue, loss of a sense of community. Travelers do not pay for the delay cost they impose on other users of a congested roadway. Likewise, they do not pay for the effects on others of the degradation of the air quality that their vehicles' emissions cause or the loss of a sense of community in their neighborhoods. As is the case with all activities that cause negative externalities, both car ownership and use levels are likely to be higher than would be socially optimal.

Recognizing this, urban planners and economists have repeatedly suggested investments and policies that encourage the use of alternatives to the private automobile for urban transportation. Cities both in the United States and around the world are trying out a multitude of transportation policy and investment alternatives with the aim of reducing car-induced externalities. However, without a solid understanding of how urban residents make their transportation and residential location choices, it is hard to tell which of these policies and investments are really doing the job and which are wasting precious city resources.

This chapter addresses the following question: What are the most effective policy levers to control car ownership and use in dense urban areas? To get at this question, this chapter explicitly models the choice to own zero, one, or two-or-more cars in the context of the related decisions of where to live and how to get to work, using the statistical framework of discrete choice econometrics. This model purposely incorporates as many variables that have clear policy relevance as possible, as well as individual characteristics of travelers as control variables. Although related work has been done, the present analysis is unique in that it focuses on both car ownership and car use while also endogenizing residential neighborhood choice. This is important because, as discussed in Chapter 1 of this dissertation, the choice of where to live is fundamentally linked to the choices of whether to own and use a car. Analyses that do not explicitly model the joint nature of these decisions may produce biased results. The only previous research known to me that jointly models the three decisions modeled here was published in 1977 (Lerman).

A second (and perhaps more important) unique aspect of this work is that it makes use of an unusually rich dataset from New York City. New York City is unusual among US cities in that it has substantial variation within the city in the availability of transportation alternatives, residential neighborhood characteristics such as density and employment opportunities, and therefore car ownership and use choices among its residents. According to the 2000 Census, only 44% of New York City households own cars, the next lowest major US city in car ownership is Washington, DC where 63% of households own cars (U.S. Census Bureau, 2000). The high variation in transportation choices made by New Yorkers allows for a more robust statistical estimation, and examination of the results for subpopulations within New York that are more urban or more suburban allows for potential extrapolation of the current results to other locations.

The remainder of this chapter is organized as follows. Section 3.1 presents a review of the existing literature on car ownership and use. Section 3.2 presents the methodology used in the estimations in this chapter. This section relies heavily on the background statistical and economic theory reviewed in Chapter 2 of this dissertation. Section 3.3 introduces the estimated models and highlights specific policy-relevant findings, and finally Section 3.4 concludes with suggestions for future research directions in this area.

3.1 Existing Literature on Car Ownership and Use

Much of the research on car ownership in the US focuses on the decision of *which* vehicle to purchase/own, rather than the decision of *whether* to own a vehicle (e.g. Manski and Sherman, 1980; Mannering and Winston, 1985; Goldberg, 1995). In most of the US, this is a sensible approach, since almost every household owns at least one vehicle. The present analysis focuses on the latter question, adding to the relatively sparse literature in this area.

Modeling the "whether" of car ownership is a difficult task. Because cars are durable goods, car ownership is a complex decision requiring the consumer to dynamically optimize by comparing the expected utility from life as a car owner to that of life as a non-owner. A large number of variables come into play in this decision process, most of them somehow related to either income or the relative "prices" of transportation alternatives, where "prices" refer to not only money prices, but also time "prices", comfort "prices", convenience "prices", etc.

Some studies based on geographically aggregated data rely almost entirely on income to explain car ownership levels (e.g. Ingram and Liu, 1999; Dargay and Gately, 1999), largely because the aggregation in their data dilutes the explanatory power of other variables. While these models forecast aggregate car ownership reasonably well, they offer little ability to evaluate policies aimed at redirecting existing trends.

For policy analysis, it is necessary to include in the model both the time and money prices of substitutes (i.e. transit) and complements (i.e. parking services) for cars as well as urban land use characteristics that are highly relevant to determining car ownership levels in cities. It is the studies such as the present dissertation that rely on spatially disaggregate data that have a better chance of shedding light on these effects. A few such studies are briefly reviewed here.

De Jong (1990) postulates and empirically estimates a model of the demand for cars and vehicle kilometers traveled by zero- and single-car households in the Netherlands using household survey data from 1985. In his model, car costs are divided into fixed and variable components, and utility is maximized subject to a budget constraint in which available income has been reduced by the fixed cost of car ownership. This utility is compared to that which could be achieved as a non-owner, and the higher utility indicates the chosen status. Schimek (1996) uses a two-stage procedure to estimate jointly the demand for vehicles and the demand for vehicle kilometers traveled using 1990 household survey data from across the US. In his analysis, Schimek focuses on the effect of population density on car ownership and vehicle kilometers traveled, and finds only a small effect.

While these models are based on disaggregate data, and therefore could potentially estimate the effects on car ownership and use of policy-sensitive variables, the authors do not include many such variables in their models. Schimek includes the policy-sensitive variables of transit availability and population density, while de Jong includes only household-specific variables that cannot be easily changed by policy.

Train (1980) estimated a nested logit model of the choices of car ownership and commute mode using 1975 data from the San Francisco Bay Area. In the creation of the model presented in this paper, direction was taken from Train. There are, however, a number of differences between Train's work and that presented here. These include contextual differences such as the year, the different physical contexts of the two cities, and the difference between the car ownership and use levels in the two datasets. In Train's data, 93 percent of surveyed households owned a car, and 81 percent used a private car for their commute trip. The corresponding values for the 1997-98 data from New York City are 61 and 30 percent, respectively. The other large difference between Train's work and the present paper is in the complexity of the model itself. Train used a nested logit structure to model two interdependent choices, while the model estimated in this paper is a joint choice logit model of three interdependent choices.

Lerman (1977) produced an impressive early attempt at a joint model of the choices of housing type, residential location, car ownership, and commute mode. He used data from Washington, D.C. from 1968, and his main focus was on the residential location decision. The present work takes direction from Lerman. Unfortunately, Lerman did not report elasticities, and therefore the results presented here cannot be directly compared to his.

3.2 Methodology

The model at the heart of this chapter is a multinomial logit model of the joint choice of residential neighborhood, car ownership status, and commute transport mode. These three sub-choices are fundamentally interrelated in the following way. In a world without transaction costs, one can imagine that these three choices would be made simultaneously. Everyone would daily choose his or her residential neighborhood as well as transport modes for each trip. Car ownership choices would be inseparable from the choice of mode, and residential location choices would be some compromise between household members based on all the locations they needed to go to on that day.

In the real world, two of these three decisions entail large transaction costs associated with change. Both changing one's residential neighborhood and changing one's car ownership status are highly costly activities in terms of both time and money. Because of these high transaction costs, many researchers have modeled commute mode choice as if residential neighborhood and car ownership status were exogenous variables. This approach may often yield reasonable results. However, it does not allow the researcher to test hypotheses about the effects on mode choice of variables that theoretically are related to car ownership or residential neighborhood choice.

This chapter – along with Appendix A of this dissertation – present the results from 12 discrete choice models of these decisions, all estimated using the same data set. The most general of the models presented in this chapter treat all choices as endogenous - the choices of commute mode, car ownership status, and residential neighborhood are modeled as a single joint choice. This structure represents a world in which each commuter has dynamically optimized his or her choices so that the set of three choices is chosen to yield the highest utility. This means that each commuter, in considering both where to live and whether to own a car, has taken into account the full discounted time and money cost of commuting for the expected duration of residence and car ownership that will results from that decision. It is the results of this joint model that will be the primary focus of this chapter.

Recall from Chaper 1 that part of the objective of this dissertation is to identify how related these three decisions are, and also to estimate how far off the policy-relevant implications will be for models that do not make all of these decisions endogenous. As such, it is relevant to estimate all possible sub-choice models, and compare their results with those from the most general models. These model comparisons will foster a better understanding of the biases that might be present in models that do not endogenize all three of these choices. Therefore, in addition to the three-choice model described above, this chapter also discusses estimation results for the following model specifications:

- Model of the Joint Choice of Residential Location and Car Ownership Status
- Two Nested Models of the Choices of Residential Location and Car Ownership Status (one with each choice as the upper level of the model)
- Model of the Joint Choice of Residential Location and Commute Mode
- Two Nested Models of the Choices of Residential Location and Commute Mode (one with each choice as the upper level of the model)
- Model of the Joint Choice of Car Ownership Status and Commute Mode
- Two Nested Models of the Choices of Car Ownership Status and Commute Mode (one with each choice as the upper level of the model)
- Model of the Choice of Car Ownership Status
- Model of the Choice of Commute Mode

The full estimation and calculated elasticity results for these models are given in Appendix A.

Most of the multinomial logit models described above and presented in this chapter (see Chapter 2 of this dissertation for a review of multinomial logit model theory) have compound choice sets. This means that each alternative in the choice set is composed of more than one sub-choice alternative. In the most general of models, each element of the compound choice set contains one mode alternative, one car ownership status alternative, and one residential location alternative as defined by a census tract. For example, one alternative is walk to work, own zero cars, live in census tract 23, and a separate alternative would be walk to work, own one car, live in census tract 23. The choice set for the model estimated here has 7 commute mode alternatives, 3 car ownership status alternatives, and over 2000 residential census tract alternatives. Therefore, even though each sub-choice has a manageable number of alternatives, the compound choice set is unmanageably large with over 40,000 alternatives. As described in Chapter 2, the choice sets are reduced to be computationally manageable by taking a random sample of the residential location sub-alternatives to be the choice set in the statistical model. For each commuter, the compound choice set for the most general model includes that included all 20 feasible mode-car ownership combinations (the combination of car driver and zero-car owner was removed) and 11 possible census tract locations, making a modeled choice set of 220 compound alternatives.

3.2.1 Nested versus joint choice models

As discussed in Chapter 2, the one serious limitation of the multinomial logit model is that it assumes that the model satisfies the Independence of Irrelevant Alternatives (IIA) property. Violation of the IIA assumption is especially likely in a joint choice situation such as that modeled here. For instance, in the present application, it makes sense that there would be some correlation among the mode sub-alternatives that all have the same residential location and car ownership status. If walk to work, own one car, live in census tract 23 were removed (perhaps because the commuter sprained her ankle), it would be disproportionately likely that she would choose drive to work, own one car, live in census tract 23, rather than any alternative that would require her to change car ownership status or residential location.

For each of the model specifications presented here that endogenizes two of the sub-choices, both nested and non-nested versions are estimated, and results from the versions were compared to see which model best fits the data. In most of the nested versions of these models, the estimates of the inclusive value coefficients are largely either not significantly different from one or were substantially larger than one. In one case, the estimated inclusive value parameter was negative. The first case indicates that a multinomial logit model fits the data as well as a nested logit specification. The cases of inclusive value parameters that are either greater than one or negative may be inconsistent with utility theory. Section 3.3.5 of this dissertation will discuss in more detail the comparison of nested and non-nested models estimated here.

3.2.2 Limitations of the models

There are two additional limitations of the multinomial logit model used here that deserve mention: the assumption of homogeneous preferences across the sample and the lack of accounting for the presence of spatial autocorrelation. Removal of these limitations by use of a more sophisticated error structure is a possible area for future research. However, because the model results are quite stable across most of the specifications that are estimated in this dissertation, I do not expect the main policy implications of the results that appear in this chapter to change appreciably by this type of change to the model.

In addition to these statistical limitations, the model in this chapter is also limited by a couple of simplifications of the choice framework. For instance, multiple-worker households are not modeled differently from single-worker households, even though the relationship in a multiple-worker household between residential neighborhood choice and travel choices is likely to involve a compromise between the workers.

Table 3.1: Shares of Car Ownership and Commute Mode in Sample Used in this Chapter

	NYC	Manhattan	Other Boros	Staten Isl.
Number of Commute Trips (observations)	2728	1128	866	734
Car Ownership				
0	40%	67%	35%	6%
1	40%	28%	44%	34%
2	20%	5%	21%	59%
Mode Choice				
Walk/Bike	8%	20%	6%	2%
Taxi	1%	3%	1%	0%
Auto Passenger	4%	2%	4%	5%
Auto Driver	25%	5%	27%	57%
Bus	16%	13%	16%	23%
Subway with Walk Access	38%	49%	38%	4%
Subway with Other Access	8%	8%	8%	8%

a. These are the neighborhood-weighted shares of NYC commuting households in the estimation sample that own 0, 1, or 2-or-more cars.

I make the simplifying assumption that the choice of residential neighborhood yields the highest possible utility for all workers in the household. Another simplification made here is that although this model explicitly explains the choice of residential neighborhood, it does not also endogenize the choice of work location. There has been some work done that indicates that it may be important to endogenize work location as well (Waddell, 1993), but due to the high level of complexity of the current model, a decision was made to leave the work location decision as exogenous. Incorporating these factors into the model is another potential area for future research.

There were a few possible determinants of mode choice that were either impossible or too costly to estimate for the alternatives not chosen, and therefore had to be left out of the model. Two of these that stand out are the number of transfers for transit trips and the fact that trip-chaining is not modeled as a determinant of choice because only the home-to-work trip is modeled.

Table 3.1 summarizes the distribution of the choice of car ownership status and commute mode in the sample used for the estimation in this paper. Figures 3.1 and 3.2 represent this information graphically.

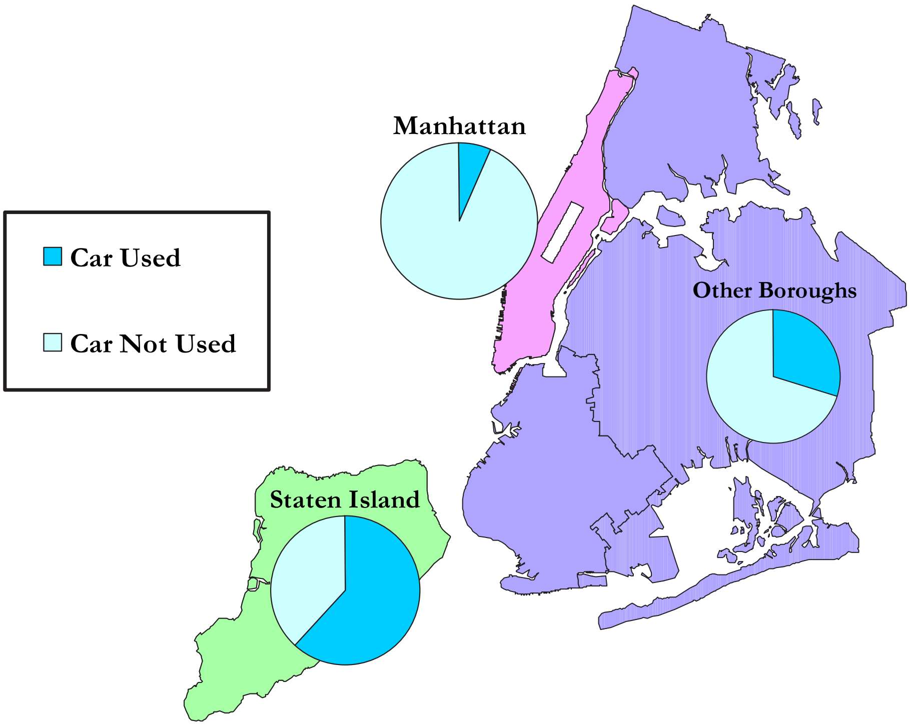

Figure 3.1: Distribution of Car Use for Commuting in Sample

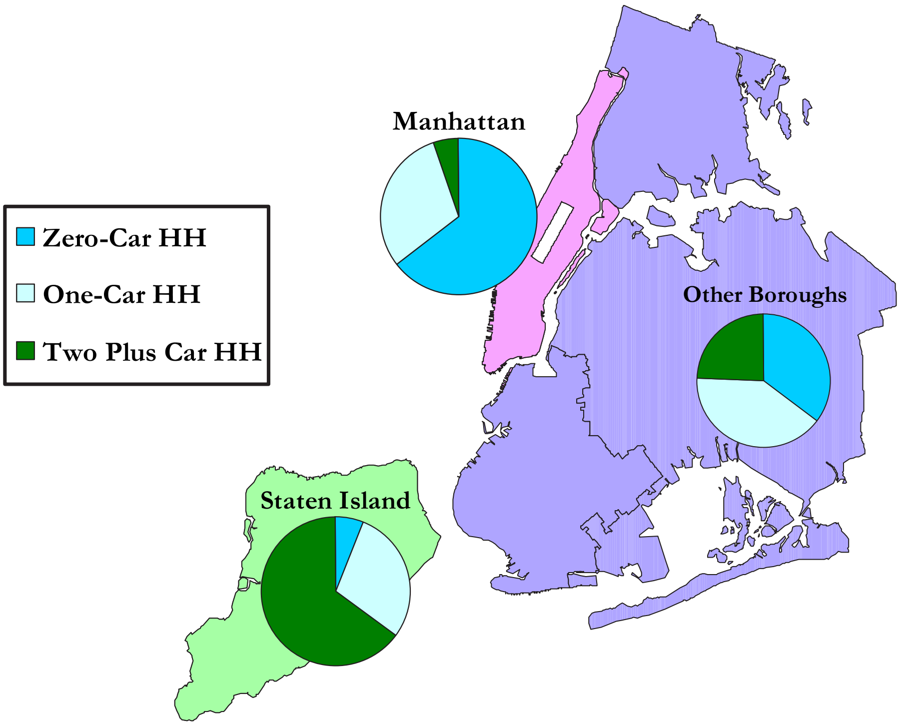

Figure 3.2: Distribution of Car Ownership in Sample

It is interesting to note that there is substantial variation in car ownership and car use (see highlighted area) even within New York City. Manhattan exhibits extremely low car use for commuting, while more than half of Staten Island commuters use cars.

3.3 Results

Table 3.3 presents the estimated coefficients for the multinomial logit model of the joint choice of residential location, car ownership status, and commute mode. Tables 3.4 and 3.5 present the elasticities that were estimated using this model. The tables in Appendix A detail the estimated coefficients and corresponding elasticity estimates of the other 11 models that were estimated for this chapter.

This section will use these findings to inform both a discussion of model selection and a discussion of the implications for policy of the selected model. First will be an overarching explanation of how the explanatory variables were chosen for the models, and how to interpret the estimated coefficients. Second will be an interpretation of the estimated coefficients of the selected model. Third will be a comparison of the elasticity results from the selected model in this chapter to those found in similar studies in the literature. Finally, there will be a detailed discussion of how this model was selected, identifying the limitations of the less complex of the models and illustrating how the less complex model results could misinform policy. The model selection section is last because it is based on a comparison of both the estimated coefficients and the elasticities between the models, and therefore a full description of the selected model and its implications for policy will be helpful in understanding the model selection process.

3.3.1 Explanatory variables and how to interpret their estimated coefficients

Explanatory variables included in the model were chosen based on a combination of data availability and economic theory. Variables that influence commute mode choice and car ownership status are meant to represent the relative "prices" of the alternatives in money, time, and convenience. Variables that influence residential location are meant to capture the relative attractiveness of neighborhoods in terms of attributes such as cost, transit access, and local availability of services. Additional variables that influence residential location choice include characteristics of the inhabitants of each location. As will be immediately apparent from examination of the tables of estimated coefficients, there are some included variables that are not statistically significant. These variables remain in the model because they were statistically significant in alternative model specifications and/or there is theoretical basis for their inclusion.

Many of the coefficients in the model are estimated separately for low- and high-income commuters, and some are estimated separately for commuters with children. Segmenting the model in this way explicitly allows for some structured heterogeneity of preferences. A low-income commuter is defined as coming from a household that earns an income per household member of less than $25,000. High-income commuters come from households that earn more than $25,000 per household member.

The independent variables in all of the results tables are divided into groups based on which sub-choice within the dependent variable that they are likely to affect most: commute mode choice, car ownership status choice, and residential location choice. It is worth emphasizing that this grouping of variables is for exposition purposes only; there is no such grouping of variables in the actual model estimation process. In the model estimation, all of the included independent variables explain the dependent variable that is the compound choice.

Many of the independent variables are interaction variables, and they should be interpreted according to the following examples. It may be helpful to refer to Table 3.3 while reading this section. The generic variables have the most intuitive interpretations. A variable is generic if a single variable takes values for all alternatives. A generic variable in the current model is “Commute Cost Not Including Parking Costs”, and its negative sign for both low- and high-income commuters indicates that as commute cost for any alternative rises, the utility of that alternative falls.

Coefficients on alternative-specific variables are interpreted to have meaning only for the alternative specified. For instance, the negative coefficient on “Subway

Lines At Home for Bus" means that as the number of subway lines near home rises, the utility of alternatives that include the bus mode goes down. In another example, the positive coefficient on "Household Size if Two or More Cars in HH" means that as the household size rises, the utility of having two or more cars in the household rises.

The final type of variable is an interaction between a characteristic of an alternative and a characteristic of the individual. Almost all of the variables in the residential location choice section of the model fall into this category. Their interpretations are all analogous to the following: the negative sign on the coefficient of "Neighborhood Percent White if Non-White HH" means that for non-white commuters, the percent of households who are white in a given census tract reduces the utility of that residential location.

The signs of the coefficients of a multinomial logit model can be interpreted intuitively as in the above examples. The magnitudes of individual coefficients, however, have meaning only when considered relative to each other.

In the one- and two-choice models (see Appendix A), there are some alternative-specific variables that do not appear in the full joint choice model. The purpose of these variables is to control for the choice(s) that are not endogenous in these models. For example, in the car ownership status choice variables section of Table A.1, the positive sign on the variable "Auto Commute Mode if One Car" indicates that auto commuters are more likely to be in one-car households. In this model, the commute mode is not endogenous, and is therefore taken as given.

In addition to the coefficients that are listed in the tables at the end of this chapter, each model also includes alternative-specific constants. These are dummy variables that serve the purpose of normalizing the model so that it will be sure to at least reproduce the sample shares of the actual choices of the sample. Usually, there are J-1 alternative-specific constants, where J is the total number of (compound) alternatives in the model. However, the current model includes only 11 of the 2000-or-so residential location alternatives for each commuter. To avoid the impossibility of estimating approximately 40,000 alternative-specific constants, the residential location alternatives are aggregated into three alternative groups: Manhattan, Staten Island, and the Rest of the City residential locations. The estimated alternative-specific constants are therefore still J-1, but now J is three (for the three aggregate residential locations) times the number of car ownership status and commute mode choice compound alternatives. In the full model, this means that J is 3*20=60, and 59 alternative-specific constants are included in the estimation.

3.3.2 Interpreting estimated coefficients in the joint model of residential location, car ownership status, and commute mode choice

The result of the model selection exercise is that the chosen model is the multinomial logit model of the joint choices of residential location, car ownership status, and commute mode choice. Here, the estimated coefficients for this chosen model are discussed in some detail.

Most of the statistically significant coefficients in the commute mode choice category of Table 3.3 have the expected signs. Higher travel costs and travel times lower the utility of the alternative. For lower-income commuters, the point estimate of the effect of travel cost on their utility is greater than the effect of travel cost on higher-income commuters. This is consistent with theory, as money is more valuable for lower-income commuters than for higher-income commuters. The exception to this is that the point estimate of the variable Parking Cost At Work is slightly larger for higher income commuters. The difference between the coefficient estimates for low- and high-income commuters for this variable is not significant, however.

A similar story can be told for the point estimates of travel time, where higher-income commuters have higher point estimates for these coefficients than do lower-income commuters. This is also consistent with theory, since higher-income commuters are likely to have higher values of time than lower-income commuters. These differences between lower- and higher-income commuter coefficient estimates are not statistically significant, however.

As for the coefficients on variables that are not segregated by income, both bus and auto are lower-utility commute mode alternatives where there is higher subway line availability. Where there are more subway lines available near work, subway is a higher-utility commute mode alternative. This makes sense, as subway line availability should make subway a more attractive mode alternative while reducing the attractiveness of all other commute modes. Where there are more subway lines available near home, however, subway is estimated to be a lower-utility commute mode alternative. I offer the following as a possible explanation. Areas with the highest number of subway lines in New York City are also areas with the highest walk accessibility. Therefore, although this model predicts intuitively that New Yorkers will switch away from bus and auto commuting in areas with high subway access, it may be that in some of these areas, the switch is to walking rather than to riding the subway.

In the car ownership choice category, all of the signs on the statistically significant coefficient estimates are as expected. Higher car insurance prices lower the utility of the car-owning alternatives, more strongly for commuters from lower income households. Higher income increases the utility of these alternatives, again more strongly for commuters from lower income households. For both income categories, higher incomes have a stronger effect on owning two or more cars than on owning one car. Greater availability of subway lines at home reduces the utility of owning a car for lower income commuters, but is insignificant for higher income commuters. Living farther from midtown Manhattan raises the utility of owning a car, and living in a higher density area (both in terms of population and retail density) lowers the utility of owning a car. Commuters from larger households have a higher utility of car ownership.

In the residential location choice category of variables as well, most of the statistically significant signs on the estimated coefficients make intuitive sense. Higher rent reduces the utility of a location, and higher neighborhood income increases its utility. A higher neighborhood percentage of people who are racially different from the commuter's household reduces the utility of the alternative. Higher subway line availability raises the utility of that alternative.

Higher population density - all else equal - increases the utility of the neighborhood alternative. This may not be the expected direction of effect in other parts of the country, but within New York City, it makes sense that the most desirable neighborhoods would also have the highest density. Compared to households without children, this model indicates that households with children prefer to live in locations that are farther from midtown Manhattan and have fewer subway lines available. This makes sense as well, since households with children have less disposable income and are more likely to need a car to carry the family around. For that reason, they gravitate toward more car-friendly neighborhoods that have less transit service and might be a bit cheaper to live in.

There are a few counterintuitive signs on the estimated coefficients in this section of the model, however. The negative signs on the coefficients for "NH Percent Owner-Occupied" look strange at first, since it is normally found that neighborhoods with a higher percent of owner-occupied housing are more desirable. The reason given for this is that to protect their investment, homeowners are more likely to take a long-term view of and a serious interest in their neighborhood. To understand the present model's negative estimated coefficients, one simply needs to note that the coefficient on this variable is strongly positive when estimated only on the homeowning segment of the sample. This means that the net negative coefficients are relevant only to the renting segment of the sample, for whom perhaps the percent of their neighbors who are homeowners is not important.

The two remaining variables in the residential location choice section of the model are more mysterious. Statistically significant model coefficients indicate that lower-income commuters prefer to live farther from midtown Manhattan, and that higher-income commuters prefer to live closer. This difference may be explained by the difference in housing prices per square foot that depends on how close the housing is to midtown Manhattan. This effect is only imperfectly controlled for by the "Rent Per Income Per Household Size" variable because this variable does not account for the physical size of the housing.

The negative coefficients on retail density are also somewhat counterintuitive.

High retail density may have two opposing effects on a neighborhood. The first is the ability to easily access local stores and restaurants on foot, and theory predicts that this should have a positive effect on the utility of a location. The second is the higher traffic and noise that come with higher retail density, and theory predicts that this should have a negative effect on the utility of a location. My hypothesis is that the latter effect is swamping the former in this analysis, and the model estimates suggest that retail density lowers the utility of a location – people want some access to retail, but do not want to live in a retail-dominated neighborhood.

3.3.3 Elasticities

Tables 3.4 and 3.5 present the elasticities of car ownership and use for commuting with respect to a number of variables in the estimated full joint choice model. Tables A.6 through A.20 in Appendix A of this dissertation provide corresponding elasticity estimates for the remaining 11 estimated models. Because they are estimated using discrete choice models, recall that these elasticities are the percent change in the probability of choosing a particular alternative when an independent variable is increased by one percent. Although they are not identical, these elasticities can be compared to demand elasticities because they can also be interpreted as the percent change in the market share (similar to demand) of the particular alternative when an independent variable is increased by one percent. For a detailed description of how the elasticities presented in this chapter were calculated, please refer to Chapter 2, Section 2.4.

The elasticities are shown for the entire sample and then separately for Manhattan residents, Staten Island residents, and the residents of the other boroughs. Table 3.4 presents elasticity estimates for all income levels, while Table 3.5 separates the sample into low- and high-income groups. These subsample elasticities were calculated by extracting the subset of the sample that actually chose to live in each location and were in each income category, and calculating the probability-weighted elasticities for each of the subsamples. Note that the borough-level and income-specific elasticities are calculated from the model estimated using the entire sample, and that, as is evident in the coefficient tables, separate coefficients are not estimated for each of the city's subregions. By not estimating different coefficients for each area of the city, the model assumes that, after controlling for commuter socioeconomic characteristics, preferences are similar across the city.[1] The differences we observe in their choice behavior (see Table 3.1) are assumed to come from differences in the transportation-land use contexts across city boroughs.

Focusing on Tables 3.4 and 3.5, it is easy to see which of the variables, if changed by one percent, would have the largest effect on commuters' choices of car ownership and commuting by car.

Car use for commuting

The variables that have the largest impact on car use for commuting in all of the boroughs of New York and for all income levels are the travel cost and travel time for both car commuting and non-car commuting. The elasticity with respect to non-car commute time - the largest elasticity in the table - is in the neighborhood of 1. Taken literally, this means that if non-car commute time were reduced (rather than increased) by one percent, commuters would have a one percent lower probability of commuting by car. Interpreting this result using the market share analogy, it means that if non-car commute time were reduced by one percent for everyone in the city, the market share of car as a mode for commuting would fall by one percent.

One might expect that the effect of car commute time changes would be similar and opposite of the effect of non-car commute time changes. A quick examination of Table 3.4 will show that although the effect of increasing car commute time does decrease the probability of car use for commuting, the effect is consistently smaller than the effect of decreasing non-car commute time. The major reason for this is

[1]Economic theory dictates that this assumption should be true, provided that the model has adequately controlled for socioeconomic differences between areas and that the most important explanatory variables are well-specified and included in the model. The extent to which this assumption is actually valid is a question for future research.

that travel time is divided into three time categories. For car commute time, the time change is only in the Riding Time variable. For non-car commute time, the time change is in all three categories of time: Walking Time, Waiting Time, and Riding Time. Table 3.6 contains the decomposition of the car use for commuting elasticity with respect to non-car commute time into the three components of time. This table illustrates that a substantial portion of the differences in the elasticities with respect to car commute time and non-car commute time is accounted for by the Walking Time and Waiting Time portions of non-car commute time. The elasticity of car use for commuting with respect to non-car riding time is still larger in absolute magnitude than that with respect to car commute time, but the magnitudes are closer. This is to be expected because the magnitude of a one percent change in riding time for car and non-car mode alternatives is dependent on the actual riding time for the two mode groups, and a larger magnitude change in the variable of interest will lead to a larger change in the estimated probability of choosing to commute by car. An examination of Figures 2.16 through 2.18 reveals that in fact, the mean riding time for auto commuters is 27 minutes, while the mean riding time for transit riders is more than 40 minutes.

Commute cost for both car and non-car modes also has a substantial impact on the choice of New Yorkers to commute by car. In the case of commute cost, the elasticity of car use for commuting is consistently larger with respect to car cost than non-car cost. This reason for this is that the mean travel cost (not including parking) for auto commuters is larger than that for non-car commuters. Auto commuters pay an average of \$2.27, while the mean travel cost for non-car commuters is \$1.54. As in the case of travel time, a one percent increase in a larger base amount is a larger absolute increase in travel cost, leading to a larger change in the probability of choosing to commute by car.

There are two additional variables that appear to have the potential to substantially affect New Yorkers' choice of the car as their commute mode, with elasticities of approximately -0.2. These are the home population density and subway availability at work. According to the present model, the remaining two evaluated variables -

subway lines at home and parking cost at work have very little effect on New Yorkers' choice of the car for commuting. There appears to be little difference between low- and high-income commuters in the elasticities of car use for commuting with respect to any of the variables evaluated.

Car ownership

Turning to the elasticities of car ownership, one of the variables to which car ownership is most sensitive is income. The income elasticities were calculated for the purpose of comparing this model with models in the existing literature, and as a point of comparison for the elasticities of car ownership with respect to other variables. Elasticities with respect to income are not relevant to any questions of policy.

Putting income aside then, the policy-relevant variables that have the largest effect on the choice of car ownership status are commute costs and times by car and by non-car modes, and home population density. These effects are estimated separately as the effect on the probability of being in a zero-car household, being in a one-car household, and being in a household with two or more cars. Recall that these elasticities are weighted by the original probabilities. This means that the elasticity estimate - for example - for a zero-car household is dominated by the changes in probability that zero-car households experience when faced with a one percent increase in the variable of interest.

It is interesting to compare the elasticities of zero-car ownership calculated separately for Manhattan and Staten Island. The zero-car ownership elasticities with respect to every tested variable are larger in magnitude in Staten Island than in Manhattan. This makes sense, as Staten Island's zero-car households are likely to be close to a threshold point beyond which they would need a car for daily trips. This means that small changes in the independent variables could cause zero-car Staten Islander households to switch to become one-car households. In contrast, Manhattan's zero-car households are not as likely to be close to such a threshold point. As for differences between income groups, these are seen only in the elasticities of zero-car ownership with respect to population density, with lower income households being less sensitive to population density than higher income households.

The effects of all the variables on the probability of being in a one-car household are small. One interesting thing to note, however, is that for both the effect of home population density and the effect of income, the effects are in different directions in different boroughs of the city. In the case of home population density, the effect is in different directions for low- and high-income households. In the case of income, the effect is positive everywhere except in Staten Island. This is because Staten Island is the most suburban of the boroughs of New York City, and it is the norm among Staten Islanders to own cars. Then, it is logical that if incomes rise in Staten Island, larger moves will be from one- to two-car households than from zero- to one-car households. This explains the negative sign on the elasticity of one-car ownership with respect to income in Staten Island.

The relative magnitudes of the effects of variables on two-or-more car ownership are similar to the zero-car ownership case, but the directions of the effects have switched. Home population density, non-car commute time, and car commute time have the largest magnitude effects on two-or-more car ownership. In the case of population density, the effect is larger for lower-income households than it is for higher-income households.

Once again, it is interesting to compare the elasticity estimates of two-or-more car ownership for Manhattan and Staten Island households. In the case of two-or-more car ownership, the elasticity magnitudes for every tested variable are larger for Manhattan households than for Staten Island households. This also makes sense because Manhattanite households that own two-or-more cars are likely to be close to a threshold point beyond which owning multiple vehicles is prohibitively expensive and/or inconvenient. Since owning at least one vehicle in Staten Island is the norm, Staten Islanders are less likely to be close to this threshold. Thus, a small change in an independent variable is likely to have a larger effect on the probability of owning two-or-more cars in Manhattan than in Staten Island.

Discussion

Overall, most of the elasticity estimates estimated using the selected model and presented in Table 3.4 and Table 3.5 make intuitive sense and are broadly consistent with the range of estimates found in the literature, with the Staten Island elasticity estimates being closest to estimates of elasticities in less dense cities. This in itself is interesting because one might expect these numbers to be more different due to the often-cited uniqueness of the land use and transportation system in New York City. This consistency with other literature is encouraging as well for the generalizability of the results from the current model. The elasticity estimates in the tables in Appendix A are provided for model comparison purposes, and will be discussed later in this chapter.

The income elasticity of car ownership and the travel cost elasticity of car use are the numbers most commonly found in the literature, and they are used here for comparison purposes. Using aggregate data, Ingram and Liu (1997) found the income elasticity in global cities to be 0.5 for all levels of car ownership. Using household-level data, Schimek (1996) estimated this same number to be 0.221 for the US, and de Jong (1990) found a value of 0.42 for the Netherlands. Using a discrete choice framework, Bhat and Pulugurta (1998) find for Boston in 1991 that the income elasticities of car ownership were -0.938 for zero-car households, -0.189 for one-car households, and 0.281 for two-car households.

The elasticities of car ownership presented in this dissertation are broadly consistent with these previous estimates. In direct comparison to Bhat and Pulugurta's work, they indicate that Staten Islanders are similar to Bostonites. The relevant comparison with Bhat and Pulugurta's work is found in Table A.19, which presents the elasticities of car ownership calculated from a model of the choice of car ownership only. These elasticities of car ownership with respect to income on Staten Island are -1.26, -0.51, and 0.33 for zero-, one-, and two-or-more-car households, respectively. Note that these elasticities become substantially smaller in magnitude when calculated using the full joint model. The elasticities found in Table 3.4 of car ownership with respect to income on Staten Island are -0.65, -0.06, and 0.30 for zero-, one-, and two-or-more-car households, respectively.

Zhang (2004) finds that the travel cost (including parking) elasticity of car use is -0.144 in Boston, and -0.242 in Hong Kong. These findings of Zhang are comparable to the results presented in Table A.18 of -0.28 for Staten Island, and -0.31 for the whole of New York City for travel cost elasticity of car use. Note here that the elasticities of car use for commuting with respect to travel cost are much larger when calculated using the full joint model. Additional estimates of travel cost elasticity of car use from the literature include Asensio's (2002) estimate of -0.092 for Barcelona and Hensher and Ton's (2000) estimates of less than -1 for Sydney and Melbourne.

Using his mode choice only model, Zhang (2004) finds that the population density elasticity of car use is -0.044 in Boston and -0.039 in Hong Kong, and Schimek (1996) finds this elasticity to be -0.069. In the models estimated for this dissertation, population density is not an explanatory variable in the mode choice segment of the model. However, for comparison purposes to Zhang's study, a mode choice only model run was conducted with population density for the auto modes as an explanatory variable. The calculated car use elasticity with respect to population density from this model is -0.05, entirely consistent with the existing literature.

The elasticity estimate presented in this chapter that is distinct from estimates in the literature is the car ownership elasticity with respect to population density. Schimek finds that the population density elasticity of car ownership is -0.057. In contrast to this relatively modest estimate, the estimates presented in both Table 3.4 and the elasticity tables in Appendix A are surprisingly large. In fact, for higher income commuters, population density appears to be the only potential policy lever in this model that substantially influences car ownership choice.

In itself, I would not expect population density to have a strong effect on car ownership or use. It is the fact that population density is correlated with variables that directly affect the time, money, and convenience "prices" of cars and their substitutes and complements that makes it such a powerful explanatory variable. These car-relevant correlates of population density include traffic congestion, parking availability and cost, transit availability, and local availability of retail, services, and employment.

What is a bit mysterious about this finding of large population density elasticities of car ownership and use is that this model controls for many of these car-relevant correlates of population density. The main variables that are not adequately controlled for in this model are parking cost and availability and traffic congestion. Traffic congestion is partially controlled for through time-of-day dependent travel speeds that were used to calculate ride time. This may not be a sufficient control, however, since a large part of the problem of traffic congestion is not the reduced speeds, but rather the threat of total gridlock. In the case of parking cost and availability, adequate data was simply not available.

3.3.4 Parking cost - the missing link?

Numerous attempts were made to obtain home parking cost data and include it in the model. However, as discussed in Chapter 2 of this dissertation, because options for parking in New York City (and their associated costs) are so varied, it is difficult to accurately "guesstimate" monthly parking costs for households. First of all, on-street parking is priced *much* lower in dollars than off-street parking, but on-street parkers must pay with their time searching for a free space. Manhattan has "Alternate Side Parking Regulations", meaning that on-street parkers in Manhattan must move their vehicles every business day. On-street parkers in most of the rest of the city must move their vehicles only 1-2 times each week for street cleaning. Off-street parking search time and car-moving to comply with city regulations, in contrast, is virtually zero, but the monetary cost is high (between $150 and $600 per month in much of Manhattan). This means that car owners with lower incomes will be more likely to choose on-street parking and car owners with high incomes will be more likely to choose off-street parking. There is a large middle income bracket, however, where car owners will sometimes choose on- and sometimes choose off-street parking.

Even if the problem were simplified so that all car owners chose off-street parking, the high variation in off-street parking costs makes it difficult to accurately guess how much a particular household would pay. The cost variation, of course, is correlated with city neighborhood, but if a household does not use a car regularly, it might decide to park the car in a cheaper part of the city than where it has chosen to live. If the vehicle being parked is expensive, garage parking might be deemed necessary, while a less-expensive vehicle could be parked in a low-security outdoor lot without a problem.

Because of this complexity in time and money costs for parking, estimated coefficients for home parking cost based on "guesstimated" data were too sensitive to the underlying assumptions to be reliable, and therefore home parking cost was removed from the model. Unfortunately, this omission might lead to bias in the estimated coefficients of variables that are correlated with home parking cost because these coefficients will act as partial proxies for the omitted variable.

This is exactly what may be happening in the case of population density. Part of the high values reported here for the population density elasticities of car ownership and use for commuting may be a result of the fact that parking prices and availability are likely to be highly correlated with population density. This means that these elasticities may actually represent the substantial effect of parking prices and availability rather than the population density itself.

3.3.5 Model Selection Part 1: Comparison of the joint choice model to individual models of the sub-choices

There are two main aspects of model selection addressed in this section. First and more importantly, this section addresses the question of the importance of a model of the choices of residential location, car ownership status, and commute mode choice be a single model that endogenizes all three decisions. The second aspect of model selection addressed here is the question of whether it is necessary to use a nested logit or if a joint logit model will suffice for this estimation.

For the first of the model selection questions identified above, six multinomial logit choice models were estimated that include all of the possible sub-models of the full joint choice models. These were as follows:

1. Joint choice of residential location, car ownership status, and commute mode
2. Joint choice of residential location and car ownership status
3. Joint choice of residential location and commute mode
4. Joint choice of car ownership status and commute mode
5. Choice of car ownership status
6. Choice of commute mode

The estimation results for these models can be seen in Table 3.3 and in Tables A.1 through A.5. To answer the question of the importance of modeling these choices jointly, the main evaluation method is to compare statistical significance and sign differences between the estimated coefficients in the models. The model that is most consistent with theory is the preferred model. One way to compare the models statistically is to compare the distribution of each model's predicted probabilities for the alternatives that were actually chosen. This study employs both of these methods.

The signs of the model coefficients are most consistent with theory in the full joint model of the choices of residential location, car ownership, and commute mode. The main changes in the coefficient signs that appear in the alternate models listed above are in the coefficients on "Riding Time". For the subset of models that does not include the choice of residential location as part of the dependent variable, the estimated coefficients of "Riding Time" are positive, indicating that longer riding times are more desirable. It is possible that this is true for some range of riding times (Redmond and Mokhtarian, 2001), but it is not likely that in general, longer riding times make an alternative more desirable. Instead, the explanation I offer is that because trip distance and riding time are closely related and longer trips may be associated with more desirable residential locations, riding time appears to be have a positive effect on utility when residential location choice is not included as part of the choice.

In addition, there are two other coefficients that are statistically significant and have different signs in some of the alternate models than in the full model. The first of these is the coefficient on car insurance for high income households. In the full model, as expected, it is negative. Nobody should prefer to pay more for car insurance. However, in both the joint model of car ownership and mode choice (Table A.3) and the single choice model of car ownership (Table A.5), the coefficient on car insurance for high income households is positive. The second coefficient that changes sign in one of the alternate models is that on the variable "NH Miles From Midtown Manhattan", again for high income households. In the full model, this coefficient is negative, indicating that high income households prefer to live closer to midtown. In the model of residential location and mode choice (Table A.2), this coefficient is positive.

There are many coefficients that are statistically significant in the full model and statistically insignificant in some of the alternate models. There are a few coefficients that are significant in one of the alternate models and statistically insignificant in the full model. The fact that more of the coefficients are stastistically significant in the full model, though, adds to the evidence that the full model provides the best fit to the data.

This evidence based on coefficient signs and statistical significance indicates strongly that residential location should be included as part of the dependent variable in a model of mode choice in New York City. Additional evidence of this is provided by comparison of each model's predicted probabilities for the alternatives that were actually chosen. This is done by comparing the average predicted probability for the chosen alternative in the joint choice model with the product of the average predicted probabilities for the chosen sub-alternatives from the single choice models.[2]

[2]This is the correct comparison to make. There is also another method that is tempting to try, but is incorrect. This is to compare the average predicted probability for each chosen sub-alternative in the joint choice model with the average predicted probability of the chosen sub-alternative in each single choice model. This second method will yield the result that the single choice models outperform the joint choice models because the joint choice models are trying to predict something much more complicated, and effective prediction of each sub-choice is compromised to achieve the

This comparison method is relatively simple. First, the joint choice model is estimated. Then, the resulting predicted probabilities for each individual's chosen compound alternative are averaged. Since the model is estimated using neighborhood weights, the averages here are weighted as well, using this same weighting scheme. For the comparison, it is necessary to also estimate the single choice models for each sub-choice, and calculate the weighted average of these predicted probabilities for each individual's chosen sub-alternative. The goodness-of-fit comparison is between the weighted average probability for the compound alternative and the product of the weighted average probabilities for the sub-alternatives. The following mathematical expression represents the comparison (without the weighting).

$$\frac{\sum_n \sum_j y_{nj} P_n(lcm)}{N} \; versus \; \frac{\sum_n \sum_l \sum_c \sum_m y_{nl} P_n(l) * y_{nc} P_n(c) * y_{nm} P_n(m)}{N}$$

where: l signifies the location choice,
c signifies the car ownership choice,
m signifies the mode choice,
$y_{nj} = 1$ if individual n chooses compound alternative j
$= 0$ otherwise, and
$y_{nl},\ y_{nc},\ and\ y_{nm}$ are defined in an analogous manner.

The results of this comparison can be seen in Table 3.2. By this measure, the joint choice models perform better for both the full compound choice case and for the location-mode choice case. For the car-mode and the location-car choice cases, the separate models perform better than the joint choice model. This is consistent with the model selection discussion above that focuses on coefficient estimates in that it re-emphasizes the importance of jointly modeling the choices of residential location and commute mode. Because the current research is focused on car ownership status as well as car use for commuting, I chose to continue to include as endogenous the choice of car ownership status in the present model. This choice was made in spite of the evidence presented here that suggests that its inclusion may actually compromise model goodness-of-fit.

---

best prediction of the joint choice.

Table 3.2: Goodness-of-Fit Comparison

Weighted Average Predicted Probabilities for Chosen Alternative		
Model	Joint	Product of Separate
Full Compound Choice	0.050	0.047
Location-Mode Compound Choice	0.104	0.081
Car-Mode Compound Choice	0.206	0.226
Location-Car Compound Choice	0.100	0.102

3.3.6 Model Selection Part 2: Joint versus nested logit specifications

Turning to the question of whether to use a nested logit or a multinomial joint logit model specification, the evidence to be used is in the inclusive value estimates presented in Tables A.1 through A.3 together with a comparison of the elasticity estimates from these models. Due to the large number of variables in the full model (117 variables) together with the large number of alternatives in the model (220 variables), it was not possible to estimate a nested version of the full model.[3] To evaluate whether or not a nested logit model would perform better than the joint multinomial logit, nested and non-nested versions of each two-sub-choice model were estimated and compared.

As is explained in Chapter 2, the inclusive value estimates in nested logit models are the main indicators of whether a nested specification is needed. If the estimates of the inclusive value parameters are 1.0, the nested logit specification is identical to the joint logit specification. If the estimates of the inclusive value parameters are signficantly greater than 1.0, this indicates that the probability of choosing alternatives within a nest are positively correlated so that they move together, and there is greater substitution between nests than within nests (Train et al., 1987; Gangrade et al., 2002). Since the nested specification is meant to put alternatives that are close substitutes into nests together, one suggested solution to the problem of inclusive

[3]Stata 8.1 could not estimate a nested version of the model on even a portion of this dataset. The limitation of Gauss appears to be the physical size of the full dataset - it is too large for Gauss to load all at once. NLogit limits nested logit models to a maximum of 100 elemental alternatives.

value estimates that are signficantly greater than 1.0 is to “switch” the nesting structure (Train, 2005). Tables A.1 through A.3 each include the joint multinomial logit specification alongside estimations using each of the two possible nesting structures that represent the two sub-choices of each model. In each case, one of the nesting structures yields inclusive value estimates that are significantly greater than 1.0, while the other yields estimates that are either approximately 1.0 or significantly less than 1.0.

The two-sub-choice model of residential location and car ownership is presented in Table A.1. The inclusive value parameter estimate is negative for the nesting structure where each nest has residential location in common (the 11-nest case). In no case is a negative inclusive value estimate consistent with basic utility theory, and this model is rejected. In the case where each nest has car ownership level in common (the 3-nest case), the inclusive value parameter estimates are each approximate 2.0, a result that indicates that within each nest, the alternatives are complements rather than substitutes. Neither of these two-sub-choice nested models appears to be the correct model, but a comparison of the elasticity results is done as well to provide an additional check to make sure that the joint logit model is consistent with the nested logit model results. Tables A.11 through A.14 present the car ownership elasticities calculated from these three versions of the model. Since the model with 11 nests has already been discarded due to lack of consistency with theory, the relevant comparison is between the values in Table A.11 and Table A.14. Most of the elasticities are stable across the model specifications, but the one that varies the most is that with respect to population density.

The two-sub-choice model of residential location and commute mode is presented in Table A.2. This time, the specification in which each nest has residential location in common (the 11-nest case) yields an inclusive value parameter estimate of approximately 2.0, while the alternative specification in which each nest has commute mode in common (the 7-nest case) yields estimates close to 1.0. Three out of the seven inclusive value estimates for this second specification are significantly different from 1.0, however, so it is instructive to once again compare the elasticities in Table A.15 with those in Table A.17. These tables indicate that the difference in calculated elasticities between the joint logit model and the preferred nested logit model are small.

The final set of two-sub-choice models looks at the car ownership and commute mode choices. Here, the specification in which each nest has commute mode in common (the 7-nest case) yields inclusive value estimates of approximately 2.0. The alternative specification in which each nest has car ownership level in common yields inclusive value estimates that are between 0 and 1.0, with one of the estimates being significantly different from 1.0. Looking to Tables A.6 and A.8 to compare elasticities calculated from the joint logit model and the preferred nested specification, the elasticity estimates are almost identical.

In spite of not being able to estimate a three-level nested logit model directly, the comparisons conducted here of the two-level models here provide strong evidence that a three-level nested model is not likely to provide results that are substantially different from the joint multinomial logit model of the three sub-choices.

3.4 Conclusion

This chapter has attempted to identify the most effective policy levers to control car ownership and use in dense urban areas such as the study site of New York City. This is accomplished through the creation of a discrete choice econometric model of car ownership that endogenizes the choices of both residential neighborhood and commute mode. The model purposefully includes more policy-sensitive variables than previous studies, and produces a number of policy-relevant elasticity results for New York City: a dense, transit-rich city in the United States.

Broad consistency of the results presented here with those of previous studies is heartening for the possibility of generalizing these results to other cities using a simulation approach. This is an area for future research.

The main departure from the literature in the results of this paper is in the elasticities of car ownership with respect to population density. New Yorkers appear to be quite sensitive in their car ownership and use choices to changes in population density. As discussed above, this result could be interpreted as sensitivity to parking prices rather than to population density.

The results presented here indicate the effectiveness of policy-sensitive variables on car ownership and use for commuting. For instance, these results indicate that the most effective way to reduce car use for commuting is to decrease commute time for non-car modes. To reduce car ownership among all New Yorkers, the most effective policy-sensitive variable appears to be population density. For policymakers, however, effectiveness is not the only consideration. The cost must also be considered. And effectiveness is not necessarily equivalent to cost-effectiveness. To use these results to inform policy, they must be interacted with cost information about competing policy alternatives. This is another area for future research.

Table 3.3: Multinomial logit model of the Full Joint Choice of Residential Location, Car Ownership Status, and Commute Mode

	Coefficient	S.E.	Coefficient	S.E.
COMMUTE MODE CHOICE VARIABLES	**Income < $25,000 per HH member**		**Income > $25,000 per HH member**	
Commute Cost Not Including Parking Costs	-0.468***	0.031	-0.414***	0.035
Parking Cost At Work	-0.024*	0.014	-0.026*	0.014
Walking Time	-2.238***	0.130	-2.501***	0.185
Waiting Time	-3.736**	1.447	-6.159***	1.900
Riding Time	-1.477***	0.110	-1.735***	0.167
Not Segregated By Income				
Subway Lines At Home for Bus	-0.249***	0.064		
Subway Lines At Home for Subway	-0.096*	0.051		
Subway Lines At Home for Auto	-0.111*	0.058		
Subway Lines At Work for Bus	-0.217***	0.049		
Subway Lines At Work for Subway	0.230***	0.044		
Subway Lines At Work for Auto	-0.026	0.047		
CAR OWNERSHIP STATUS CHOICE VARIABLES	**Income < $25,000 per HH member**		**Income > $25,000 per HH member**	
Car Insurance Cost	-0.582***	0.081	-0.372***	0.133
Income if One Car	0.528***	0.097	0.123***	0.039
Income if Two or More Cars	1.002***	0.105	0.152***	0.046
Subway Lines at Home if One Car	-0.117**	0.048	-0.004	0.054
Subway Lines at Home if Two or More Cars	-0.005	0.059	0.006	0.072
Miles to Midtown Manhattan if One Car	0.038**	0.018	0.169***	0.029
Miles to Midtown Manhattan if Two or More Cars			0.152***	0.028
Retail Density at Home if One Car	-1.111	1.027	0.429	0.588
Retail Density at Home if Two or More Cars	-4.902**	2.379	-3.098**	1.360
Population Density at Home if One Car			-0.070***	0.019
Population Density at Home if Two or More Cars	-0.275***	0.023	-0.125***	0.027
Not Segregated By Income				
Household Size if One Car	0.152***	0.041		
Household Size if Two or More Cars	0.514***	0.042		

	Coefficient	S.E.	Coefficient	S.E.
RESIDENTIAL LOCATION CHOICE VARIABLES	**Income < $25,000 per HH member**		**Income > $25,000 per HH member**	
Rent Per Income Per Household Size	-0.050	0.100	-2.014**	0.779
NH Percent Owner-Occupied	-1.425***	0.237	-1.413***	0.321
NH Population Density	0.120***	0.009	0.176***	0.016
NH Miles From Midtown Manhattan	0.043***	0.016	-0.062**	0.029
NH Retail Density	-1.776***	0.587	-0.352	0.476
NH Subway Line Availability	0.223***	0.058	0.140**	0.062
NH Median Income	-0.036	0.029	0.220***	0.035
Not Segregated By Income				
NH Miles From Midtown Manhattan if Kids in HH	0.051***	0.016		
NH Subway Line Availability if Children in HH	-0.115***	0.042		
NH Percent White if Non-White HH	-2.337***	0.160		
NH Percent Non-White if White HH	-2.838***	0.138		
NH Percent Owner-Occupied if Homeowner	2.913***	0.209		
PLUS ALTERNATIVE-SPECIFIC CONSTANTS[a]				
ESTIMATION SUMMARY INFORMATION				
Observations	2728			
Alternatives[b]	220			
Pseudo R-squared	0.29			

* significant at 10%; ** significant at 5%; *** significant at 1%

a. There are 59 alternative specific constants in this model, representing all combinations of commute mode and car ownership, and three residential location groups (Manhattan, Staten Island, and the Rest of New York City).

b. The 220 compound alternatives consist of 7 mode alternatives, 3 car ownership status alternatives, and 11 census tract alternatives sampled from the full set of over 2000 possible census tracts. The compound alternatives that included both Auto Driver for the mode and Zero Car Household for the car ownership status were removed from the set.

Table 3.4: Elasticities of car ownership and car use for commuting in Full Joint Model

	Car Use	**Zero Car**	**One Car**	**Two+ Car**
FIVE BOROUGHS OF NEW YORK CITY				
Population Density (home)	-0.22	0.34	0.11	-0.64
Subway Lines (home)	0.02	0.05	-0.06	0.02
Subway Lines (home and work)	-0.20	0.11	-0.06	-0.06
Car Commute Cost (w/o parking)	-0.57	0.18	-0.04	-0.22
Car Commute Cost (incl. parking)	-0.61	0.20	-0.04	-0.23
Non-Car Commute Cost	0.40	-0.10	0.04	0.09
Car Commute Time	-0.40	0.13	-0.03	-0.14
Non-Car Commute Time	0.96	-0.26	0.05	0.32
Income	n/a	-0.42	0.11	0.45
MANHATTAN ONLY				
Population Density (home)	-0.38	0.41	-0.18	-0.95
Subway Lines (home)	0.03	0.03	-0.04	0.0
Subway Lines (home and work)	-0.45	0.10	-0.08	-0.14
Car Commute Cost (w/o parking)	-0.95	0.12	-0.04	-0.33
Car Commute Cost (incl. parking)	-1.09	0.15	-0.05	-0.37
Non-Car Commute Cost	0.54	-0.05	0.04	0.06
Car Commute Time	-0.61	0.08	-0.03	-0.18
Non-Car Commute Time	1.30	-0.07	0.0	0.24
Income	n/a	-0.36	0.26	0.54
STATEN ISLAND ONLY				
Population Density (home)	-0.12	0.46	0.11	-0.26
Subway Lines (home)	0.0	0.04	-0.03	0.0
Subway Lines (home and work)	-0.07	0.13	-0.02	-0.04
Car Commute Cost (w/o parking)	-0.40	0.31	0.02	-0.14
Car Commute Cost (incl. parking)	-0.42	0.33	0.02	-0.15
Non-Car Commute Cost	0.28	-0.17	0.0	0.07
Car Commute Time	-0.27	0.21	0.0	-0.09
Non-Car Commute Time	0.67	-0.46	-0.08	0.24
Income	n/a	-0.65	-0.06	0.30

	Car Use	**Zero Car**	**One Car**	**Two+ Car**
REST OF NEW YORK CITY				
Population Density (home)	-0.21	0.31	0.15	-0.65
Subway Lines (home)	-0.02	0.05	-0.07	0.03
Subway Lines (home and work)	-0.19	0.11	-0.06	-0.06
Car Commute Cost (w/o parking)	-0.55	0.20	-0.04	-0.21
Car Commute Cost (incl. parking)	-0.58	0.21	-0.04	-0.23
Non-Car Commute Cost	0.41	-0.12	0.05	0.09
Car Commute Time	-0.40	0.14	-0.03	-0.15
Non-Car Commute Time	0.96	-0.31	0.06	0.34
Income	n/a	-0.43	0.09	0.46

Table 3.5: Elasticities of car ownership and car use for commuting in Full Joint Model By Income Level

	Car Use		**Zero Car**		**One Car**		**Two+ Car**	
	Low	High	Low	High	Low	High	Low	High
FIVE BOROUGHS OF NEW YORK CITY								
Population Density (home)	-0.23	-0.20	0.25	0.56	0.25	-0.17	-0.77	-0.38
Subway Lines (home)	-0.02	0.0	0.06	0.0	-0.10	0.0	0.04	0.0
Subway Lines (home and work)	-0.18	-0.24	0.11	0.10	-0.09	0.0	-0.05	-0.10
Car Commute Cost (w/o parking)	-0.56	-0.59	0.18	0.19	-0.06	0.0	-0.22	-0.21
Car Commute Cost (incl. parking)	-0.58	-0.66	0.19	0.21	-0.06	0.0	-0.23	-0.23
Non-Car Commute Cost	0.42	0.37	-0.11	-0.08	0.05	0.03	0.11	0.05
Car Commute Time	-0.38	-0.45	0.12	0.14	-0.04	0.0	-0.14	-0.14
Non-Car Commute Time	0.94	1.00	-0.29	-0.19	0.09	-0.04	0.34	0.28
Income	n/a	n/a	-0.44	-0.36	0.10	0.14	0.59	0.19
MANHATTAN ONLY								
Population Density (home)	-0.33	-0.42	0.19	0.55	0.17	-0.37	-1.23	-0.78
Subway Lines (home)	-0.04	-0.03	0.07	0.0	-0.12	0.0	0.04	0.0
Subway Lines (home and work)	-0.34	-0.52	0.11	0.10	-0.13	-0.05	-0.08	-0.17
Car Commute Cost (w/o parking)	-0.95	-0.95	0.13	0.12	-0.07	-0.03	-0.35	-0.32
Car Commute Cost (incl. parking)	-1.03	-1.12	0.14	0.15	-0.08	-0.04	-0.37	-0.36
Non-Car Commute Cost	0.56	0.53	-0.06	-0.05	0.06	0.04	0.10	0.04
Car Commute Time	-0.54	-0.66	0.07	0.08	-0.04	-0.02	-0.18	-0.19
Non-Car Commute Time	1.22	1.35	-0.13	-0.03	0.08	-0.04	0.30	0.21
Income	n/a	n/a	-0.38	-0.34	0.26	0.26	0.84	0.37

	Car Use		Zero Car		One Car		Two+ Car	
	Low	High	Low	High	Low	High	Low	High
STATEN ISLAND ONLY								
Population Density (home)	-0.14	-0.09	0.38	0.65	0.27	-0.04	-0.34	-0.16
Subway Lines (home)	0.0	0.0	0.05	0.0	-0.06	0.0	0.02	0.0
Subway Lines (home and work)	-0.06	-0.09	0.12	0.17	-0.04	0.0	-0.03	-0.06
Car Commute Cost (w/o parking)	-0.41	-0.39	0.30	0.34	0.0	0.04	-0.14	-0.13
Car Commute Cost (incl. parking)	-0.43	-0.42	0.32	0.37	0.0	0.04	-0.15	-0.14
Non-Car Commute Cost	0.31	0.25	-0.19	-0.14	0.0	0.0	0.08	0.05
Car Commute Time	-0.25	-0.29	0.19	0.24	0.0	0.02	-0.09	-0.09
Non-Car Commute Time	0.68	0.66	-0.48	-0.41	-0.05	-0.11	0.26	0.22
Income	n/a	n/a	-0.78	-0.40	-0.14	-0.03	0.45	0.10
REST OF NEW YORK CITY								
Population Density (home)	-0.23	-0.17	0.26	0.55	0.25	-0.09	-0.79	-0.34
Subway Lines (home)	-0.02	0.0	0.06	0.0	-0.10	0.0	0.04	0.0
Subway Lines (home and work)	-0.18	-0.22	0.11	0.11	-0.09	0.0	-0.05	-0.09
Car Commute Cost (w/o parking)	-0.55	-0.56	0.19	0.24	-0.06	0.0	-0.22	-0.20
Car Commute Cost (incl. parking)	-0.57	-0.61	0.20	0.26	-0.06	0.0	-0.23	-0.21
Non-Car Commute Cost	0.42	0.36	-0.12	-0.11	0.06	0.02	0.11	0.06
Car Commute Time	-0.38	-0.44	0.13	0.19	-0.04	0.0	-0.15	-0.14
Non-Car Commute Time	0.95	0.99	-0.30	-0.32	0.10	-0.03	0.35	0.31
Income	n/a	n/a	-0.44	-0.37	0.09	0.09	0.59	0.17

Table 3.6: Decomposition of elasticities of car use for commuting with respect to travel time

	Car Use for Commuting		
	All Income	**Low**	**High**
FIVE BOROUGHS OF NEW YORK CITY			
Car Commute Time (Riding Time Only)	-0.40	-0.38	-0.45
Non-Car Commute Time	0.96	0.94	1.00
Non-Car Walking Commute Time	0.25	0.26	0.23
Non-Car Waiting Commute Time	0.15	0.12	0.18
Non-Car Riding Commute Time	0.57	0.56	0.58
MANHATTAN ONLY			
Car Commute Time (Riding Time Only)	-0.61	-0.54	-0.66
Non-Car Commute Time	1.30	1.22	1.35
Non-Car Walking Commute Time	0.34	0.32	0.35
Non-Car Waiting Commute Time	0.22	0.16	0.26
Non-Car Riding Commute Time	0.73	0.72	0.74
STATEN ISLAND ONLY			
Car Commute Time (Riding Time Only)	-0.27	-0.25	-0.29
Non-Car Commute Time	0.67	0.68	0.66
Non-Car Walking Commute Time	0.15	0.17	0.13
Non-Car Waiting Commute Time	0.10	0.08	0.11
Non-Car Riding Commute Time	0.42	0.43	0.42
REST OF NEW YORK CITY			
Car Commute Time (Riding Time Only)	-0.40	-0.38	-0.44
Non-Car Commute Time	0.96	0.95	0.99
Non-Car Walking Commute Time	0.25	0.26	0.22
Non-Car Waiting Commute Time	0.14	0.12	0.18
Non-Car Riding Commute Time	0.57	0.56	0.59

Chapter 4

Understanding Spatial Patterns of Behavioral Response to Policy Change: A Case Study of Transport Choices in New York City Using Discrete Choice Econometrics and GIS

Local policies such as zoning changes or parking regulations are often implemented to different degrees in different neighborhoods of cities. City investments in infrastructure are almost never spatially uniform. Even policies that are implemented uniformly across a city will illicit different behavioral responses from residents based on differences in the existing built environment. It is important for city decision makers to understand how residents in different neighborhoods may respond differentially to these policies and investments. The analysis in this chapter aims to contribute to improving our understanding of the spatial heterogeneity of behavior using New York City as a case study of spatial variation in behavioral response to policy change.

Specifically, this chapter asks two questions:

- What is the spatial distribution of behavior change in response to uniform changes in factors that explain residents' choices of car ownership level and commute mode?
- How do residents respond to spatially differentiated changes in these factors?

To answer these questions, the joint discrete choice model of New Yorkers' choices of residential neighborhood, car ownership status, and transport mode from Chapter 3 is used. Recall that this model treats all choices as endogenous, modeling the three choices as a single joint choice. The discrete choice model is linked to a GIS database, allowing for calculation of spatially-differentiated model results and simulation of spatially-explicit policy scenarios. This tool is able to produce not only information about the behavioral response to transportation and land use policy changes, but also maps of where the people whose behavior is most sensitive to policy are located within New York City. It is this capability that is the focus of the present chapter.

The model results include estimates for each individual of the probability that each compound alternative is chosen. These probabilities can be averaged by neighborhood to create maps that illustrate the spatial heterogeneity of the model results. All of the maps in this chapter that illustrate model predicted probabilities are those probabilities *conditional on residential neighborhood choice.*

This chapter begins by demonstrating the spatial validity of the model, comparing maps illustrating the spatial heterogeneity in the 2000 Census, the present sample, and the predicted probabilities from the model. The focus of the next section is on modeling the spatial heterogeneity of response to spatially uniform change in selected independent variables that represent plausible future scenarios for the city. The section following that focuses on mapping the predicted response to spatially-differentiated changes in selected independent variables. The final section concludes the chapter.

4.1 Spatial validation of the model: The choice of commute mode

Before beginning simulation analysis, this section of the chapter demonstrates the spatial validity of the original model prediction by comparing the prediction to both the original RT-HIS data and to Census data, starting with a comparison or commute mode choices and continuing by comparing car ownership choices. In each set of map comparisons, the neighborhoods of the map that illustrates the data taken from the RT-HIS sample are divided into quantiles. The other maps use the same percentages as the color categories to allow for easy map comparison.

Figures 4.1 through 4.3 illustrate the actual and predicted percent of commuters in each neighborhood in NYC who commute by car. Figure 4.1 shows the spatial distribution of car commuting that was reported by respondents to the 2000 Census. In this first figure, the subway lines are laid over the background neighborhoods to illustrate the clear negative spatial relationship between car commuting and subway line availability. Figure 4.2 shows this same distribution as reported by respondents to the RT-HIS, the main data source used in this dissertation. Figure 4.3 shows the probability of commuting by car in each neighborhood of New York as predicted by the joint multinomial logit model from Chapter 3 of this dissertation.

Interestingly, it appears that the model predicted percent of commuters using cars is closer to the census "true" population percentages than the sample that forms the basis for the model. This is likely explained by the regularity of the spatial pattern of the census data. It is a regular pattern where neighborhoods farther from midtown Manhattan are more likely to commute by car. Since a number of the important variables in the model relate to distance from midtown Manhattan and travel time to work, it makes sense that the model would predict a similarly regular spatial pattern of commute mode choices. This is despite the fact that the model is actually based on the somewhat less regular sample data illustrated in Figure 4.2.

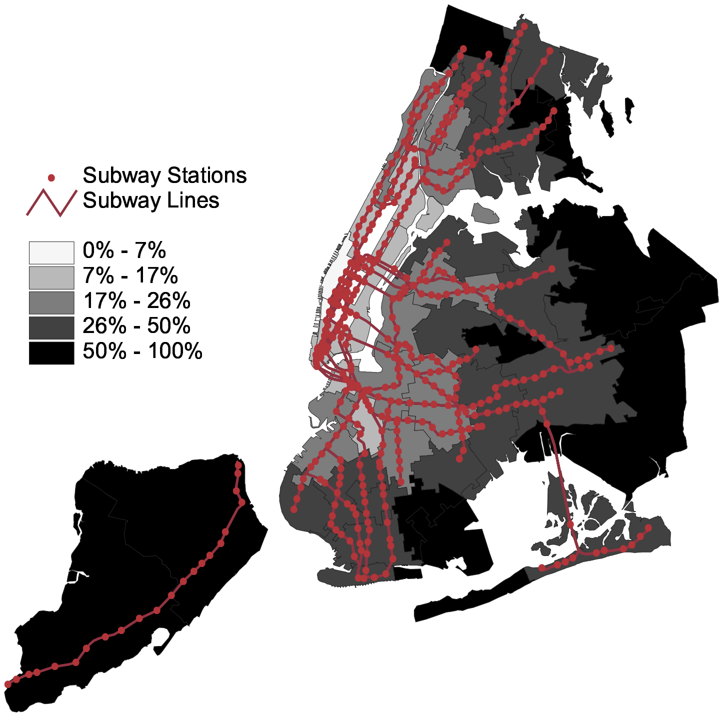

Figure 4.1: 2000 Census Percent of Commuters Using Cars in NYC Neighborhoods

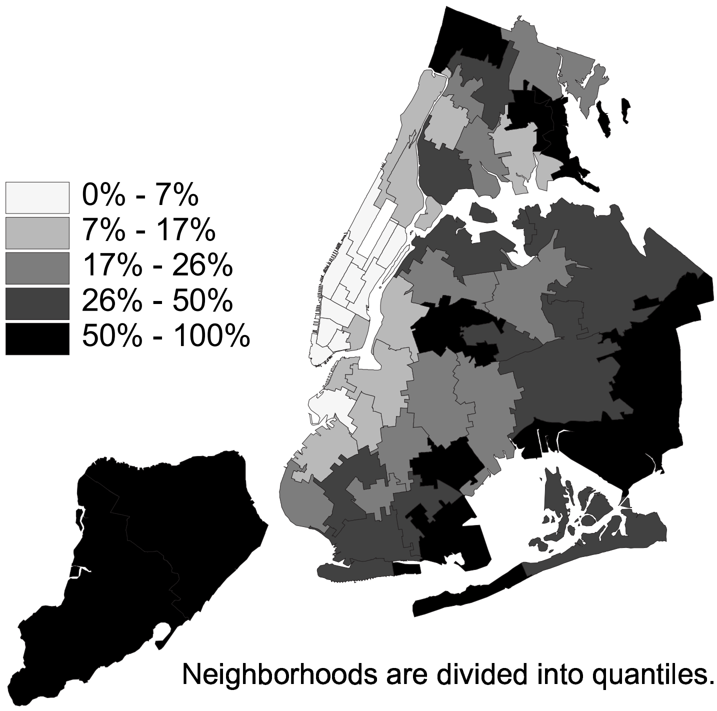

Figure 4.2: Sample Percent of Commuters Using Cars in NYC Neighborhoods

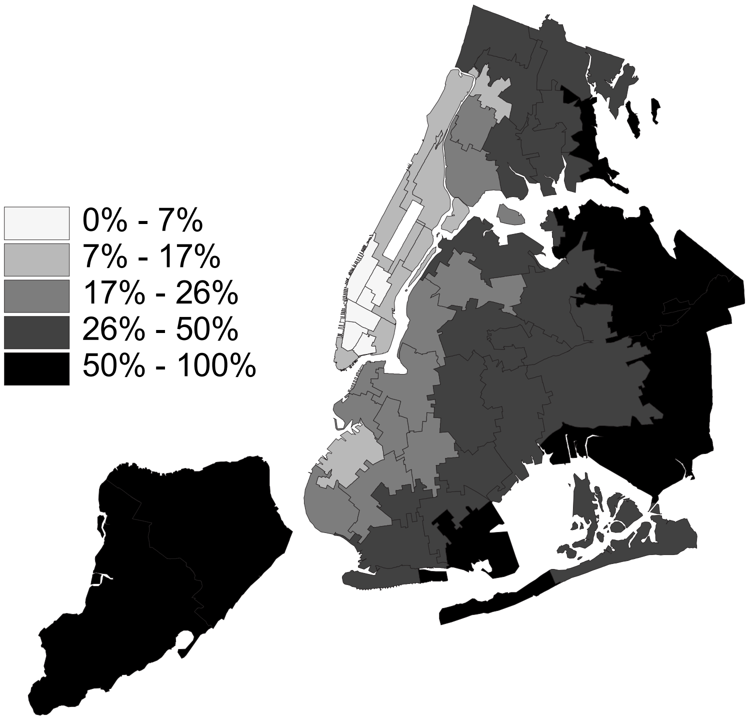

Figure 4.3: Predicted Percent of Commuters Using Cars in NYC Neighborhoods

Figures 4.4 through 4.6 depict data from the same three sources, this time focusing on transit as the commute mode choice. Again, the map that is based on 2000 Census data is overlaid by the subway line map, this time to illustrate the strong positive correlation between subway line availability and choosing transit for the commute.

Here, it appears that both the sample and the model based on that sample produce less accurate results than those regarding the car commute mode. Still, I would argue that the basic pattern is consistent, with the outlying areas of the city having lower transit use for commuting than the more central areas.

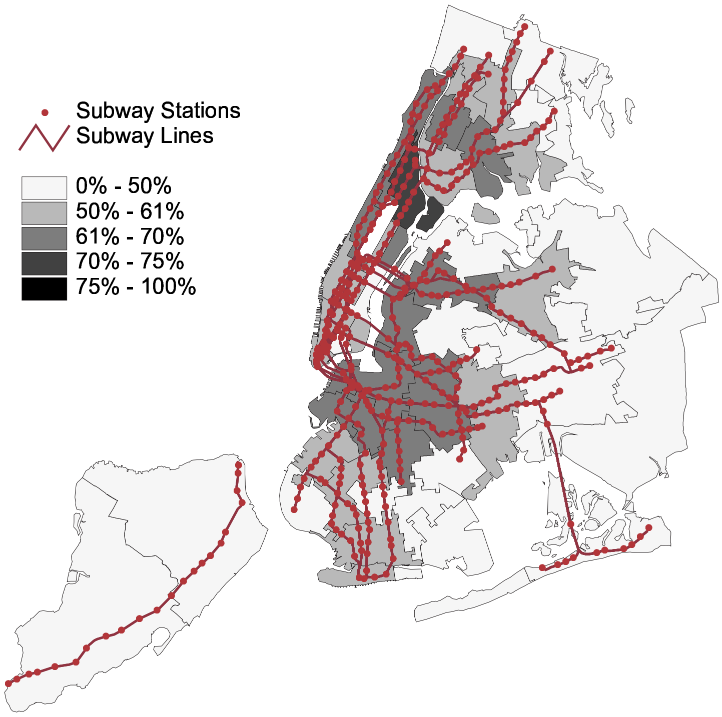

Figure 4.4: 2000 Census Percent of Commuters Using Transit in NYC Neighborhoods

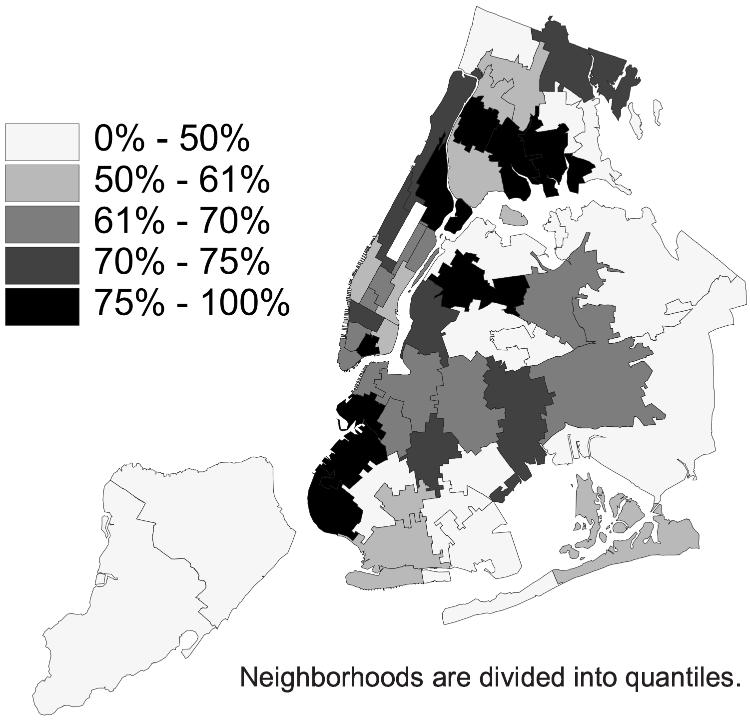

Figure 4.5: Sample Percent of Commuters Using Transit in NYC Neighborhoods

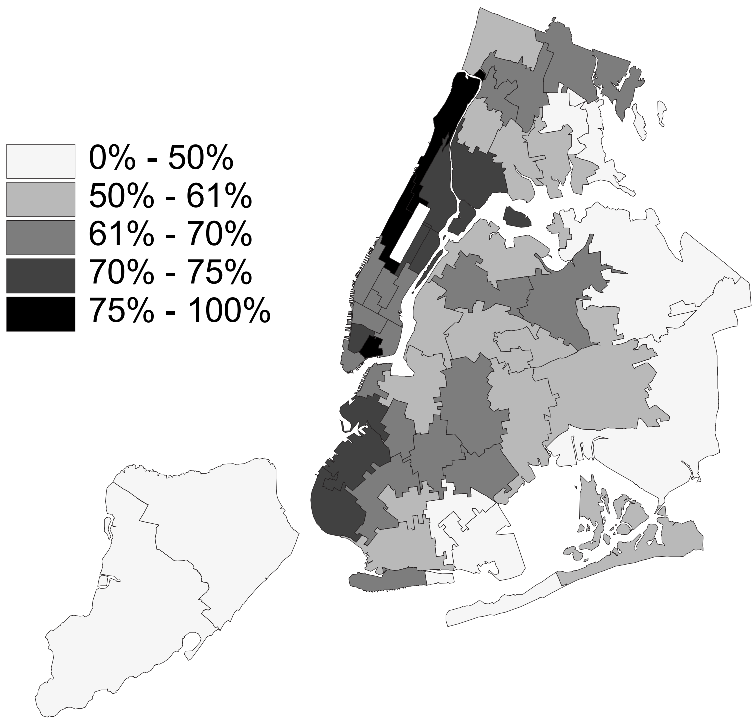

Figure 4.6: Predicted Percent of Commuters Using Transit in NYC Neighborhoods

In the final set of map comparisons to spatially validate the model's ability to predict commute mode choice, Figures 4.7 through 4.9 illustrate the predicted percentages of commuters who walk to work. As in the case of predicting car commuting, the model seems to predict the 2000 Census better than the RT-HIS data that it is based on. Again, this is likely to be because the model tends to have a spatial smoothing effect. As in the case of commuting by car, the actual pattern of the choice to walk to work is a spatially smooth pattern that appears to radiate outward from midtown Manhattan.

Overall, the maps in this section have illustrated that the model does indeed predict spatial heterogeneity in commute mode choices that corresponds to that observed in the real world. The correspondence is strongest for the car and walk mode choices, but is present for the transit mode choice as well.

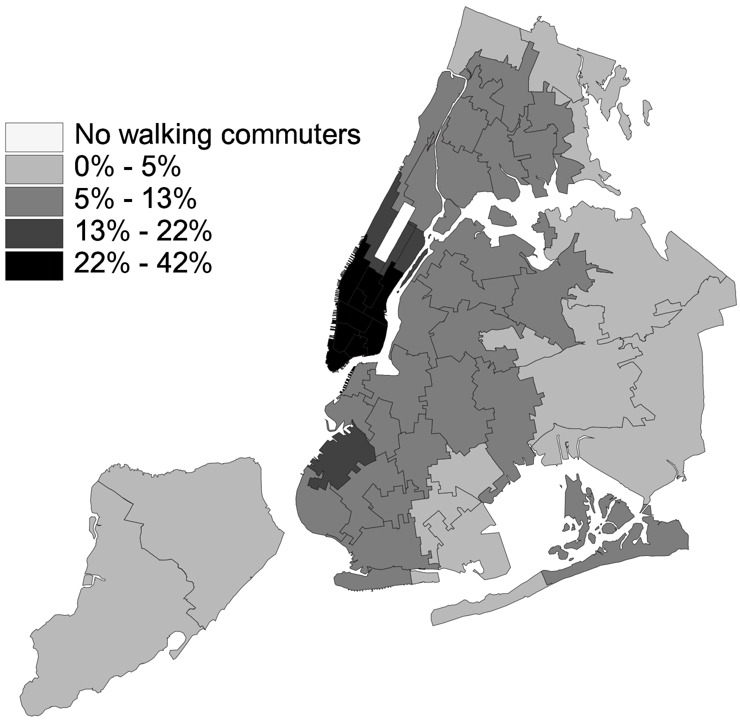

Figure 4.7: 2000 Census Percent of Commuters Walking in NYC Neighborhoods

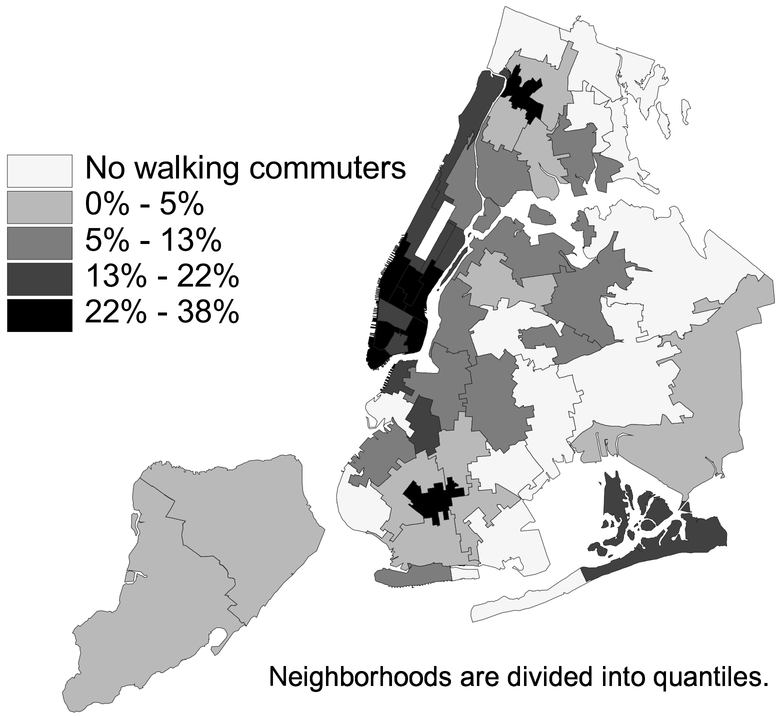

Figure 4.8: Sample Percent of Commuters Walking in NYC Neighborhoods

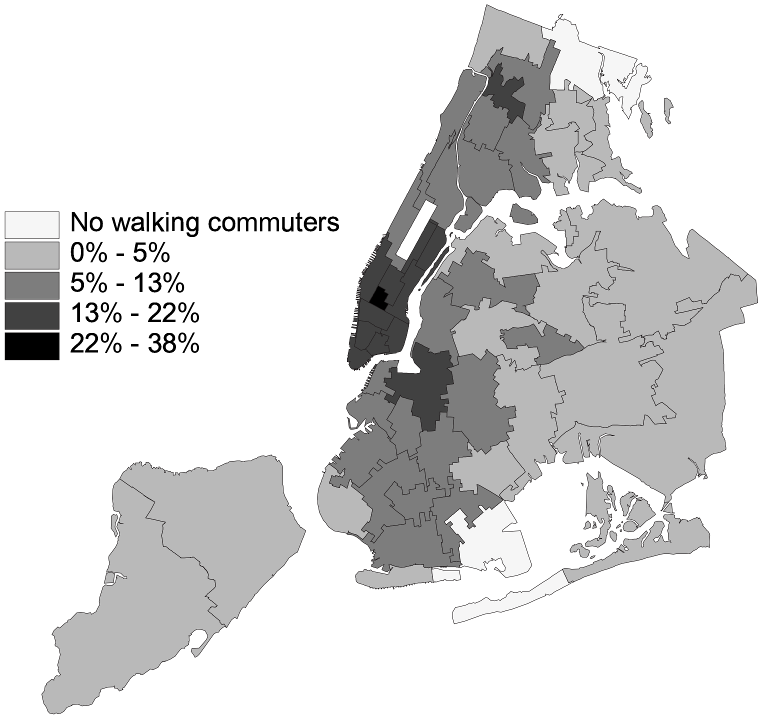

Figure 4.9: Predicted Percent of Commuters Walking in NYC Neighborhoods

4.2 Spatial validation of the model: The choice of car ownership status

Turning to the choice of car ownership status, Figures 4.10 and 4.11 display the actual number of cars per household as reported by the 2000 Census and in the portion of the RT-HIS dataset that was used for the modeling in Chapter 3 of this dissertation. These maps cannot be directly compared to the model results because the model predicts the percentage of commuters in each neighborhood who live in households with zero, one, or two-or-more cars. The 2000 Census, on the other hand, provides household vehicle ownership information in the aggregate - the number of available vehicles in each census tract. To create Figure 4.10, this number was divided by the number of occupied housing units to obtain the average number of available vehicles per household by neighborhood. The number of vehicles per household is one of the variables in the RT-HIS dataset, so Figure 4.11 was created to provide a direct comparison to the Census.

The RT-HIS dataset appears to be populated across the entire city with households that have more vehicles than are reported in the Census. This could be a real result; it could be true that respondents to the RT-HIS are systematically wealthier than the average New Yorker, and therefore have slightly higher car ownership. Another possibility is that this systematic result is the product of the difference in the way that the data were created. In any event, the basic spatial pattern of car ownership in the two data sources is extremely similar.

Figures 4.12 through 4.17 illustrate the direct comparison between the actual car ownership levels in the dataset and the car ownership levels predicted by the model. In the zero-car ownership category (Figures 4.12 and 4.13), the two maps look almsot identical. In the one-car ownership category (Figures 4.14 and 4.15), the correspondence between the sample data and the model's prediction is less perfect. This is likely a reflection of the lack of precision of many of the coefficient estimates in the model that correspond to one-car households. It may be that the choice to own one car in New York City is not well predicted by the model because the factors that determine one-car ownership may not be closely related to the local transportation-land use context. In the two-or-more-car ownership category (Figures 4.16 and 4.17), the model performs well again. The sample data has a somewhat more irregular pattern to it than the percentages predicted by the model, but the overall trend in both maps is the same.

The maps in this section illustrate that the statistical model from Chapter 3 predicts spatial heterogeneity in car ownership status quite well for zero- and two-or-more-car households. The model is less reliable in its prediction of the spatial heterogeneity of the car ownership choices of one-car households.

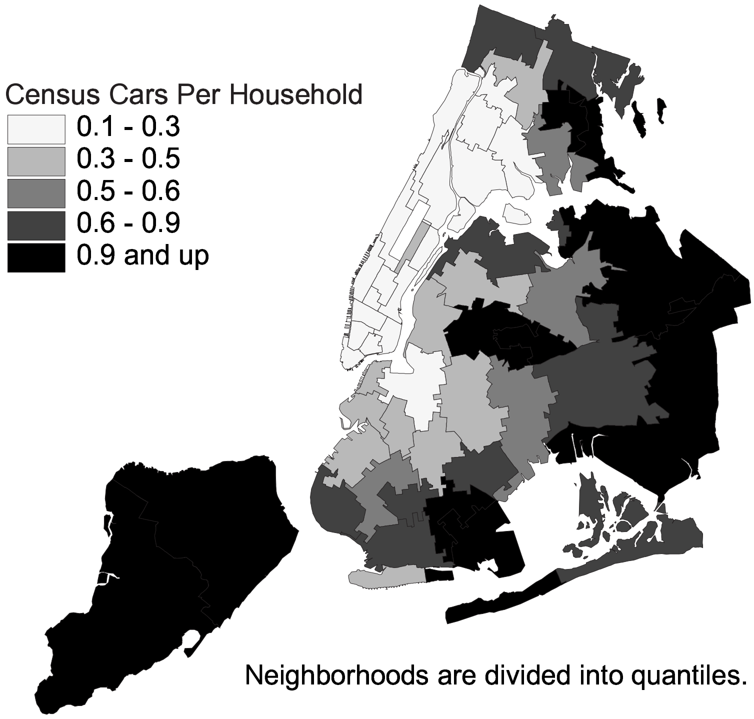

Figure 4.10: Number of Available Vehicles Per Occupied Housing Unit from the 2000 Census

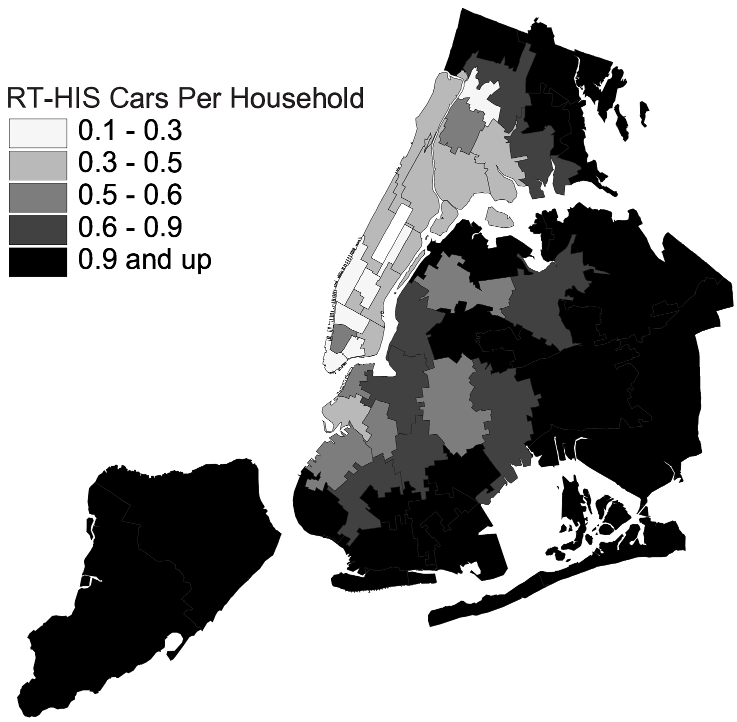

Figure 4.11: Number of Cars Per Household from the RT-HIS Sample Used in Chapter 3

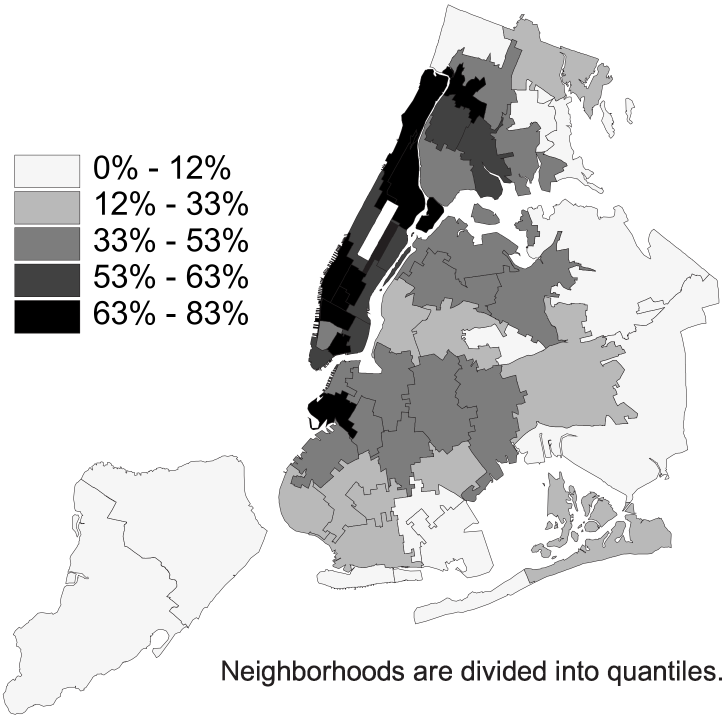

Figure 4.12: Sample Percent of Commuters who live in Car-Free Households in NYC Neighborhoods

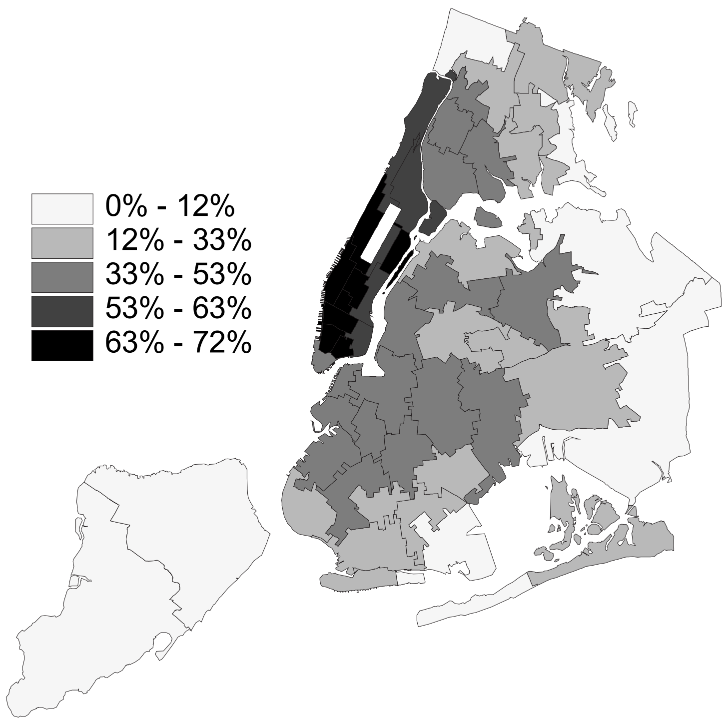

Figure 4.13: Predicted Probability of Commuters living in Car-Free Households in NYC Neighborhoods

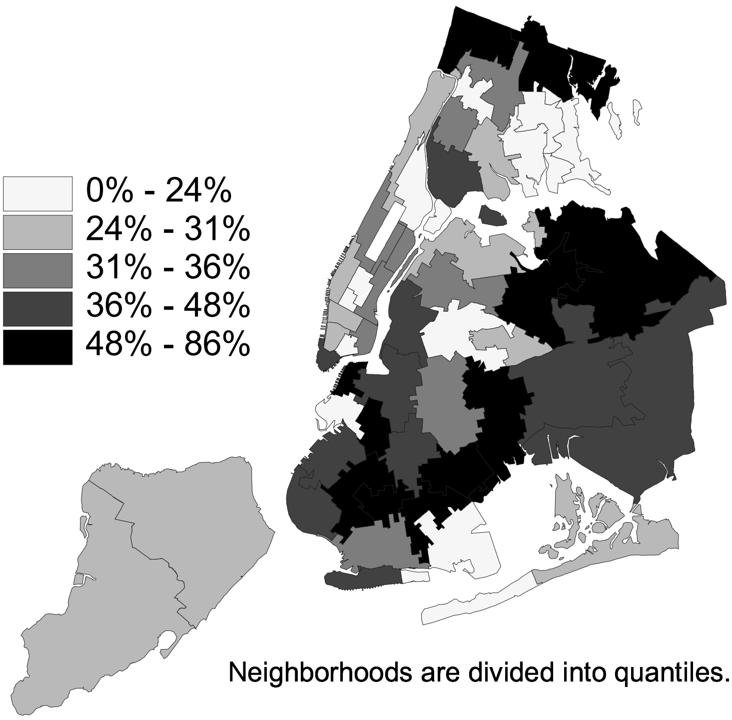

Figure 4.14: Sample Percent of Commuters who live in One-Car Households in NYC Neighborhoods

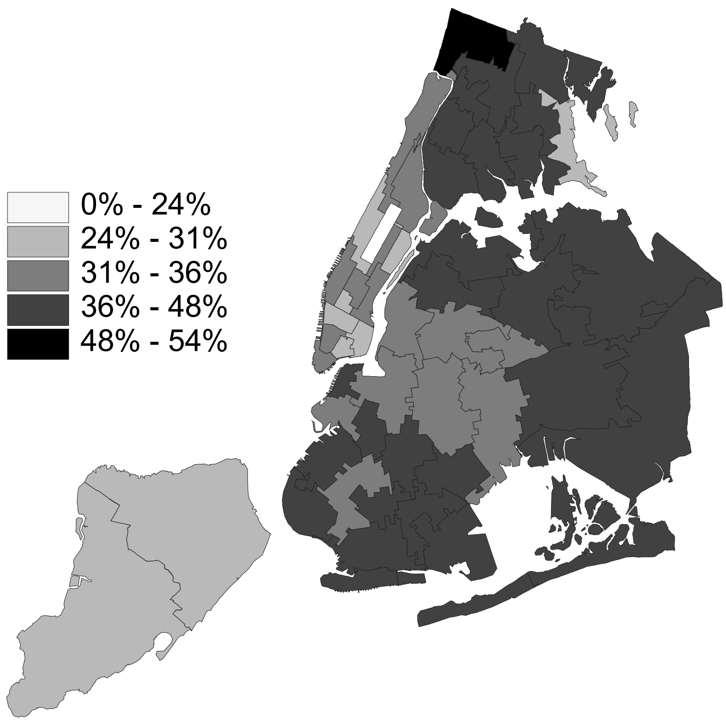

Figure 4.15: Predicted Probability of Commuters living in One-Car Households in NYC Neighborhoods

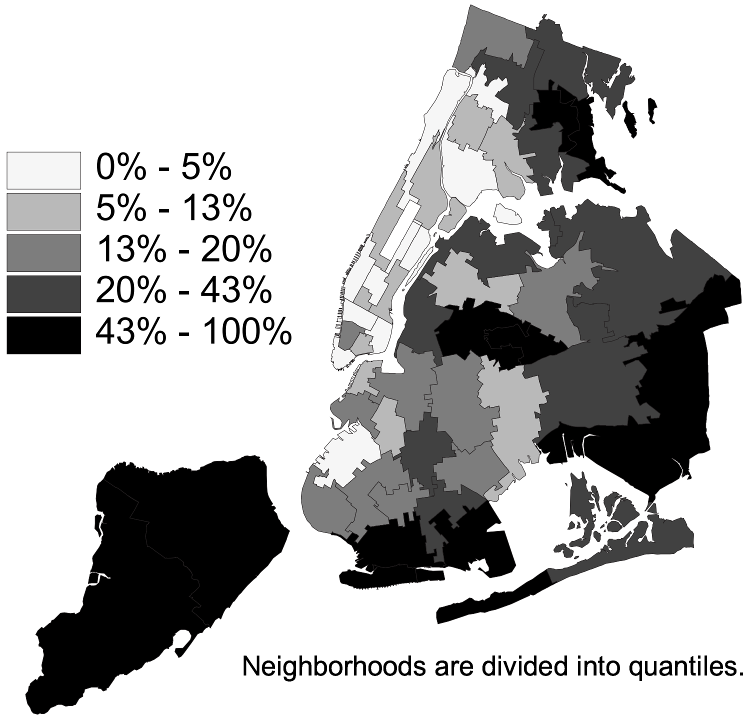

Figure 4.16: Sample Percent of Commuters who live in Two-Or-More-Car Households in NYC Neighborhoods

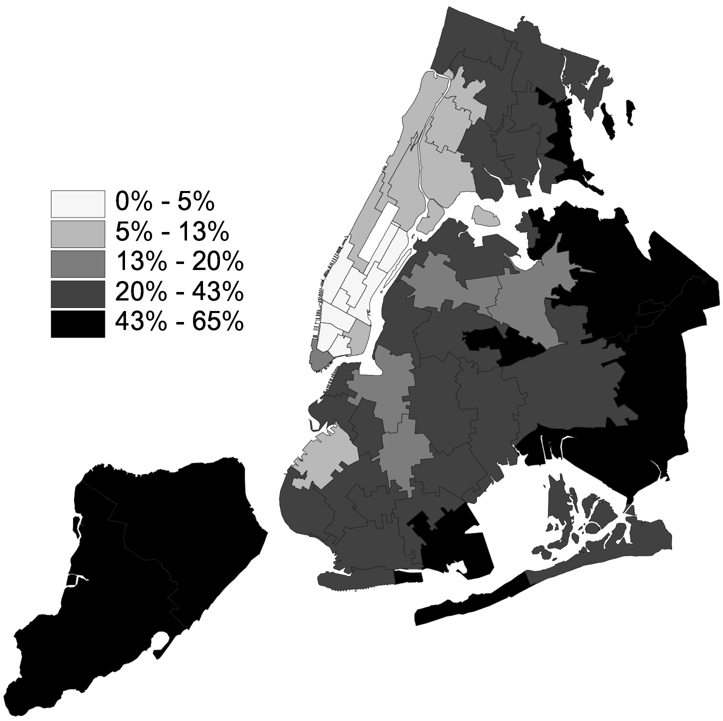

Figure 4.17: Predicted Probability of Commuters living in Two-Or-More-Car Households in NYC Neighborhoods

4.3 Simulations of spatially-uniform changes in independent variables

As mentioned in the introduction to this chapter, even spatially-uniform changes in the transportation-land use system produce spatially-differentiated behavioral results. This section of the chapter explores three spatially-uniform changes in the transportation-land use system in an attempt to better understand the spatial heterogeneity of the likely responses of the residents of New York City.

The first simulated change is a 25 percent increase in the cost of commuting by car. This has more than become a reality in recent times, as gasoline prices have approximately doubled over the past couple of years.

The second simulated change is a 25 percent increase in the time necessary to commute by car. This reflects increasing traffic congestion over time, another phenomenon that is common in many urban areas as both employment centers and residences become increasingly decentralized. This decentralization causes increases in traffic congestion both because people have physically greater distances to travel and because public transportation systems are less effective at providing access in decentralized land use settings.

The final spatially-uniform change simulated here is a doubling of transit headways. This means that buses and trains would arrive at half their current frequency, and transit waiting times would double. With transit ridership as high as it is in New York City, this is not a likely future scenario for the city. However, it is unfortunately representative of a national trend that we have been seeing over the past few decades as transit has been losing market share to the private auto, and it is therefore interesting to look at the effect such a change might have on the commute mode decisions of New Yorkers.

4.3.1 A twenty-five percent increase in car travel cost

As reported in Chapter 3 of this dissertation, results from a spatially uniform analysis conducted using this model indicate that in New York City, the factors that have the largest impact on commuter mode choices are the relative time and money costs of the available modes. When the travel cost for the car mode is increased by 25 percent, the model predicts that the mode share of cars for commuting will fall by 3.2 percent, the mode share of transit will increase by 3.0 percent and the mode share of walking will increase by 0.1 percent.

As illustrated in Figures 4.18 and 4.19, a spatially differentiated simulation shows that this response is far from uniform. Commuters living farther from the central business district are more likely to switch away from commuting by car in response to percent changes in the cost of car commuting. Because transit is the closest substitute for the car for most New York commuters, most of this reduction in car commuting means an increase in transit commuting. As Figure 4.19 shows, the largest increases in transit use for commuting as a result of a 25 percent increase in car commute costs occur in areas farthest from Manhattan's business district. Because walking to work is not a close substitute for commuting by car (at least in New York City!), the changes in walking to work resulting from an increase in car travel cost were extremely small.

4.3.2 A twenty-five percent increase in car travel time

When the travel time for the car mode is increased by 25 percent (perhaps due to an overall increase in traffic congestion), the present model predicts that the mode share of cars for commuting will fall overall by 2.2 percent, the mode share of transit will increase by 2.0 percent, and the mode share of walking will increase by 0.1 percent.

Once again, looking at the results in a spatially disaggregate way clearly shows that this response is not spatially uniform. As car commute time rises, those commuters who live farthest from Manhattan's business district switch away from their cars and toward transit commuting. It is interesting to note that the effect on commute mode choice of increasing car commute time appears to be slightly smaller than that of increasing car commute cost. Again, the changes in walking to work resulting from an increase in car travel time were extremely small.

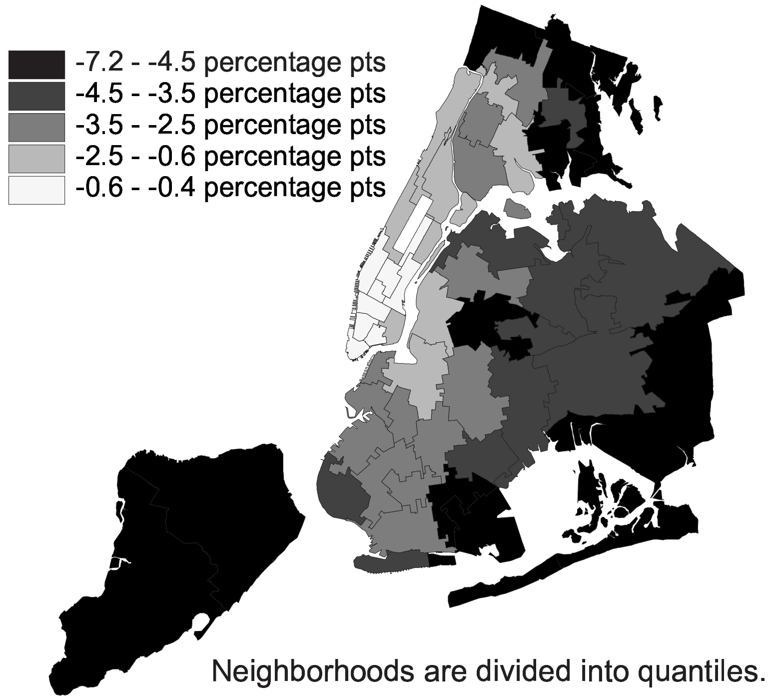

Figure 4.18: Change in Predicted Car Use for Commuting After 25 Percent Increase in Car Commute Cost

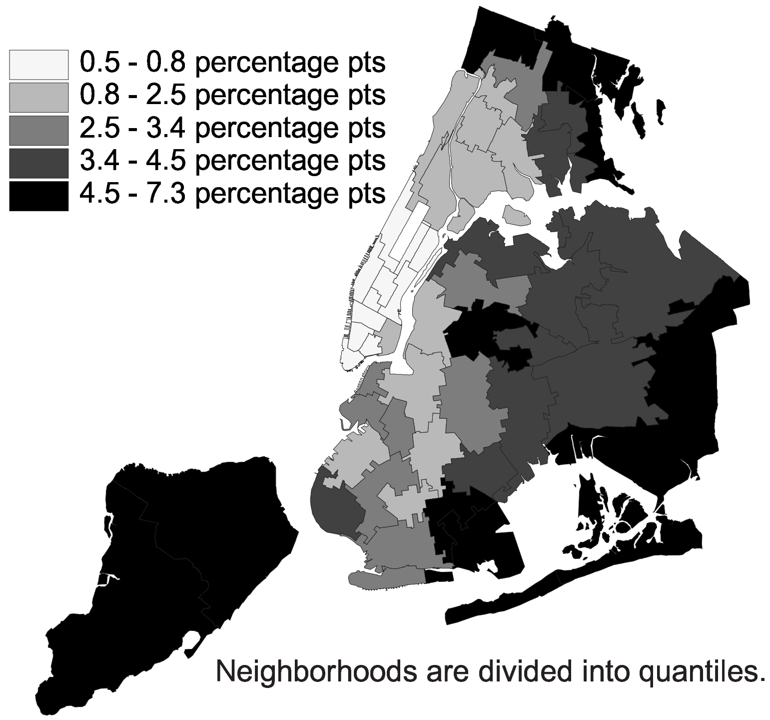

Figure 4.19: Change in Predicted Transit Use for Commuting After 25 Percent Increase in Car Commute Cost

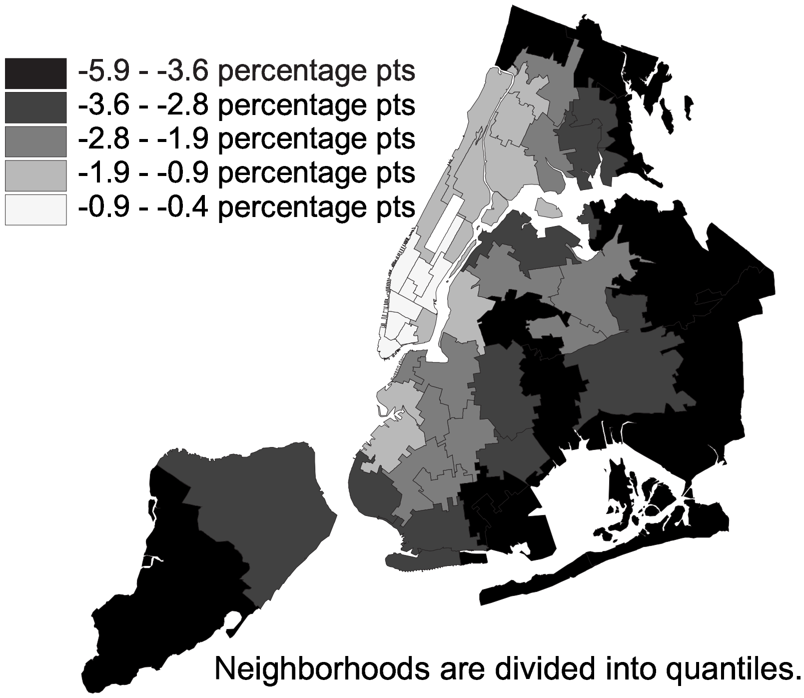

Figure 4.20: Change in Predicted Car Use for Commuting After 25 Percent Increase in Car Commute Time

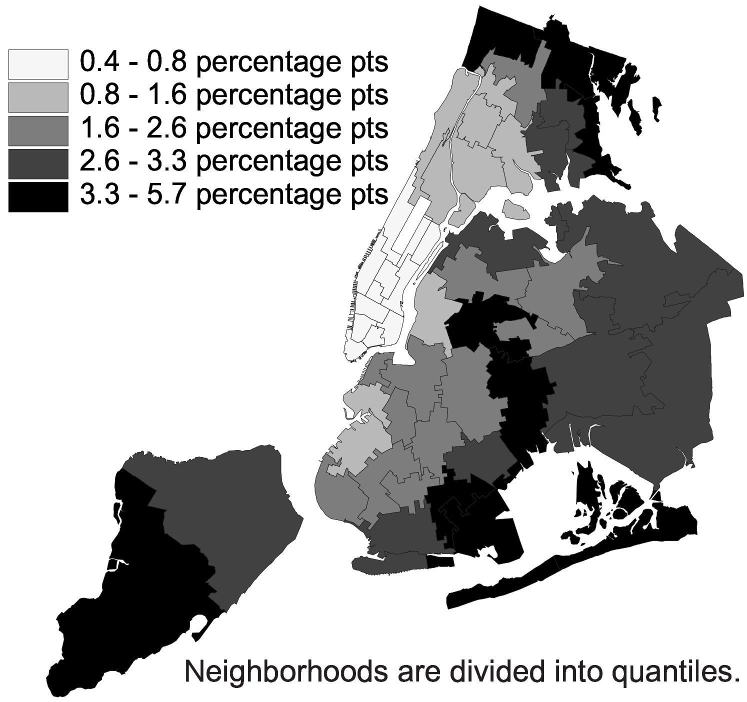

Figure 4.21: Change in Predicted Transit Use for Commuting After 25 Percent Increase in Car Commute Time

4.3.3 A doubling of transit headways

Figures 4.22 through 4.24 illustrate the last of the spatially-uniform changes simulated here – a doubing of transit headways. In the previous two simulations, the change affected the car mode. Because transit is the main substitute for the car, maps illustrating only the changes in car and transit commuting are shown in this chapter.

Here, an additional map is shown that represents the effect of the doubling of transit headways on the choice to walk to work. The difference is that in this case, the change directly affects transit. It turns out that both the car and walking are close substitutes for transit, depending on which part of New York you look at. As depicted in Figure 4.22, the increase in car commuting in the lower half of Manhattan is small despite the fact that this same area is where the largest decrease in transit commuting takes place (see Figure 4.23). The reason for this becomes clear by looking at Figure 4.24, which shows a large increase in the percent of commuters who arrive at work on foot in the lower half of Manhattan.

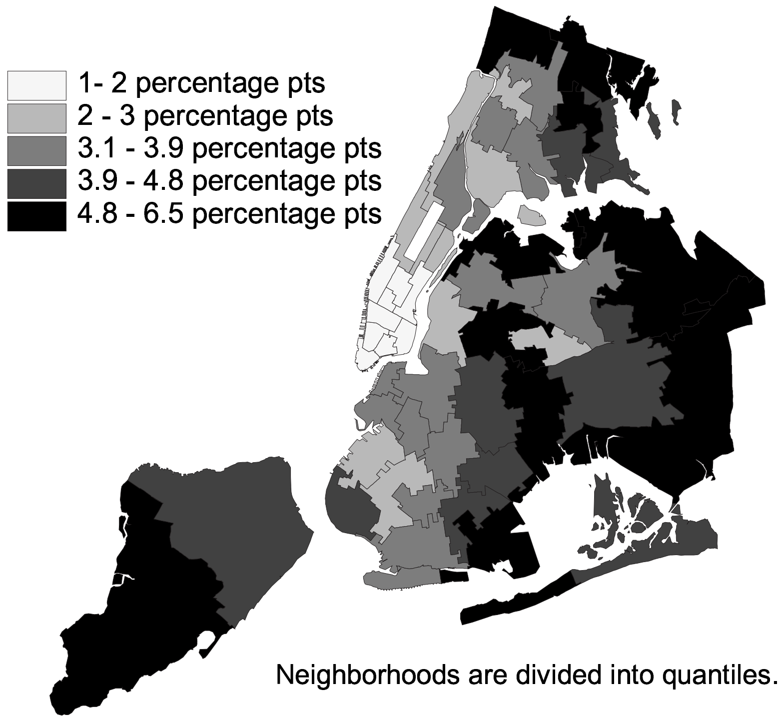

Figure 4.22: Change in Predicted Probability of Commuting By Car After Doubling of Transit Headways

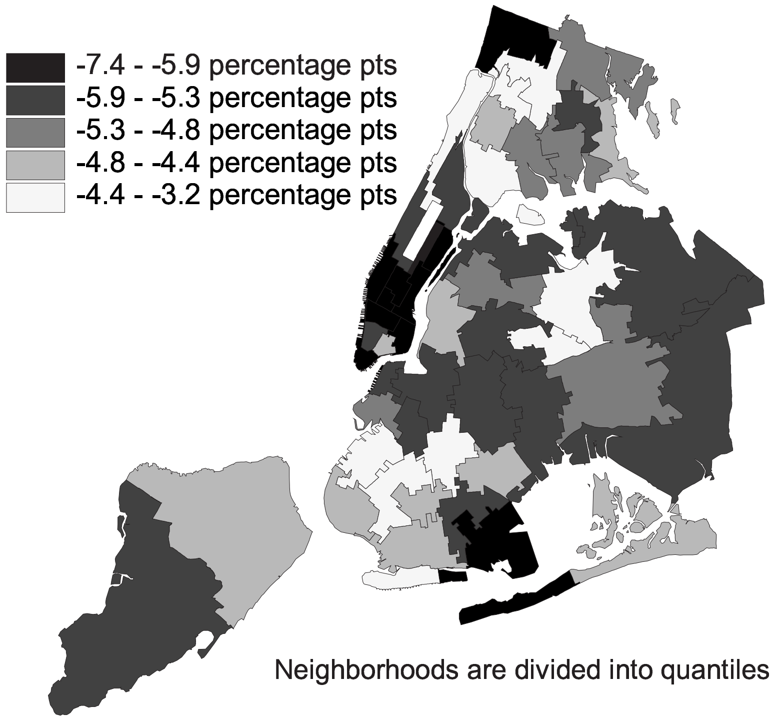

Figure 4.23: Change in Predicted Probability of Commuting By Transit After Doubling of Transit Headways

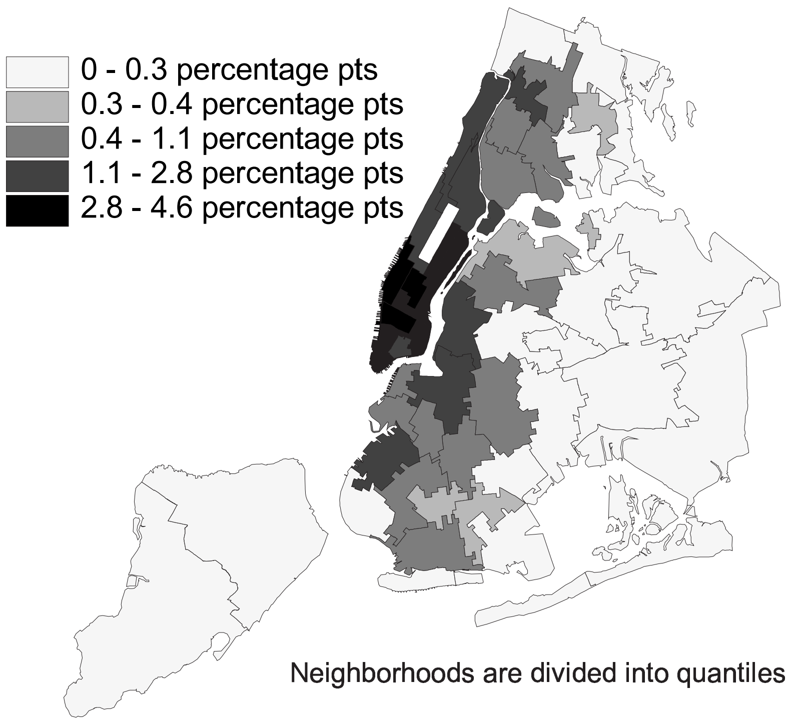

Figure 4.24: Change in Predicted Probability of Commuting On Foot After Doubling of Transit Headways

4.4 Simulations of spatially-differentiated changes in independent variables

Looking at spatially-uniform changes in the transportation-land use system is appealing for its methodological simplicity as well as the relatively straightforward interpretation of its results. However, many important changes to the transportation-land use system in the real world occur in spatially-heterogeneous ways. This section of the chapter identifies three potentially interesting changes to New York's transportation-land use system that occur heterogeneously across the neighborhoods.

The first set of changes relates to the cost of parking a car at work. In the data underlying the statistical model of Chapter 3, the cost to park a car at work is \$15 per day for the business district of Manhattan and \$0 for all other areas of New York City. While this is clearly a simplification of reality, it does capture the essence of the daytime parking price scheme in the city. Because the cost to park a vehicle in this situation swamps all other costs of commuting, many sustainable cities advocates have suggested raising parking prices at employment centers that are accessible by transport modes other than the car. This scenario is crudely represented by raising parking prices to Manhattan business district levels throughout the city. A second scenario modeled here is reducing all parking prices in New York City to \$0. This is clearly not a scenario that will actually happen - parking prices will remain high in Manhattan. However, since most other employment centers in the United States actually do provide free parking, it is interesting to see how this simulation model predicts New Yorkers would react to such a situation.

The second change modeled here is to continue the linear population density trend for each census tract between the 1990 and the 2000 Census. Population density is an important variable in the present statistical model, contributing to both how New Yorkers choose their residential location and their car ownership status.

The third and final spatially-heterogeneous change modeled in this section is the building of the much-discussed Second Avenue Subway line. Since the 1920's, New York City has been talking about building an additional north-south subway line in Manhattan that runs along Second Avenue. This project has been started and stopped a number of times, and is still under serious consideration by the city.

4.4.1 Playing with parking costs at work

Figure 4.25 illustrates the spatial pattern of the effect of increasing parking charges outside of the outlined "Original parking price area" to Manhattan business district rates. Figure 4.26 illustrates the effect of making parking free at the work location throughout New York City.

The maps in this section represent the data differently than those in the rest of this chapter. Here, individual mode choices are assigned to *work* locations rather than home locations. To interpret these maps, then, the darker neighborhoods indicate a larger change in the probability of commuting by car for commuters who *work* in that neighborhood. The reason for this is that these simulations focus on a change in the parking price at the work location.

Since these maps are oriented to the work locations, it makes sense that there is no change in behavior for those commuters who work in the areas where parking prices did not change. In Figure 4.25, this area is in the business district of Manhattan. In Figure 4.26, this area is everywhere in the city except for the business district of Manhattan. Figure 4.25 shows some variation in the commute mode choice response when parking prices are raised to Manhattan levels. The people who are most sensitive to high parking prices appear to live in a band of neighborhoods that stretches across Brooklyn and Queens. Note, however, that the variation in parking price sensitivity depicted in this map is actually only a couple of percentage points across the city.

It is also interesting to note that in comparing the two scenarios that raising parking costs outside of Manhattan has a larger effect on commute mode choice than making parking free in Manhattan does. This is also to be expected, since both walking and transit are so convenient in Manhattan that even free parking would not entice many people into their cars.

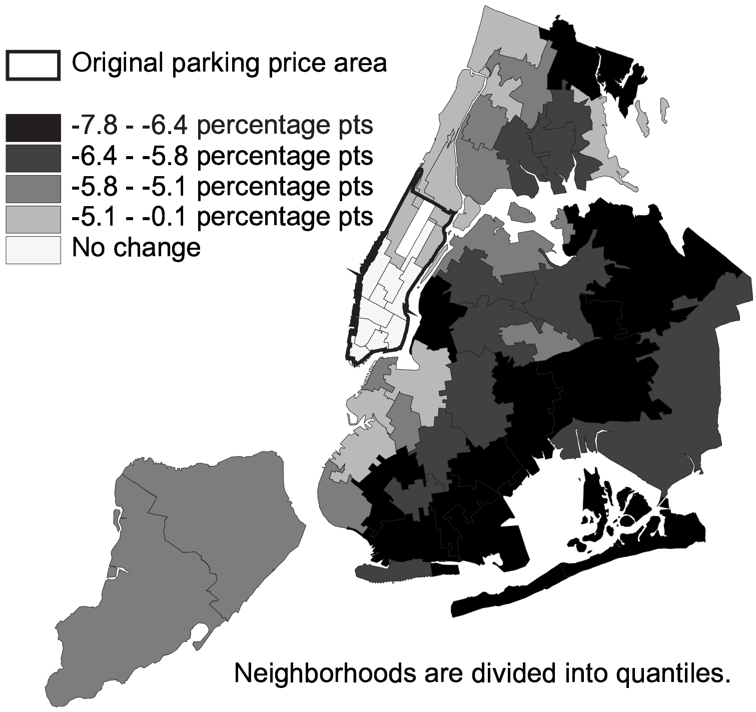

Figure 4.25: Change in Probability of Commuting By Car When Parking Prices At Work Are Manhattan Rates All Over New York City

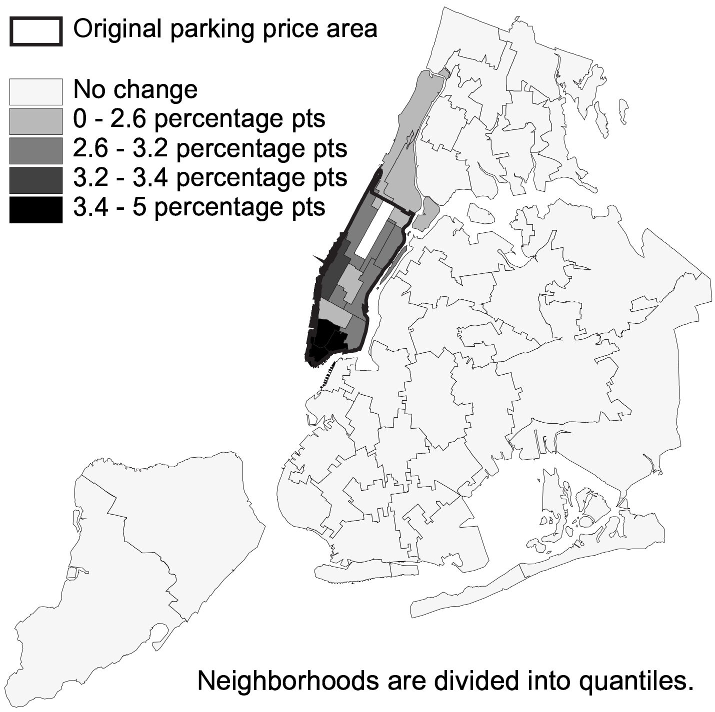

Figure 4.26: Change in Probability of Commuting By Car When Parking At Work is Free All Over New York City

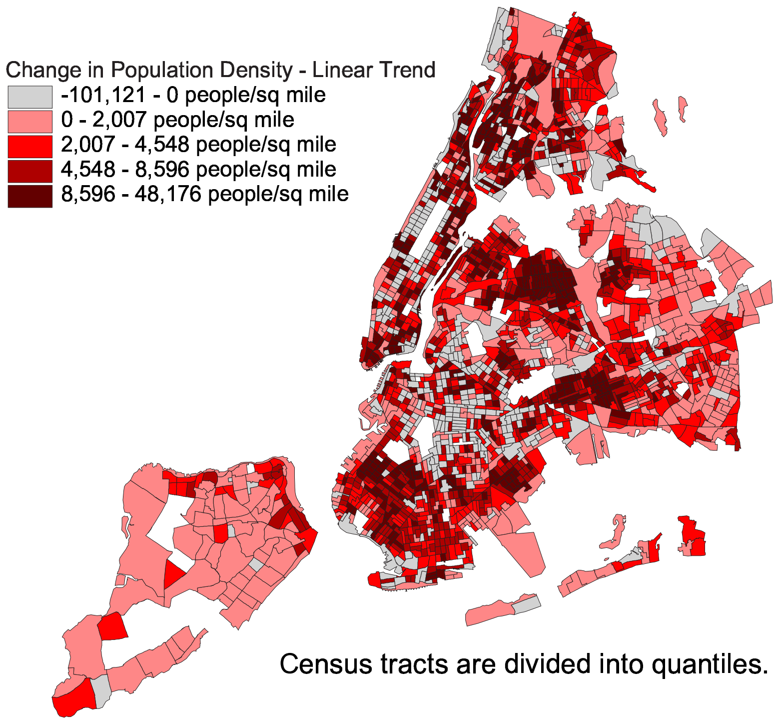

Figure 4.27: Linear Trend in Population Density by New York City Census Tract: 1990-2000

4.4.2 Continuation of a linear population density trend

The maps in Figures 4.28 through 4.30 again are depicted according to residential location. They illustrate the changes in the percent of people living in each neighborhood who will choose to own zero, one, or two cars when a linear trend in population density is continued out to 2010 in each census tract of the city. The population density linear trend is simply the difference in population density between the 1990 and 2000 Census, and is shown in the map in Figure 4.27. All of the maps in this section are in color because in the trend, some areas gain population and other areas lose population, leading to both positive and negative effects across the neighborhoods on the choice of car ownership status as well. Negative changes are shown in grayscale, and positive changes are shown in a red monochrome color scale.

Overall, New York City's trend is one of population growth. This is made clear by the overwhelming number of census tracts colored in shades of red in Figure 4.27. This is reflected also in Figures 4.28 through 4.30. The predicted percent of commuters in zero-car households rises throughout most of the city, the predicted percent

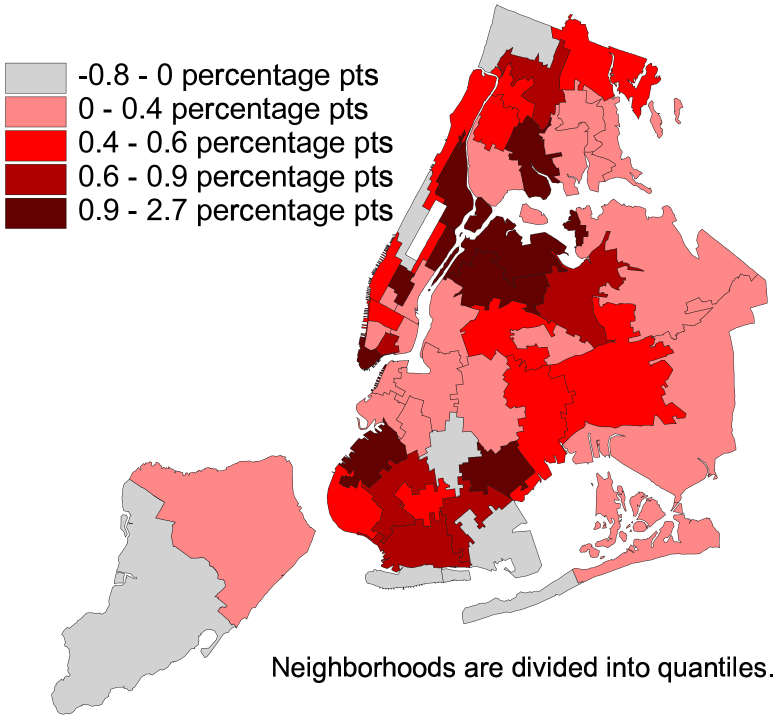

Figure 4.28: Change in Probability of Being in a Zero-Car Household After Continuation of Linear Population Density Trend

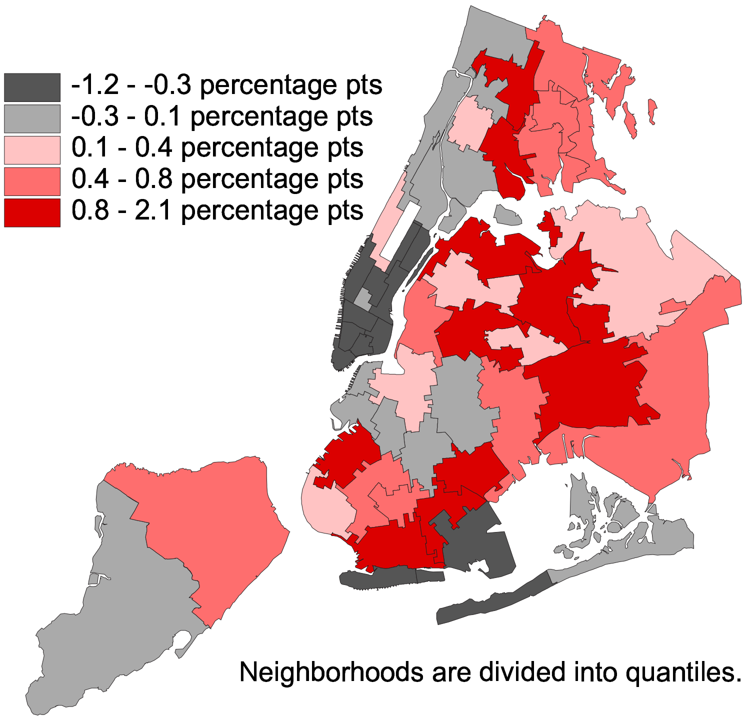

Figure 4.29: Change in Probability of Being in a One-Car Household After Continuation of Linear Population Density Trend

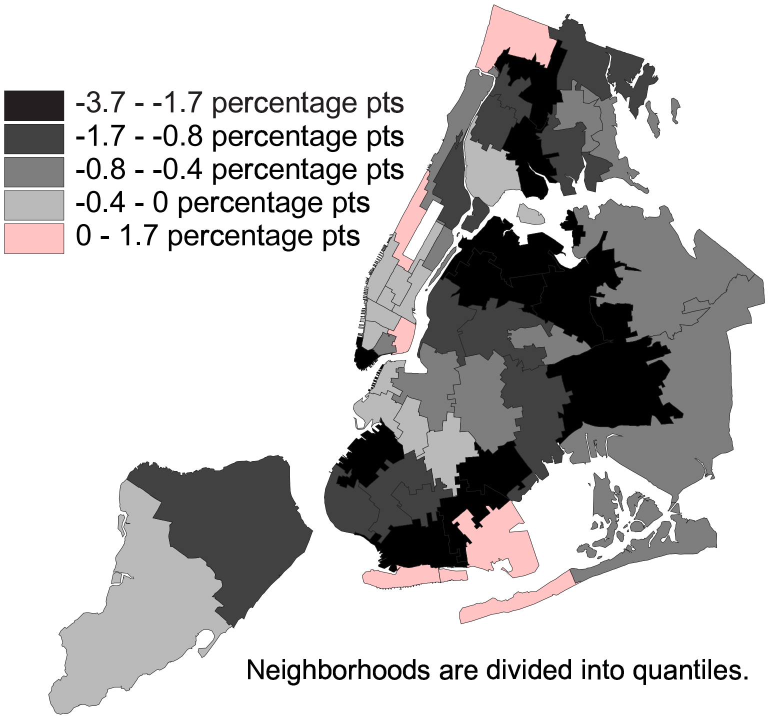

Figure 4.30: Change in Probability of Being in a Two-Or-More-Car Household After Continuation of Linear Population Density Trend

of commuters in one-car households drops in Manhattan and rises throughout most of the rest of the city, and the predicted percent of commuters in two-or-more-car households drops for most neighborhoods in the city.

4.4.3 The building of the Second Avenue subway

The final simulation performed in this chapter is probably the one that will be most interesting for New York planners. Figure 4.31 illustrates the change in the predicted transit commute mode share by home neighborhood that would result from the building of the Second Avenue subway. Predictably, the largest change occurs in the area of Manhattan close to Second Avenue (the east side of Manhattan island on the map). The reason that changes in transit commute mode are predicted outside of this area at all is that there are many work locations in this area, and transit use for commuting is partially influenced (in both the discrete choice model from Chapter 3 and in reality) by transit availability near work.

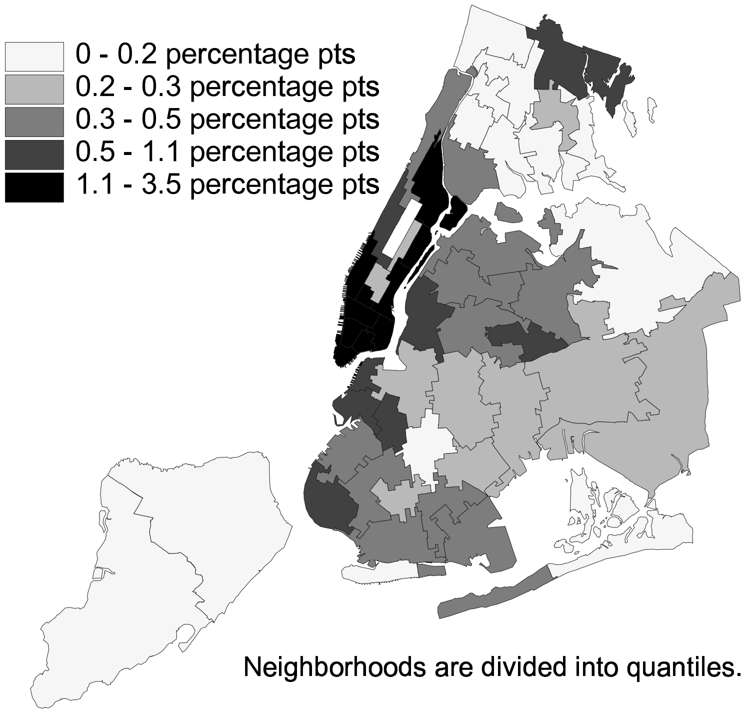

Figure 4.31: Change in Probability of Commuting by Transit After Building of Second Avenue Subway

4.5 Conclusion

This chapter explores the importance of understanding spatial patterns of behavior using case studies of changes in the transportation-land use system of New York City. These spatial simulations are made possible by the model estimated in Chapter 3 of this dissertation together with GIS technology. The maps in this chapter provide convincing evidence that even if a policy is implemented in a uniform manner across New York, there can be high level of spatial heterogeneity in the behavioral response to that policy. To the extent that policy makers care about this spatial heterogeneity in the behavioral response, models that are spatially-explicit such as this one can be useful tools to gain insights into the likely spatial distribution of responses to policies.

Chapter 5

Walk Trips, Car Ownership, and Residential Location Choice in New York City: A Study of the Interrelated Nature of Decisions

Walking behavior, car ownership, and residential location choice are integrally linked. All else equal, the more money people have, the more cars they will buy. The more cars they buy, the less they walk, bike, and use transit. The less they walk, bike, and use transit, the less people will consider local access to services and public transit in their residential location choice. The less urban residents consider local access to services and public transit in their residential location choice, the more spread out our cities become. In this way, the downward spiral of non-motorized transport modes and transit and the upward spiral of car use accelerates. But all else does not have to be equal. In particular, public policies can - and do - vary.

5.1 The Self-Selection Question

In recent years, there has been a growing interest in the relationship between the built environment and physical activity. This interest is fueled by a desire to promote public health, and the observation that people who are more physically active are generally healthier.

The policymaker who would like to encourage walking as a transportation mode to promote public health has a problem. As described above, the close linkage between the choices of non-motorized transport modes, car ownership, and residential location makes it difficult to disentangle the effects of policy on any one of these choices. This problem has been described in the literature as the "self-selection" problem.

Self-selection in the context of this chapter is the idea that perhaps people who choose to live in more traditional neighborhoods (higher density, pedestrian friendly places) and/or not to own cars make these choices because they like to walk. This would mean that the observed correlation between traditional neighborhoods and walking occurs at least partly because natural walkers are attracted to these places rather than because these places are turning non-walkers into walkers. The observed correlation between car ownership status and walking would be at least partly explained by natural walkers choosing lower car ownership than non-walkers.

For the promoter of pedestrian-friendly built environments as a public health initiative, this is bad news. To the extent that self-selection explains the behavioral change among the residents of a neighborhood that has become more pedestrian-friendly, the public health objective would not be accomplished (although the natural walkers who are taking advantage of the new infrastructure sure do appreciate it). Instead of turning non-walkers into walkers, the newly pedestrian-friendly built environment is simply initiating a reorganization of the population, with natural walkers moving into the neighborhood as those who love their cars move out.

This chapter focuses on the choice of walking level in New York City. It proposes a methodology to split the total effect of the built environment on the choice of walking level into the portion that is due to residential self-selection and the portion that is not. This information is crucial for the policymaker who aims to promote physical activity through built environment change.

Further, this chapter shows that a similar methodology can split the total effect of the built environment on the choice of walking level into the portion that will be realized in the short-term (more-or-less immediately after the built environment has changed), and that portion that will be realized only in the long term. This is accomplished through identifying how much of the total effect of built environment change on walking is actually due to the effect of the built environment on car ownership status and residential location, which in turn affect walking behavior. Since changes in both car ownership status and residential location involve high transaction costs, these changes are likely to be long-term changes, and their effects on walking behavior will not be seen immediately after the built environment change.

This chapter uses the explanatory variable population density to represent the built environment, and asks the following three questions:

1. What is the long-term effect of neighborhood population density on the propensity to walk among New York City residents?

2. How much of this long-term effect is likely to be realized in the short-term?

3. How much of this effect can be explained by "self-selection"?

4. How much do results differ between a model that endogenizes the choices of residential location and car ownership status and one that does not?

To answer these questions, this chapter presents the results of two discrete choice models of the choice of walking level that are based on the New York City data set described in Chapter 2 of this dissertation. In one of these models, residential location and car ownership status are treated as endogenous choice variables along with the choice of walking level. In the other, the choices of residential location and car ownership status are taken as exogenous to the choice of walking level. This second model is presented because it is substantially easier to estimate, and as such the comparison between this and the more complex model is relevant to inform future research.

Many researchers model only mode choice, despite the clear conceptual link between mode choice and the choices of car ownership and residential location. Comparison of the current empirical results indicates that, if a researcher is only interested in estimating the full, long-term effect of the built environment on walking levels, this may be a fine approach. Taking the extra effort to make these choices endogenous may not be necessary.[1] If, however, the researcher would like to understand the component parts of the full effect, estimating a model that treats residential location and car ownership status as endogenous is necessary.

The results of the model estimated here indicate that for most areas of New York, approximately half of the total effect of population density on walking level choice can be attributed to residential self-selection. Further, these results indicate that the short-term effect on walking level of an increase in population density will be between one- and two-thirds of the long-term effect.

5.2 Existing Literature

In the past decade, there has been an explosion of research on the relationship between the built environment and travel behavior. Many of these studies have documented the reasonably strong relationship between specific land use patterns and walking behavior (e.g. Cervero and Gorham, 1995; Cervero and Radisch, 1996; Kitamura et al., 1997). Those studies that have used discrete choice models of transport mode choice to explore this relationship have found that when socioeconomic control variables are included in the model, the effect of land use variables on mode choice becomes small (e.g. Rodriguez and Joo, 2004; Cervero and Duncan, 2003). The consensus from this body of work appears to be that land use patterns have a small but statistically significant influence on walking behavior, and that this effect is stronger for non-work trips.

A question that has come out of this literature is to what extent the observed correlation between particular land use patterns and particular travel behavior patterns is a product of "self-selection". A number of researchers have employed various strategies in their attempts to identify which of these causal structures is most important. Three strategies stand out as most prominent: using longitudinal data for which there has been a change in the environment, using attitudinal survey questions

[1] It is a question for future research whether this empirical result is generalizable to other situations.

in addition to travel diaries and using this information to control for preferences, and comparing populations that have similar socioeconomic characteristics, but live in different neighborhood types. Here, I summarize a few of those studies that are most relevant to the present work.

The first group of studies is those that examine paired neighborhoods (e.g. Cervero and Gorham, 1995; Cervero and Radisch, 1996). These studies tend to find stronger direct influences of land use patterns on travel behavior than those employing other methods, but perhaps this is because self-selection is only partially accounted for by the socioeconomic characteristics of the residents.

Krizek (2003) aims to test for self-selection by using longitudinal data from households who moved from one neighborhood to another during the survey period. He finds that travel behavior measures such as vehicle miles traveled are affected by urban form changes, but that the effect on mode choices is less clear. This could be because Krizek's data comes from the Central Puget Sound region, where alternatives to the automobile as a transport mode are not prevalent in most neighborhoods.

Other studies have incorporated attitudinal variables into statistical models of travel behavior, aiming to move the self-selection portion of the effect from the land use characteristics onto these attitudinal variables (Kitamura et al., 1997; Bagley and Mokhtarian, 2002; Handy et al., 2006). These studies find that once attitudes are explicitly included, they tend to explain most of the variation in travel behavior, and the estimated direct effect of land use patterns on travel behavior is either virtually non-existent (Kitamura et al., 1997; Bagley and Mokhtarian, 2002) or small (Handy et al., 2006).

The current study adds to this literature in two ways. First, previous studies have used data from areas that are largely car-dependent and that have few neighborhoods that are truly high density. The data used in this chapter were collected in New York City, where both walking levels and neighborhood density exhibit high variation. The mean population density of census tracts in this data set is more than 50,000 people per square mile, and the standard deviation of this mean is close to 40,000. Close to 30 percent of the sample used here walked for half or more of their trips on the travel survey day, but almost 60 percent did not take a single walk-only trip on that day. This high variation in both the dependent variable and the independent variable most relevant to this study allows for a particularly robust estimation of the relationship between walking behavior and land use pattern.

Second, the present study of the relationship between land use and travel behavior is unusual because it models both car ownership status and residential location choice explicitly. This allows – for the first time that I am aware of – for quantification of the relative contributions of the self-selection effect and the direct effect of land use patterns on walking choice.

5.3 Data

People in New York City walk more - both longer distances and more trips - than people in any other major metropolitan area in the United States. According to the 1995 Nationwide Personal Transportation Survey, New Yorkers reported walking an average of 1.7 miles each day, while the average for the rest of the country was only 1.2 miles walked per day. As far as trip numbers, 41 percent of trips in Manhattan and 31 percent of trips in New York City overall were walk trips. In the rest of the country, this figure was less than 5 percent. Residents of New York City are also less likely to own cars than other Americans. In Manhattan, 65 percent of households do not own cars, compared with 38 percent in the other boroughs of New York City, and only 3 percent in the United States outside of the New York metropolitan area (U.S. Department of Transportation, 1997). Figures 5.1 and 5.2 represent the shares of walking level and car ownership in the sample used for this chapter's analysis.

As in Chapter 3 of this dissertation, the main data source used here is the Regional Travel - Household Interview Survey (RT-HIS). Because this chapter includes all adults in the sample and is not restricted to commuting adults, the sample size is larger than that of the Chapter 3 models, including 4,382 individuals who reside in the five boroughs of New York City.

The portion of the data set that is used to estimate the model sections of

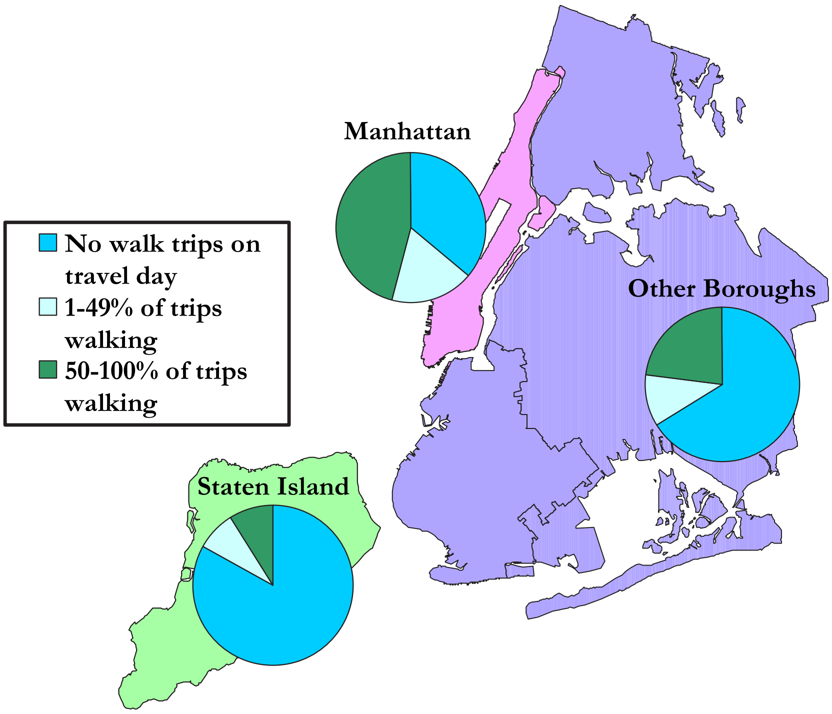

Figure 5.1: Distribution of Walking Level in Sample

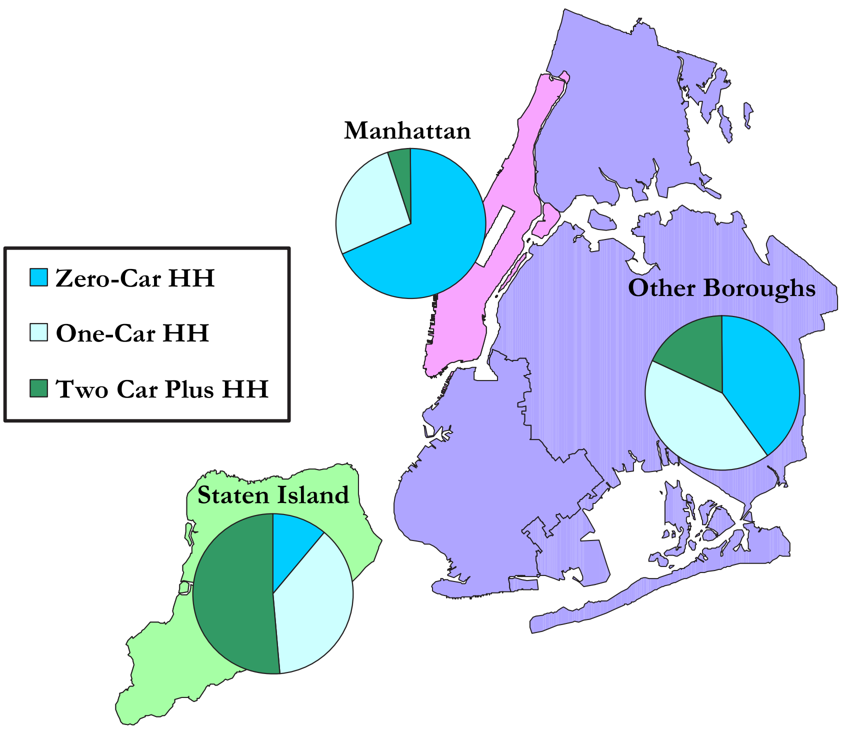

Figure 5.2: Distribution of Car Ownership in Sample

car ownership status and residential location was created exactly as described in Chapter 2 of this dissertation. The portion of the data set that is used to estimate the walking level section of the models in this chapter requires a bit more description. The dependent variable of main interest in this chapter is walking level. This variable indicates whether the percent of all trips made on the assigned survey travel day by each individual that were walk trips was zero, 1 to 49 percent, or 50 percent or more. Note that a walk trip is defined here as a trip completed entirely on foot. Trips that may have had a walking component - such as transit trips - are not counted here as walking trips. Note also that this variable accounts only for the number of trips; there is no accounting for distance walked.

Mode choice-style studies of the choice to walk will often include explanatory variables such as trip distance, the level of street connectivity, and the presence of sidewalks. The models estimated here are different because the dependent variable is the percent of trips that are walk trips rather than a choice between modes for a given trip. This unconventional dependent variable was chosen in an attempt to model the choice to walk for all trips, and not single out specific trip purposes. This different dependent variable makes many of the commonly-used explanatory variables for walk mode choice models not useful.

For this reason, the models estimated here include variables that are themselves representative of the set of trips taken on the travel day rather than being representative of a particular trip. These variables include the total number of trips taken on the travel day, the percent of trips for each trip purpose, the percent of trips taken at night (between 8pm and 6am), and the average number of travelers for the day's trips. In addition, there is one variable that represents the travel day – "Rain on travel day" – and three variables that represent characteristics of the home neighborhood of the traveler – "Home retail density", "Miles from home to midtown", and "Home population density".

Table 5.1 summarizes the distribution of the choice of car ownership status and walking level in the sample used in this chapter's analysis. Because much of the results section of this chapter focuses on the independent variable of population density, the

Table 5.1: Shares of Walking Level and Car Ownership in Current Dataset

	NYC	**Manhattan**	**Other Boros**	**Staten Island**
Number of Households	2828	1297	839	692
Car Ownership				
0	42%	69%	38%	10%
1	40%	26%	43%	37%
2	18%	5%	19%	52%
Number of Adults	4382	1814	1356	1212
Walking Level				
Zero walk trips on travel day	63%	38%	67%	83%
1-49% of trips walking	11%	17%	10%	8%
50-100% of trips walking	26%	45%	23%	9%
Average Population Density (thousands)	62	113	53	13

average population density for each area of the city is also given. The main things to note in Table 5.1 are that there are many individuals in the sample who walk for at least half of their trips, and there is substantial variation in the distribution of walking level, car ownership, and population density even within New York City.

5.3.1 Methodology

Two discrete choice models are presented in this chapter, both attempting to explain people's choice of walking level. The first is a multinomial logit model of the joint choice of walking level, car ownership status, and residential location. The second is a multinomial logit model of the choice of walking level, taking both car ownership status and residential location as exogenous. The full estimation results are provided in Tables 5.2 and 5.3, and a full description of the multinomial logit model can be found in Chapter 2 of this dissertation.

As in the model in Chapter 3 of this dissertation, the joint choice model has a compound choice set. This choice set includes 3 walking level alternatives, 3 car ownership status alternatives, and 2200 residential census tract alternatives. Again, the compound choice set is unmanageably large, with 3x3x2200 = 19,800 alternatives!

To reduce this choice set to one that is computationally manageable, I once again follow McFadden (1978) by taking a sample of the alternatives to be the choice set in the statistical model. The sampling methodology used here is identical to that used in Chapter 3. First, the chosen alternative was set aside for each individual to ensure that every individual's choice set included his or her actual choice. Then, 10 additional census tracts were randomly sampled for each person, making 11 home census tract alternatives in each choice set. Finally, the compound choice set was created for each commuter that included 9 walk level-car ownership status combinations and 11 home census tracts, making a modeled compound choice set of 99 alternatives.

The model presented in Table 5.3 is a simpler model of walking level that takes both car ownership level and residential location as given. This model is presented here because, being a single choice model, it is more typical of models estimated in the literature. It also presents an interesting comparison to the full three-choice model. The explanatory variables included in this model are identical to those included in the "Walking level choice" section of the full joint choice model, plus a few variables that explicitly control for residential location and car ownership level.

5.3.2 Elasticities

Recall that an elasticity in a discrete choice model is defined as the percent change in the probability of choosing a particular alternative when a particular independent variable is increased by one percent. Since the estimated probabilities for the alternatives can be viewed as estimates of the market shares of the alternatives, these elasticities can be interpreted as the change in the market shares of the alternatives that arise from a one percent change in an independent variable. The elasticities presented in this chapter are probability-weighted elasticities, calculated according to the methodology described in Chapter 2 of this dissertation.

The results of only two choice models are presented in this chapter, but there are four tables of calculated elasticities. The two additional tables of elasticities (Tables 5.5 and 5.6) are based on conditional estimated probabilities from the joint choice model, rather than the unconditional estimated probabilities.

For Table 5.5, this means that the probabilities for each individual's actual choices of residential location and car ownership status were extracted from the probabilities for the full set of 99 compound alternatives. For each individual, this extraction process yields only three probabilities - one for each walking level at their chosen residential location and car ownership level. These probabilities are then scaled so that they sum to 1, making them conditional probabilities. From that point on, the elasticity calculation methodology is exactly as described in Chapter 2. The elasticities presented in Table 5.6 were calculated using an analogous methodology, this time based only on the probabilities for each individual for alternatives that included their actual residential location.

5.3.3 Confidence intervals

The confidence intervals that appear in the elasticity tables in this chapter were simulated. The reason that simulation was necessary is that there are nine different explanatory variables in the full joint model that represent population density. The population density within each compound alternative is constant. The nine variables are included because each represents population density interacted with either a sub-choice such as walking level or car ownership level or with a characteristic of the traveler such as income. Whenever there are multiple explanatory variables that are involved in the calculation of a result such as an elasticity, it is necessary to take into account both their individual coefficient variances and their covariances with each other.

The simulation process is straightforward, and is a simplified version of that suggested by Hensher and Greene (2001) in section 4.10 of their working paper. First, the model is estimated and the variance-covariance matrix of the parameters is extracted. From this matrix, the 9x9 portion that corresponds to the population density explanatory variables is separated. This 9x9 matrix is called M. The next step is to take the Cholesky factor of M. Taking the Cholesky factor is akin to taking the square root of a matrix. This matrix is a triangular 9x9 matrix and is called L. This matrix L is used along with a random number generator to create random draws from the joint distribution of these nine estimated population density model parameters. The following is the expression used.

$$\hat{\beta} = \beta + Lu$$

where: $\hat{\beta}$ is the 9x1 vector of simulated population density parameter estimates,
β is the original 9x1 vector of population density parameter estimates,
L is the 9x9 Cholesky factor of the portion of the variance-covariance matrix of parameters that corresponds to the population density parameters, and
u is a 9x1 vector of random standard normal deviates.

Using each simulated set of population density parameters $\hat{\beta}$, elasticities are calculated. In the case of this analysis, 100 sets of parameter estimates were simulated and the resulting elasticities were calculated. The 95% confidence intervals listed in Tables 5.4 through 5.8 are calculated to be 1.96 times the standard deviation of these samples of 100. Likewise, the 90% confidence intervals in these tables are 1.645 times the sample standard deviations.

5.4 Results

The estimated coefficients for the full model estimated here are given in Table 5.2, and those for the single-choice model are in Table 5.3. In both the single-choice model of walking level choice and the section of the joint model that largely explains walking level choice, the signs of the statistically significant estimated coefficients that are characteristics of the trips made on the travel day are as expected. The more trips taken by an individual, the more likely it is that some of them are walk trips. This is reasonable because many of the trips made on foot are discretionary trips. When a person makes a lot of trips in one day, more of those trips are likely to be discretionary trips, and therefore she is more likely to walk for at least some of them. Rain on the assigned travel survey day reduces the likelihood of a person walking for more than half of her trips. Very few people like to walk in the rain. For the variables that indicate percent of trips by trip purpose, all of the coefficients are negative for the low walker category and all except “Percent shopping trips” are negative for the high walker category. This makes sense because the category that is left out of the model is the percent of trips that are made for some “other” purpose. A mentioned earlier, many walk trips are discretionary trips, and often cannot be categorized into any of the seven trip purposes identified as variables in the model. A higher average number of travelers taking the trips together reduces the chances of being in the high walker category in the joint model. This could be because when people travel in groups, cars are often used. The more a person travels at night (from 8pm to 6am), the less likely she is to walk. This is likely to be both because it is colder at night than during the day and for personal safety reasons.

The estimated coefficients on the three variables explaining walking level that are characteristics of the residential location are also as expected. Higher home retail density increases the likelihood of being in the high walker category, probably because there are more retail and services destinations within walking distance of home. The farther an individual lives from midtown Manhattan, the less likely she is to walk. This also makes sense because Manhattan is a common New York destination regardless of where an individual lives in the city, and those living farther from Manhattan are likely to use transit rather than walk to make the trip. Finally, higher home population density encourages more walking. This variable is likely to be capturing some of the retail density effect since the two are highly correlated. It is also possible that areas with higher population density encourage walking because there are more people from the neighborhood likely to be out on the street.

The sections of the model that largely explain the choices of residential location and car ownership status are extremely similar in both their estimates and their interpretations to those offered in the discussion in Chapter 3 of this dissertation. The differences that do exist indicate that the estimated model in this chapter is even more robust than that presented in Chapter 3. A number of coefficients are statistically significant here that were not significantly different from zero in the Chapter 3 model. This is likely due to the fact that the sample size here is substantially larger than that which the model in Chapter 3 was based on.

5.4.1 Interpreting the elasticities

Tables 5.4 through 5.6 provide estimates based on the joint choice model of the unconditional and conditional elasticities of each walking level with respect to home population density. Table 5.8 provides elasticity estimates calculated using the choice model of walking level only. Although population density itself is not a direct determinant of the propensity to walk, it is correlated with a number of built environment factors that are direct determinants of the propensity to walk, and for which adequate data is not readily available. These include, but are not limited to, the average distance between destinations in a neighborhood, local traffic congestion levels, and parking costs at local destinations.

The values in these tables of elasticities provide insights into the answers to all of the questions posed in this chapter. The elasticities in Table 5.4 represent in some sense the long-run elasticities of walking level in New York City. This is because they are calculated based on a model that allows for flexibility in not only short-run walking level, but also in the more long-run adjustments of residential location and car ownership level that affect walking level. These long-run elasticities of walking level with respect to population density are substantial for the cases of both "Zero Walkers" and "High Walkers". The model indicates that when population density increases by one percent, the long-run effect on walking level will be to decrease the probability of residents to be in the "Zero Walker" category by 0.15 percent and to increase the probability that residents will be in the "High Walker" category by 0.29 percent (see Table 5.4).

In contrast, the elasticities in Table 5.5 represent the short-run, or more immediate, elasticities of walking level in New York City. These elasticities are calculated based on the full joint model estimates, but are conditioned specifically on both residential location and car ownership level. Comparing the estimated values in these two tables makes a clear case for a larger long-run than short-run effect for both the "Zero Walker" and the "High Walker" categories, just as theory would predict. The point estimates for the "Low Walker" category follow this pattern, but are imprecisely estimated such that there is no statistically significant difference between the values in any of the elasticity tables in this category.

The two analyses discussed thus far in this section answer the first two of the four questions posed at the start of this chapter. Answering the third question, regarding self-selection, involves an extra step. The term "self-selection" refers to people self-selecting into neighborhoods. As described above, the elasticities presented in Table 5.5 are conditional on both residential location and car ownership status. A comparison to the full model elasticities that gets directly at the question of self-selection is found in Table 5.6. This table's elasticities are conditioned on residential location only, and therefore represent the effect of population density on walking level *minus* the effect of locational self-selection.

This means that the portion of the full long-run elasticity that can be accounted for by self-selection can be directly quantified by taking the difference between the values in these two tables. Table 5.7 does just this.[2] For the "Zero Walker" category, the contribution of residential self-selection to the total elasticities is significantly different from zero everywhere but in Manhattan. In the "High Walker" category, the contribution of residential self-selection to the total elasticities is significantly different from zero in all areas of the city.

The point estimates displayed in Table 5.7 indicate that the contribution of locational self-selection to the long-run elasticity of walking level with respect to population density is large. More specifically, locational self-selection accounts for between one-third and one-half of the total effect of population density on walking level in most areas of the city. In Staten Island, the effect of self-selection appears to be the main contributing factor to the relationship between a neighborhood's population density and the walking level of its residents. Turning this statement around, it means that between one-half and two-thirds of the total effect of population density on walking level in most areas of the city can be explained by factors other than residential self-selection.

[2]The confidence intervals shown in this table are calculated according to the theory of error propagation, where the error of the difference of two values $x \pm \delta_x$ and $y \pm \delta_y$ is given by $\sqrt{\delta_x^2 + \delta_y^2}$.

This quantification of the portion of the relationship between neighborhood walkability (here represented through the coarse measure of population density) and the amount of walking done by residents of that neighborhood that is due to locational self-selection is the main contribution of this chapter to the literature. To my knowledge, the effect of locational self-selection has not previously been successfully quantified.

The final question raised in the introductory section of this chapter asks about the difference between the results obtained from the complex full joint choice model estimated here and a simpler model of walking level that takes both residential location and car ownership level as exogenous variables. This model was also estimated for this paper, and the resulting elasticities are reported in Table 5.8. For this model and dataset, the point estimates of the elasticities that are derived from the simpler model of walking level are extremely similar to the point estimates of the elasticities from the full joint model that are conditional on residential location. It is likely that this result is not generalizable, but rather an artifact of the particular dataset and model being estimated here. However, the results do indicate that the estimates of elasticities that are calculated based on a simpler model of only walking level are not substantially different from those estimated based on a more complex joint choice estimation methodology.

5.5 Conclusion

This chapter has presented the results of models of the choice of walking level in New York City, both in the context of the related choices of residential location and car ownership and on its own. The elasticities calculated from the results of these models shed light on two questions that have eluded researchers in this area. First, they identify the portion of the total elasticity effect that is likely to be realized in the short-term. Second, they separate the portion of the total elasticity that is due to residential self-selection from that which is not.

The explanatory variable used in this chapter to demonstrate this methodology is population density. The results have shown that the short-term effect on walking level of an increase in population density will be between one- and two-thirds of the long-term effect. Further, the results have shown that the portion of the total elasticity that is not due to residential self-selection is between one-half and two-thirds of the total elasticity for all areas of New York City except Staten Island.

In the long term, population density in a neighborhood can be affected by zoning-type policies, but it is not the first thing one thinks of when listing policies that aim to make a neighborhood more pedestrian-friendly. Options such as sidewalk improvements, traffic calming measures, crime prevention, or even creating pedestrian-only streets are more obvious.

The contribution to the literature of this chapter, then, is not in the specific results regarding the responsiveness of New Yorkers' walking behavior to population density, but rather in the demonstration of the methodology. The methodology demonstrated here is new to the self-selection literature. To my knowledge, it is the only methodology available that quantifies the extent of self-selection, rather than simply identifying its existence. Applying this methodology to more obvious policy options that improve the pedestrian-friendliness of a neighborhood - such as those listed above - is an area for future research.

However, the downside to this methodology is that it requires a joint model of the choice of interest (in this case walking level) and the choice of location, and modeling the choice of location is not a trivial exercise. Interestingly, the model in this chapter of walking level that takes both car ownership status and residential location as exogenous produces elasticity results that are similar to the full choice model. It is an open research question whether this finding is peculiarity of this particular data set or whether it is a result that can be generalized.

Table 5.2: Multinomial logit model of the Full Joint Choice of Residential Location, Car Ownership Status, and Commute Mode

	Coefficient	S.E.	Coefficient	S.E.
WALKING LEVEL CHOICE VARIABLES	**Walks between 1% and 50% of Trips**		**Walks 50% or more of Trips**	
Number of total trips	0.359***	0.024	0.275***	0.021
Rain on travel day	-0.154	0.216	-0.426**	0.170
Percent work trips	-0.025***	0.002	-0.014***	0.002
Percent work-related trips	-0.009*	0.005	-0.014***	0.004
Percent school trips	-0.017***	0.004	-0.006***	0.002
Percent social trips	-0.006**	0.002	-0.004**	0.002
Percent personal trips	-0.014***	0.003	-0.011***	0.002
Percent shopping trips	-0.007**	0.003	0.008***	0.002
Percent serve passenger trips	-0.017***	0.004	-0.004*	0.002
Average number of travelers	-0.108	0.079	-0.119**	0.059
Percent night trips	-0.011***	0.003	-0.004**	0.002
Home retail density	-0.083	0.559	0.961**	0.418
Miles from home to midtown	-0.092***	0.020	-0.090***	0.015
Home population density	0.040***	0.014	0.034***	0.010
CAR OWNERSHIP STATUS CHOICE VARIABLES	**Income $<$ \$25,000 per HH member**		**Income $>$ \$25,000 per HH member**	
Car Insurance Cost Per Income	-0.329***	0.068	-0.421***	0.106
Income if One Car	0.432***	0.036	0.100***	0.023
Income if Two or More Cars	0.605***	0.053	0.147***	0.024
Subway Lines at Home if One Car	-0.129***	0.037	0.054	0.047
Subway Lines at Home if Two or More Cars	-0.063	0.055	0.115**	0.054
Miles to Midtown Manhattan if One Car	0.047***	0.014	0.245***	0.026
Miles to Midtown Manhattan if Two or More Cars			0.292***	0.025
Retail Density at Home if One Car	-0.169	0.651	1.638***	0.508
Retail Density at Home if Two or More Cars	0.025	1.260	-0.100	0.934
Population Density at Home if One Car	-0.060***	0.012	-0.061***	0.016
Population Density at Home if Two or More Cars	-0.236***	0.023	-0.129***	0.020

	Coefficient	S.E.	Coefficient	S.E.
CAR OWNERSHIP STATUS CHOICE VARIABLES	**Income < $25,000 per HH member**		**Income > $25,000 per HH member**	
Not Segregated By Income, cont.				
Household Size if One Car in HH	0.076***	0.027		
Household Size if Two or More Cars in HH	0.432***	0.034		
RESIDENTIAL LOCATION CHOICE VARIABLES	**Income < $25,000 per HH member**		**Income > $25,000 per HH member**	
Rent Per Income Per Household Size	-0.041	0.071	-1.460***	0.311
NH Percent White if Non-White HH	-2.772***	0.170	-2.406***	0.221
NH Percent Non-White if White HH	-2.937***	0.130	-2.792***	0.159
NH Percent Owner-Occupied	-1.484***	0.182	-2.269***	0.236
NH Population Density	0.109***	0.010	0.140***	0.015
NH Miles From Midtown Manhattan	0.047***	0.014	-0.194***	0.028
NH Retail Density	-1.194**	0.478	-1.326**	0.590
NH Subway Line Availability	0.136***	0.025	0.039	0.039
NH Median Income	-0.078***	0.023	0.180***	0.024
Not Segregated By Income				
NH Miles From Midtown Manhattan if Children in HH	0.051***	0.011		
NH Miles From Midtown Manhattan * Number of Workers in HH	-0.009*	0.005		
NH Miles From Midtown Manhattan if HH Head ¡35	-0.020*	0.010		
NH Subway Line Availability if Children in HH	-0.134***	0.031		
NH Median Income if Children in HH	-0.100***	0.022		
NH Population Density if HH Head ¡ 35	0.024**	0.010		
NH Percent Owner-Occupied if Homeowner	3.491***	0.158		
NH Employment Density * Number of Workers in HH	-0.008	0.006		
PLUS ALTERNATIVE-SPECIFIC CONSTANTS[a]				

ESTIMATION SUMMARY INFORMATION	
Observations	4382
Alternatives[b]	99
Pseudo R-squared	0.2345

* significant at 10%; ** significant at 5%; *** significant at 1%

a. There are 26 alternative specific constants in this model, representing all combinations of walking level and car ownership, and three residential location groups (Manhattan, Staten Island, and the Rest of New York City).

b. The 99 compound alternatives consist of 3 walking level alternatives, 3 car ownership status alternatives, and 11 census tract alternatives sampled from the full set of over 2000 possible census tracts.

Table 5.3: Multinomial logit model of the Choice of Walking Level (Only)

	Coefficient	S.E.	Coefficient	S.E.
WALKING LEVEL CHOICE VARIABLES	**Walks between 1% and 50% of Trips**		**Walks 50% or more of Trips**	
Number of total trips	0.390***	0.025	0.324***	0.023
Rain on travel day	-0.305	0.222	-0.627***	0.183
Percent work trips	-0.025***	0.003	-0.014***	0.002
Percent work-related trips	-0.011**	0.005	-0.018***	0.005
Percent school trips	-0.017***	0.005	-0.005**	0.003
Percent social trips	-0.006**	0.003	-0.004**	0.002
Percent personal trips	-0.014***	0.003	-0.011***	0.002
Percent shopping trips	-0.004	0.003	0.011***	0.002
Percent serve passenger trips	-0.015***	0.004	-0.001	0.003
Average number of travelers	-0.078	0.080	-0.087	0.061
Percent night trips	-0.011***	0.003	-0.005**	0.002
Home retail density	0.601	0.609	1.651***	0.469
Miles from home to midtown	-0.068***	0.023	-0.068***	0.018
Home population density	0.061***	0.017	0.058***	0.013
Manhattan households	0.239	0.209	0.119	0.158
Staten Island households	0.388	0.322	-0.031	0.278
One Car households	-0.533***	0.130	-0.962***	0.097
Two or More Car households	-0.705***	0.165	-1.550***	0.135
PLUS ALTERNATIVE-SPECIFIC CONSTANTS[a]				
ESTIMATION SUMMARY INFORMATION				
Observations	4382			
Alternatives[b]	3			
Pseudo R-squared	0.3650			

* significant at 10%; ** significant at 5%; *** significant at 1%

a. There are 2 alternative specific constants in this model, representing walking level alternatives.

b. The 3 alternatives are the 3 walking levels.

Table 5.4: Elasticities of Walking Level with respect to Population Density in Full Joint Model

	Zero Walker	**Low Walker**	**High Walker**
FIVE BOROUGHS OF NEW YORK CITY			
Elasticity	-0.15	0.18	0.29
95% Confidence Interval	[-0.20, -0.10]	[-0.02, 0.38]	[0.19, 0.39]
90% Confidence Interval	[-0.19, -0.11]	[-0.01, 0.35]	[0.20, 0.38]
MANHATTAN ONLY			
Elasticity	-0.25	0.21	0.33
95% Confidence Interval	[-0.33, -0.17]	[-0.05, 0.47]	[0.21, 0.45]
90% Confidence Interval	[-0.32, -0.18]	[0.00, 0.42]	[0.23, 0.43]
STATEN ISLAND ONLY			
Elasticity	-0.10	0.15	0.31
95% Confidence Interval	[-0.13, -0.07]	[-0.01, 0.31]	[0.21, 0.40]
90% Confidence Interval	[-0.13, -0.07]	[0.02, 0.28]	[0.23, 0.39]
REST OF NEW YORK CITY			
Elasticity	-0.14	0.17	0.28
95% Confidence Interval	[-0.18, -0.09]	[-0.01, 0.35]	[0.18, 0.38]
90% Confidence Interval	[-0.18, -0.10]	[0.01, 0.32]	[0.20, 0.36]

Table 5.5: Elasticities of Walking Level with respect to Population Density Conditional on Both Residential Location and Car Ownership Status in Full Joint Model

	Zero Walker	**Low Walker**	**High Walker**
FIVE BOROUGHS OF NEW YORK CITY			
Elasticity	-0.07	0.12	0.10
95% Confidence Interval	[-0.11, -0.03]	[-0.05, 0.29]	[0.01, 0.19]
90% Confidence Interval	[-0.10, -0.04]	[-0.03, 0.27]	[0.02, 0.17]
MANHATTAN ONLY			
Elasticity	-0.20	0.16	0.10
95% Confidence Interval	[-0.32, -0.08]	[-0.09, 0.41]	[0.00, 0.20]
90% Confidence Interval	[-0.30, -0.10]	[-0.05, 0.37]	[0.01, 0.19]
STATEN ISLAND ONLY			
Elasticity	0.0	0.03	0.03
95% Confidence Interval	[0.00, 0.00]	[0.00, 0.06]	[0.01, 0.05]
90% Confidence Interval	[0.00, 0.00]	[0.01, 0.05]	[0.01, 0.05]
REST OF NEW YORK CITY			
Elasticity	-0.05	0.12	0.10
95% Confidence Interval	[-0.08, -0.02]	[0.00, 0.24]	[0.03, 0.17]
90% Confidence Interval	[-0.07, -0.02]	[0.02, 0.22]	[0.04, 0.16]

Table 5.6: Elasticities of Walking Level with respect to Population Density Conditional on Residential Location in Full Joint Model

	Zero Walker	**Low Walker**	**High Walker**
FIVE BOROUGHS OF NEW YORK CITY			
Elasticity	-0.10	0.11	0.15
95% Confidence Interval	[-0.14, -0.06]	[-0.07, 0.29]	[0.06, 0.24]
90% Confidence Interval	[-0.13, -0.07]	[-0.04, 0.26]	[0.08, 0.22]
MANHATTAN ONLY			
Elasticity	-0.25	0.13	0.15
95% Confidence Interval	[-0.37, -0.13]	[-0.12, 0.38]	[0.05, 0.25]
90% Confidence Interval	[-0.35, -0.15]	[-0.08, 0.34]	[0.06, 0.24]
STATEN ISLAND ONLY			
Elasticity	0.0	0.03	0.06
95% Confidence Interval	[0.00, 0.00]	[0.00, 0.06]	[0.04, 0.08]
90% Confidence Interval	[0.00, 0.00]	[0.01, 0.05]	[0.04, 0.08]
REST OF NEW YORK CITY			
Elasticity	-0.08	0.10	0.17
95% Confidence Interval	[-0.11, -0.05]	[-0.02, 0.22]	[0.10, 0.24]
90% Confidence Interval	[-0.10, -0.05]	[0.00, 0.20]	[0.11, 0.23]

Table 5.7: Self-Selection Contribution to Elasticities of Walking Level with respect to Population Density in Full Joint Model

	Zero Walker	**Low Walker**	**High Walker**
FIVE BOROUGHS OF NEW YORK CITY			
Self-Selection Elasticity Effect	-0.05	0.07	0.14
95% Confidence Interval	[-0.11, 0.01]	[-0.20, 0.34]	[0.01, 0.27]
90% Confidence Interval	[-0.10, 0.00]	[-0.15, 0.29]	[0.03, 0.25]
MANHATTAN ONLY			
Self-Selection Elasticity Effect	0.00	0.08	0.18
95% Confidence Interval	[-0.14, 0.14]	[-0.28, 0.44]	[0.02, 0.34]
90% Confidence Interval	[-0.12, 0.12]	[-0.22, 0.38]	[0.05, 0.31]
STATEN ISLAND ONLY			
Self-Selection Elasticity Effect	-0.10	0.12	0.25
95% Confidence Interval	[-0.13, -0.07]	[-0.04, 0.28]	[0.15, 0.35]
90% Confidence Interval	[-0.13, -0.07]	[-0.01, 0.25]	[0.17, 0.33]
REST OF NEW YORK CITY			
Self-Selection Elasticity Effect	-0.06	0.07	0.11
95% Confidence Interval	[-0.11, -0.01]	[-0.15, 0.29]	[-0.01, 0.23]
90% Confidence Interval	[-0.11, -0.01]	[-0.12, 0.26]	[0.01, 0.21]

Table 5.8: Elasticities of Walking Level with respect to Population Density in Walking Level (Only) Choice Model

	Zero Walker	**Low Walker**	**High Walker**
FIVE BOROUGHS OF NEW YORK CITY			
Elasticity	-0.10	0.16	0.17
95% Confidence Interval	[-0.15, -0.05]	[0.00, 0.32]	[0.08, 0.26]
90% Confidence Interval	[-0.14, -0.06]	[0.02, 0.30]	[0.09, 0.25]
MANHATTAN ONLY			
Elasticity	-0.30	0.20	0.18
95% Confidence Interval	[-0.45, -0.15]	[-0.06, 0.46]	[0.07, 0.29]
90% Confidence Interval	[-0.43, -0.17]	[-0.02, 0.42]	[0.09, 0.27]
STATEN ISLAND ONLY			
Elasticity	0.0	0.05	0.05
95% Confidence Interval	[0.00, 0.00]	[0.02, 0.08]	[0.02, 0.08]
90% Confidence Interval	[0.00, 0.00]	[0.02, 0.08]	[0.03, 0.07]
REST OF NEW YORK CITY			
Elasticity	-0.08	0.16	0.16
95% Confidence Interval	[-0.12, -0.04]	[0.02, 0.30]	[0.08, 0.24]
90% Confidence Interval	[-0.11, -0.05]	[0.04, 0.28]	[0.09, 0.23]

Chapter 6

Conclusion: Successes, Shortcomings, and Next Steps

This dissertation uses discrete choice econometrics along with GIS technology to explore how New Yorkers make choices about where they live, whether they own cars, and how they get around in their daily lives. The first of the analyses presented here focuses on the inter-relationship between the choices of residential location, car ownership status, and commute mode. The second analysis uses the model developed in the first analysis along with GIS to simulate spatially-explicit policy scenarios for the city. The final analysis included in this dissertation explores the relationship between the choice of walking as a transportation mode and the neighborhood quality of population density, specifically looking at the extent to which locational self-selection explains this relationship.

6.1 Successes and shortcomings: Chapters 3 & 4

The discrete choice model developed in Chapter 3 and used in the Chapter 4 analysis as well can be viewed as a technical success. The joint model of the decisions of New Yorkers about where to live, how many cars to own, and which mode to use for their commutes appears to fit the data well and yields results that are remarkably consistent with both economic theory and with those in the existing literature. These results indicate that the factors to which New Yorkers are most sensitive in decision to use cars for commuting are commute time and commute cost. In deciding whether to own a car, New Yorkers are heavily influenced by the population density near their homes. Population density is not likely to be the factor that has a direct causal effect on New Yorkers' choice of car ownership status, but it suggests the importance of the omitted variable of home parking price.

In itself, creating a complex empirical model that produces results that make sense is an accomplishment. The fact that there are virtually no surprises does, however, beg the question, "So what?" (which I have been asked a number of times in the course of presenting this work). I argue that the fact that the model estimated here produced results consistent with those in the existing literature is actually surprising. The models in the literature were estimated using data from cities that span a wide variety of transportation-land use system characteristics, and therefore often are not entirely consistent with each other. The model estimated in Chapter 3 of this dissertation uses data from New York City, which itself contains a substantial variety of transportation-land use system characteristics. Elasticity results were estimated from this model for the city as a whole as well as for sub-areas within the city that are more homogeneous in their transportation-land use characteristics. It is when these results for New York sub-areas are compared to those found in the literature for cities with comparable transportation-land use characteristics that the similarities are clearest.

In effect, then, the single model estimated in Chapter 3 provides results that are representative of a behavior in a wide variety of transportation-land use system settings. This result is predicted by economic theory, since there is no theoretical reason that people in different transportation-land use system settings would go about making their utilitarian decisions about car ownership and commute mode in different ways. However, it may be quite surprising to many city planners and even perhaps to many economists. It is something I continue to question as well, and in the "Directions for future research" section of this chapter, I propose a study to further test this result.

In the first section of the introduction to this dissertation, the overarching goal of the project is identified as gaining a better understanding of how New Yorkers make their environmentally-friendly decisions with the idea that this understanding could help city planners in other cities to create a "similar choice environment". This goal has been achieved here, but now that the analysis is done, it is not entirely clear how to use the results of the present model to gain insight into what might be done to improve the sustainability of the transportation-land use systems in other cities. It turns out that translating results from one choice environment to another is not a straightforward application of discrete choice models. Some ideas are provided for how this might be done in the "Directions for future research" section of this chapter.

Chapter 4 achieved its main objective. It clearly demonstrates the value of representing results in a spatially disaggregated way in order to illustrate the spatial patterns in the behavioral response to changes in the transportation-land use system.

Conducting these scenario analyses made it clear that higher spatial resolution on a number of variables could improve the results. These variables are those that were taken from the 1997 Economic Census and those generated by neighborhood using GIS. The first category of variables includes retail and employment density, and is available only by zip code from the census. However, it may be possible to obtain similar variables at a higher spatial resolution from the City of New York. The second category of variables could be created in the future at a higher spatial resolution. This exercise is extremely time-consuming, however, and is left for future iterations of the model.

In addition, having more and better-characterized variables in the original model itself would also allow for additional and more realistic scenario analyses in the future. Particularly helpful additional variables include parking price or availability, violent crime frequency, and public school quality information, all by census tract. Transit routes could be better-characterized to include more realistic neighborhood-specific waiting times and information about the number of transfers that are necessary for each route.

6.2 Successes and shortcomings: Chapter 5

The main contribution to the literature of Chapter 5 is methodological. The chapter lays out a clear, replicable methodology for quantifying the extent to which an observed choice pattern can be explained by self-selection. The application here looks specifically at locational self-selection, but the methodology is general. This work has the potential to be a substantial contribution to the joint choice modeling literature. It is a new methodological extension of joint choice models that has potential application beyond the narrow topic area of residential self-selection.

Unfortunately, the present application of the model is severely limited by the data available, and the results presented in this dissertation regarding the locational self-selection effect on the relationship between population density and walking level are not particularly useful in and of themselves. Having more policy-relevant variables in a model of walking level could make it possible to explore the effect of self-selection on the relationship between the choice to walk and neighborhood characteristics such as sidewalk availability, traffic calming implementation, or crime levels.

However, after doing this research, I question the importance of knowing the contribution of locational self-selection to the observed correlation between nonmotorized transport use and certain features of the built environment. Sidewalks and bicycle lanes are not investments we as a society should make primarily for public health reasons – we should make them because we want to foster more livable communities in our cities where neighbors interact on the street, because we are care about the personal mobility of children and others who cannot drive cars, and because we want to offer a viable alternative for local transportation to the energy-using, pollutant-emitting private car. If people are walking and biking more in neighborhoods that create a safe environment for doing so by providing appropriate infrastructure, does it matter whether these people are those who are "natural" walkers and bicyclists?

6.3 Directions for future research

There are two main directions that I would like to take this research in the near future. The first aims to further test the theory that people everywhere make their transportation choices in similar ways. The methodology proposed is to extend this work to the suburbs of New York City. It is clear that suburbanites make different transport choices than their urban counterparts. The cause of these observed choice differences is less clear. Are the basic travel preferences of suburbanites simply different from those of city-dwellers? Alternatively, are the underlying preferences of these groups similar, and the combination of differences in the built environment and available transportation options responsible for the difference in choices? The extent to which the first of these hypotheses is true has important implications for the effectiveness of transport and land use policy in suburbia versus the city. Knowing the major differences, if any, between the factors that figure prominently in the transportation choices of the two populations will help planners to better serve each of them.

The data set used for this dissertation includes information for not only the five boroughs of the city, but also for the entire metropolitan area of New York. A future application of the model developed in this research is to extend it to model the suburban areas of the New York metropolitan area. The estimated parameters and elasticity results for separate models of urban counties and suburban counties could easily be compared. These statistical tests will give evidence as to whether or not urbanites and surbanites have different preferences that lead to their different choices, helping planners to better target their populations of interest.

A second direction for future research aims to adapt the estimated transportation-land use utility function for New York to inform planners in other cities about what to expect as behavioral responses to investments and implementation of policies that affect their transportation-land use systems. The literature in this area of discrete choice model transferability exists, but is small. In my work over the next two years, I will have an opportunity to contribute to the expansion of this literature by exploring the possibility of transfering modeling techniques and/or estimated results to other cities.

Bibliography

Anas, A. (1982). *Residential location markets and urban transportation: Economic theory, econometrics, and policy analysis with discrete choice models*. Academic Press, New York.

Asensio, J. (2002). Transport mode choice by commuters to Barcelona's CBD. *Urban Studies*, 39:1881–1895.

Bagley, M. N. and Mokhtarian, P. L. (2002). The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *The Annals of Regional Science*, 36:279–297.

Ben-Akiva, M. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA.

Bhat, C. R. and Pulugurta, V. (1998). A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. *Transportation Research B*, 32:61–75.

Cervero, R. and Duncan, M. (2003). Walking, bicycling, and urban landscapes: Evidence from the San Francisco bay area. *American Journal of Public Health*, 93(9):1478–1483.

Cervero, R. and Gorham, R. (1995). Commuting in transit versus automobile neighborhoods. *Journal of the American Planning Association*, 61(2):210–225.

Cervero, R. and Radisch, C. (1996). Travel choice in pedestrian versus automobile oriented neighborhoods. *Transport Policy*, 3(3):127–141.

Dargay, J. and Gately, D. (1999). Income's effect on car and vehicle ownership, worldwide: 1960-2015. *Transportation Research Part A*, 33:101–138.

de Jong, G. (1990). An indirect utility model of car ownership and private car use. *European Economic Review*, 34:971–985.

Everitt, B. S. (1993). *Cluster Analysis, 3rd Ed.* Edward Arnold, London.

Gangrade, S., Pendyala, R. M., and McCullough, R. G. (2002). A nested logit model of commuters' activity schedules. *Journal of Transportation and Statistics*, 5(2/3).

Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: The case of the U.S. automobile industry. *Econometrica*, 63:891–951.

Handy, S., Cao, X., and Mokhtarian, P. L. (2006). Self-selection in the relationship between the built environment and walking. *Journal of the American Planning Association*, 72(1).

Hensher, D. A. and Greene, W. H. (2001). The mixed logit model: The state of practice and warnings for the unwary. Working Paper, School of Business, University of Sydney.

Hensher, D. A. and Ton, T. T. (2000). A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. *Transportation Research Part E*, 36:155–172.

Ingram, G. K. and Liu, Z. (1997). Motorization and the provision of roads in countries and cities. Technical report, World Bank. Policy Research Working Paper No. 1842.

Ingram, G. K. and Liu, Z. (1999). 10: Determinants of motorization and road provision. In *Essays in Transportation Economics and Policy*, chapter 10. Brookings Institution Press, Washington, DC.

Kitamura, R., Mokhtarian, P. L., and Laidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco bay area. *Transportation*, 24:125–158.

Krizek, K. J. (2003). Residential relocation and changes in urban travel: Does neighborhood-scale urban form matter? *Journal of the American Planning Association*, 69(3):265–281.

Lerman, S. R. (1977). Location, housing, automobile ownership, and mode to work: a joint choice model. *Transportation Research Record*, 610:6–10.

Mannering, F. and Winston, C. (1985). A dynamic empirical analysis of household vehicle ownership and utilization. *Rand Journal of Economics*, 16:215–236.

Manski, C. F. and Sherman, L. (1980). An empirical analysis of household choice among motor vehicles. *Transportation Research A*, 14:349–366.

McFadden, D. (1978). Modelling the choice of residential location. In *Spatial Interaction Theory and Planning Models*, pages 75–96. North-Holland, Amsterdam.

Onozaka, Y. (2002). Evaluating alternative methods of dealing with missing observations: an economic application. Selected paper for the American Agricultural Economics Association Annual Meeting.

Owen, D. (2004). Green Manhattan. *The New Yorker*.

Quigley, J. M. (1985). Consumer choice of dwelling, neighborhood, and public services. *Regional Science and Urban Economics*, 15(1):41–63.

Redmond, L. S. and Mokhtarian, P. L. (2001). The positive utility of the commute: Modeling ideal commute time and relative desired commute amount. *Transportation*, 28(2):179–205.

Rodriguez, D. A. and Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D*, 9:151–173.

Schimek, P. (1996). Household motor vehicle ownership and use: How much does residential density matter? *Transportation Research Record*, 1552:120–125.

Train, K. (1980). A structured logit model of auto ownership and mode choice. *The Review of Economic Studies*, 47(2):357–370.

Train, K. (2002). *Discrete Choice Methods with Simulation.* Cambridge University Press.

Train, K. (2005). Personal communication via email with Kenneth Train, May 2, 2005.

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

U.S. Census Bureau (2000). *2000 Decennial United States Census of Population and Housing.* Washington, D.C.

U.S. Department of Transportation (1997). *1995 Nationwide Personal Transportation Survey.* Federal Highway Administration, Research Triangle Park, NC. Published by the Research Triangle Institute. FHWA-PL-98-002.

Waddell, P. (1993). Exogenous workplace choice in residential location models – Is the assumption valid? *Geographical Analysis*, 25(1):65–82.

Zhang, M. (2004). The role of land use in travel mode choice: Evidence from Boston and Hong Kong. *Journal of the American Planning Association*, 70:344–360.

Appendix A

Estimated Coefficients and Calculated Elasticities for All Not-Chosen Models From Chapter 3

Table A.1: Multinomial and Nested Logit Models of the Choice of Residential Location and Car Ownership Status

	MNL Joint Model		Location-Car Nested Model (11 Nests)		Car-Location Nested Model (3 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
CAR OWNERSHIP STATUS CHOICE VARIABLES						
Low Income - Income < $25,000 per HH member						
Car Insurance Cost	-0.126	0.082	-0.027	0.041	-0.090	0.103
Income if One Car	0.718***	0.106	-0.360***	0.118	0.752***	0.111
Income if Two or More Cars	1.526***	0.135	-0.792***	0.240	1.622***	0.147
Subway Lines at Home if One Car	-0.108**	0.047	0.020	0.020	-0.201**	0.083
Subway Lines at Home if Two or More Cars	-0.003	0.057	-0.095**	0.042	0.052	0.098
Miles to Midtown Manhattan if One Car	0.022	0.018	-0.018**	0.008	0.067**	0.031
Miles to Midtown Manhattan if Two or More Cars						
Retail Density at Home if One Car	-0.272	1.020	0.136	0.436	-1.700	1.901
Retail Density at Home if Two or More Cars	-3.315	2.377	1.069	1.018	-7.387*	4.647
Population Density at Home if One Car						
Population Density at Home if Two or More Cars	-0.228***	0.023	0.153***	0.045	-0.385***	0.046
High Income - Income > $25,000 per HH member						
Car Insurance Cost	-0.117	0.131	-0.297**	0.119	-0.338	0.206
Income if One Car	0.136***	0.041	-0.080***	0.031	0.126***	0.042
Income if Two or More Cars	0.219***	0.050	-0.119***	0.043	0.191***	0.054
Subway Lines at Home if One Car	0.027	0.054	0.001	0.023	-0.030	0.082
Subway Lines at Home if Two or More Cars	0.091	0.072	0.000	0.031	-0.043	0.116
Miles to Midtown Manhattan if One Car	0.212***	0.030	-0.052***	0.019	0.320***	0.049
Miles to Midtown Manhattan if Two or More Cars	0.257***	0.029	-0.058***	0.020	0.334***	0.045
Retail Density at Home if One Car	1.206**	0.577	-0.323	0.255	1.792**	0.815

	MNL Joint Model		Location-Car Nested Model (11 Nests)		Car-Location Nested Model (3 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
CAR OWNERSHIP STATUS CHOICE VARIABLES, cont.						
High Income - Income > $25,000 per HH member, cont.						
Retail Density at Home if Two or More Cars	-1.654	1.318	0.697*	0.540	-5.018	2.494
Population Density at Home if One Car	-0.045**	0.019	0.009	0.009	-0.030	0.029
Population Density at Home if Two or More Cars	-0.066**	0.027	0.012	0.013	-0.092**	0.042
Not Segregated By Income						
Household Size if One Car	0.245***	0.045	-0.122***	0.041	0.243***	0.047
Household Size if Two or More Cars	0.777***	0.056	-0.361***	0.106	0.784***	0.060
Auto Commute Mode if One Car	2.805***	0.251	-1.093***	0.336	2.624***	0.259
Auto Commute Mode if Two or More Cars	3.495***	0.308	-1.282***	0.388	3.141***	0.328
Transit Commute Mode if One Car	-0.004	0.161	0.031	0.066	-0.150	0.170
Transit Commute Mode if Two or More Cars	-0.218	0.247	0.180	0.117	-0.484*	0.267
RESIDENTIAL LOCATION CHOICE VARIABLES						
Low Income - Income < $25,000 per HH member						
Rent Per Income Per Household Size	-0.073	0.094	-0.105	0.097	-0.208	0.189
NH Percent Owner-Occupied	-1.784***	0.231	-1.785***	0.232	-3.708***	0.621
NH Population Density	0.106***	0.011	0.068***	0.012	0.207***	0.029
NH Miles From Midtown Manhattan	-0.005	0.020	-0.018	0.019	-0.024	0.039
NH Retail Density	-1.549**	0.709	-1.963***	0.661	-2.424*	1.378
NH Subway Line Availability	0.125***	0.036	0.081**	0.032	0.216***	0.073
NH Median Income	-0.033	0.029	-0.020	0.029	-0.068	0.058

	MNL Joint Model		Location-Car Nested Model (11 Nests)		Car-Location Nested Model (3 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
RESIDENTIAL LOCATION CHOICE VARIABLES, cont.						
High Income - Income > $25,000 per HH member						
Rent Per Income Per Household Size	-2.270***	0.746	-2.307***	0.742	-4.182***	1.540
NH Percent Owner-Occupied	-1.749***	0.311	-1.856***	0.308	-3.663***	0.741
NH Population Density	0.160***	0.016	0.118***	0.013	0.285***	0.037
NH Miles From Midtown Manhattan	-0.199***	0.030	0.009	0.024	-0.304***	0.056
NH Retail Density	-0.099	0.629	0.853	0.576	0.757	1.235
NH Subway Line Availability	0.057	0.044	0.090***	0.034	0.209***	0.080
NH Median Income	0.227***	0.033	0.226***	0.033	0.442***	0.080
Not Segregated By Income						
NH Miles From Midtown Manhattan if Children	0.046***	0.016	0.055***	0.016	0.116***	0.035
NH Subway Line Availability if Children	-0.101**	0.042	-0.088**	0.042	-0.195**	0.087
NH Percent White if Non-White HH	-2.455***	0.156	-2.451***	0.155	-4.963***	0.608
NH Percent Non-White if White HH	-2.803***	0.133	-2.818***	0.132	-5.698***	0.650
NH Percent Owner-Occupied if Homeowner	2.972***	0.200	3.332***	0.201	6.266***	0.812
Auto Commute Mode if Staten Island	0.920	0.562	1.552***	0.556	2.200*	1.132
Transit Commute Mode if Staten Island	0.676	0.565	0.448	0.559	1.451	1.109
Auto Commute Mode if Manhattan	-2.230***	0.272	-3.169***	0.280	-4.663***	0.747
Transit Commute Mode if Manhattan	-1.343***	0.196	-1.290***	0.193	-2.531***	0.454
PLUS ALTERNATIVE-SPECIFIC CONSTANTS[a]						

	MNL Joint Model		Location-Car Nested Model (11 Nests)		Car-Location Nested Model (3 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
INCLUSIVE VALUE PARAMETERS[b]						
Inclusive Value Parameter Within Location			-0.397***	0.116		
Inclusive Value Parameter Within Zero Car					2.035***	0.220
Inclusive Value Parameter Within One Car					2.182***	0.252
Inclusive Value Parameter Within Two or More Car					1.925***	0.224
ESTIMATION SUMMARY INFORMATION						
Observations	2728		2728		2728	
Alternatives[c]	33		33		33	
Pseudo R-Squared	0.2116		0.2152		0.2139	

* significant at 10%; ** significant at 5%; *** significant at 1%

a. There are 8 alternative specific constants in this model, representing all combinations of car ownership and three residential location groups (Manhattan, Staten Island, and the Rest of New York City).

b. One is the null hypothesis for the value of inclusive value parameters. The asterisks here, therefore, indicate whether the estimate is significantly different from 1.

c. The 33 compound alternatives consist of 3 car ownership status alternatives and 11 census tract alternatives sampled from the full set of over 2000 possible census tracts.

Table A.2: Multinomial and Nested Logit Models of the Choice of Residential Location and Commute Mode

	MNL Joint Model		Location-Mode Nested Model (11 Nests)		Mode-Location Nested Model (7 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
COMMUTE MODE CHOICE VARIABLES						
Low Income - Income < $25,000 per HH member						
Commute Cost Not Including Parking Costs	-0.376***	0.032	-0.614***	0.058	-0.342***	0.038
Parking Cost At Work	-0.049***	0.014	-0.125***	0.030	-0.056***	0.014
Walking Time	-2.488***	0.141	-3.658***	0.244	-2.264***	0.176
Waiting Time	-5.177***	1.568	-12.388***	3.126	-5.198***	1.560
Riding Time	-1.590***	0.111	-1.150***	0.154	-1.440***	0.154
High Income - Income > $25,000 per HH member						
Commute Cost Not Including Parking Costs	-0.306***	0.037	-0.514***	0.068	-0.276***	0.040
Parking Cost At Work	-0.042***	0.014	-0.095***	0.029	-0.048***	0.014
Walking Time	-2.758***	0.204	-4.114***	0.368	-2.538***	0.228
Waiting Time	-5.388***	1.975	-13.234***	3.911	-5.517***	1.975
Riding Time	-1.874***	0.169	-1.447***	0.209	-1.700***	0.212
Not Segregated By Income						
Subway Lines At Home for Bus	-0.167**	0.065	-0.260**	0.119	-0.146**	0.058
Subway Lines At Home for Subway	-0.044	0.053	0.005	0.093	-0.036	0.049
Subway Lines At Home for Auto	-0.107*	0.059	-0.122	0.101	-0.101*	0.055
Subway Lines At Work for Bus	-0.175***	0.052	-0.228**	0.094	-0.169***	0.051
Subway Lines At Work for Subway	0.271***	0.047	0.554***	0.094	0.272***	0.046
Subway Lines At Work for Auto	0.020	0.051	0.093	0.094	0.022	0.050
One Car HH for Auto	2.890***	0.253	5.692***	0.638	2.913***	0.252

	MNL Joint Model		Location-Mode Nested Model (11 Nests)		Mode-Location Nested Model (7 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
COMMUTE MODE CHOICE VARIABLES, cont.						
Not Segregated By Income, cont.						
Two or More Car HH for Auto	3.438***	0.299	7.150***	0.798	3.496***	0.298
One Car HH for Transit	-0.109	0.170	-0.179	0.300	-0.090	0.168
Two or More Car HH for Transit	-0.403*	0.244	-0.448	0.443	-0.346	0.242
RESIDENTIAL LOCATION CHOICE VARIABLES						
Low Income - Income < $25,000 per HH member						
Rent Per Income Per Household Size						
NH Percent Owner-Occupied	-1.355***	0.240	-1.234***	0.244	-1.147***	0.232
NH Population Density	0.076***	0.011	0.075***	0.012	0.071***	0.012
NH Miles From Midtown Manhattan	0.087***	0.020	0.091***	0.021	0.080***	0.019
NH Retail Density	-3.190***	0.693	-3.513***	0.708	-2.991***	0.643
NH Subway Line Availability	0.141**	0.057	0.121	0.088	0.139***	0.052
NH Median Income	-0.020	0.029	-0.012	0.027	-0.025	0.024
High Income - Income > $25,000 per HH member						
Rent Per Income Per Household Size	-2.003***	0.783	-1.883**	0.795	-1.688**	0.723
NH Percent Owner-Occupied	-1.153***	0.322	-0.966***	0.328	-0.982***	0.300
NH Population Density	0.122***	0.013	0.120***	0.013	0.110***	0.015
NH Miles From Midtown Manhattan	0.082***	0.021	0.088***	0.023	0.074***	0.020
NH Retail Density	-0.357	0.602	-0.590	0.616	-0.318	0.544
NH Subway Line Availability	0.082	0.056	0.054	0.088	0.077	0.052
NH Median Income	0.212***	0.036	0.211***	0.036	0.186***	0.035

	MNL Joint Model		Location-Mode Nested Model (11 Nests)		Mode-Location Nested Model (7 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
RESIDENTIAL LOCATION CHOICE VARIABLES						
Not Segregated By Income						
NH Miles From Midtown Manhattan if Kids in HH	0.045**	0.017	0.050***	0.017	0.036**	0.016
NH Subway Line Availability if Children in HH	-0.119***	0.044	-0.124***	0.044	-0.107***	0.040
NH Percent White if Non-White HH	-2.355***	0.160	-2.434***	0.162	-2.110***	0.213
NH Percent Non-White if White HH	-2.763***	0.137	-2.748***	0.140	-2.496***	0.222
NH Percent Owner-Occupied if Homeowner	2.890***	0.213	2.956***	0.219	2.530***	0.268
One Car HH if Staten Island	0.876**	0.406	0.825*	0.427	0.695*	0.356
Two or More Car HH if Staten Island	1.844***	0.402	1.851***	0.436	1.529***	0.367
One Car HH if Manhattan	-0.893***	0.163	-0.835***	0.182	-0.788***	0.163
Two or More Car HH if Manhattan	-1.738***	0.264	-1.620***	0.304	-1.533***	0.273
PLUS ALTERNATIVE-SPECIFIC CONSTANTS[a]						
INCLUSIVE VALUE PARAMETERS[b]						
Inclusive Value Parameter Within Location			2.040***	0.151		
Inclusive Value Parameter Within Walk					0.726***	0.094
Inclusive Value Parameter Within Taxi					0.867	0.193
Inclusive Value Parameter Within Auto Passenger					0.772**	0.099
Inclusive Value Parameter Within Bus					0.768***	0.073
Inclusive Value Parameter Within Subway with Walk Access					0.948	0.079
Inclusive Value Parameter Within Subway with Non-Walk Access					1.066	0.130
Inclusive Value Parameter Within Auto Driver					0.896	0.077

	MNL Joint Model	Location-Mode Nested Model (11 Nests)	Mode-Location Nested Model (7 Nests)
ESTIMATION SUMMARY INFORMATION			
Observations	2728	2728	2728
Alternatives[c]	77	77	77
Pseudo R-Squared	0.3250	0.3291	0.3258

* significant at 10%; ** significant at 5%; *** significant at 1%

a. There are 20 alternative specific constants in this model, representing all combinations of commute mode and three residential location groups (Manhattan, Staten Island, and the Rest of New York City).

b. One is the null hypothesis for the value of inclusive value parameters. The asterisks here, therefore, indicate whether the estimate is significantly different from 1.

c. The 77 compound alternatives consist of 7 mode alternatives and 11 census tract alternatives sampled from the full set of over 2000 possible census tracts.

Table A.3: Multinomial and Nested Logit Models of the Choice of Car Ownership Status and Commute Mode

	MNL Joint Model		Car-Mode Nested Model (3 Nests)		Mode-Car Nested Model (7 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
COMMUTE MODE CHOICE VARIABLES						
Low Income - Income < $25,000 per HH member						
Commute Cost Not Including Parking Costs	-0.163***	0.030	-0.133***	0.031	-0.168***	0.030
Parking Cost At Work	-0.089***	0.014	-0.082***	0.016	-0.090***	0.014
Walking Time	-2.003***	0.160	-1.581***	0.301	-2.030***	0.162
Waiting Time	-6.363***	1.533	-5.220***	1.482	-6.568***	1.549
Riding Time	0.822***	0.209	0.662***	0.200	0.782***	0.211
High Income - Income > $25,000 per HH member						
Commute Cost Not Including Parking Costs	-0.140***	0.036	-0.116***	0.034	-0.142***	0.037
Parking Cost At Work	-0.078***	0.014	-0.069***	0.015	-0.080***	0.014
Walking Time	-2.455***	0.220	-1.935***	0.380	-2.374***	0.223
Waiting Time	-7.035***	1.935	-5.969***	1.824	-6.818***	1.940
Riding Time	0.455	0.306	0.412	0.254	0.547*	0.310
Not Segregated By Income						
Subway Lines At Home for Bus	-0.308***	0.070	-0.236***	0.069	-0.308***	0.070
Subway Lines At Home for Subway	-0.075	0.059	-0.057	0.047	-0.076	0.059
Subway Lines At Home for Auto	-0.185***	0.066	-0.140**	0.059	-0.172**	0.068
Subway Lines At Work for Bus	-0.136**	0.055	-0.108**	0.046	-0.134**	0.055
Subway Lines At Work for Subway	0.292***	0.050	0.226***	0.057	0.294***	0.050
Subway Lines At Work for Auto	0.013	0.053	-0.001	0.043	0.014	0.054
Staten Island HH for Auto	-0.078	0.564	0.020	0.500	-0.313	0.582

	MNL Joint Model		Car-Mode Nested Model (3 Nests)		Mode-Car Nested Model (7 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
COMMUTE MODE CHOICE VARIABLES, cont.						
Not Segregated By Income, cont.						
Staten Island HH for Transit	-0.791	0.583	-0.580	0.520	-0.867	0.594
Manhattan HH for Auto	-1.137***	0.269	-0.863***	0.268	-1.025***	0.280
Manhattan HH for Transit	-0.677***	0.199	-0.501***	0.182	-0.715***	0.202
CAR OWNERSHIP STATUS CHOICE VARIABLES						
Low Income - Income < $25,000 per HH member						
Car Insurance Cost	-0.168*	0.096	-0.172**	0.096	-0.265**	0.133
Income if One Car in HH	0.878***	0.105	0.870***	0.105	1.112***	0.160
Income if Two or More Cars in HH	1.929***	0.140	1.920***	0.140	2.388***	0.266
Subway Lines at Home if One Car in HH	-0.062	0.050	-0.070*	0.050	-0.065	0.067
Subway Lines at Home if Two or More Cars in HH	0.176**	0.071	0.156**	0.071	0.204**	0.090
Miles to Midtown Manhattan if One Car in HH	0.045**	0.017	0.046***	0.017	0.056***	0.022
Miles to Midtown Manhattan if Two or More Cars in HH						
Retail Density at Home if One Car in HH	-0.278	1.065	-0.281	1.064	-0.291	1.401
Retail Density at Home if Two or More Cars in HH	-2.164	2.301	-2.136	2.305	-2.620	2.933
Population Density at Home if One Car in HH						
Population Density at Home if Two or More Cars in HH	-0.364***	0.032	-0.365***	0.032	-0.432***	0.059
High Income - Income > $25,000 per HH member						
Car Insurance Cost	0.524***	0.199	0.506**	0.199	0.707***	0.270
Income if One Car in HH	0.208***	0.044	0.206***	0.044	0.287***	0.067

	MNL Joint Model		Car-Mode Nested Model (3 Nests)		Mode-Car Nested Model (7 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
CAR OWNERSHIP STATUS CHOICE VARIABLES, cont.						
High Income - Income > $25,000 per HH member, cont.						
Income if Two or More Cars in HH	0.329***	0.054	0.326***	0.054	0.424***	0.083
Subway Lines at Home if One Car in HH	-0.029	0.058	-0.033	0.058	-0.058	0.082
Subway Lines at Home if Two or More Cars in HH	-0.007	0.079	-0.021	0.080	-0.018	0.107
Miles to Midtown Manhattan if One Car in HH	0.142***	0.033	0.145***	0.033	0.177***	0.045
Miles to Midtown Manhattan if Two or More Cars in HH	0.158***	0.034	0.162***	0.034	0.196***	0.047
Retail Density at Home if One Car in HH	0.944	0.593	0.951	0.592	1.060	0.815
Retail Density at Home if Two or More Cars in HH	-1.660	1.276	-1.614	1.277	-2.766	1.813
Population Density at Home if One Car in HH	-0.054**	0.022	-0.055**	0.022	-0.078**	0.032
Population Density at Home if Two or More Cars in HH	-0.081**	0.031	-0.084***	0.031	-0.106**	0.043
Not Segregated By Income						
Household Size if One Car in HH	0.333***	0.045	0.331***	0.045	0.435***	0.068
Household Size if Two or More Cars in HH	0.939***	0.057	0.938***	0.057	1.143***	0.122
Staten Island if One Car in HH	0.314	0.417	0.306	0.414	0.573	0.540
Staten Island if Two or More Cars in HH	0.992**	0.408	0.986**	0.403	1.416**	0.549
Manhattan if One Car in HH	-0.487**	0.200	-0.526***	0.200	-0.677**	0.278
Manhattan if Two or More Cars in HH	-0.480	0.345	-0.531	0.344	-0.684	0.461
PLUS ALTERNATIVE-SPECIFIC CONSTANTS[a]						

	MNL Joint Model		Car-Mode Nested Model (3 Nests)		Mode-Car Nested Model (7 Nests)	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
INCLUSIVE VALUE PARAMETERS[b]						
Inclusive Value Parameter Within Zero-Car HH			0.716**	0.142		
Inclusive Value Parameter Within One-Car HH			0.809	0.138		
Inclusive Value Parameter Within Two-or-More-Car HH			0.931	0.160		
Inclusive Value Parameter Within Walk					1.623**	0.288
Inclusive Value Parameter Within Taxi					2.043	1.702
Inclusive Value Parameter Within Auto Passenger					1.274	0.312
Inclusive Value Parameter Within Bus					1.301*	0.180
Inclusive Value Parameter Within Subway with Walk Access					1.511**	0.209
Inclusive Value Parameter Within Subway with Non-Walk Access					0.851	0.151
Inclusive Value Parameter Within Auto Driver					1.090	0.168
ESTIMATION SUMMARY INFORMATION						
Observations	2728		2728		2728	
Alternatives[c]	20		20		20	
Pseudo R-Squared	0.3239		0.3243		0.3254	

* significant at 10%; ** significant at 5%; *** significant at 1%

a. There are 19 alternative specific constants in this model, representing all combinations of commute mode alternatives and car ownership status levels.

b. One is the null hypothesis for the value of inclusive value parameters. The asterisks here, therefore, indicate whether the estimate is significantly different from 1.

c. The 20 compound alternatives consist of 7 mode alternatives and 3 car ownership status alternatives, where the compound alternatives that included both Auto Driver for the mode and Zero Car Household for the car ownership status were removed from the choice set.

Table A.4: Multinomial logit model of the Choice of Commute Mode

	Coefficient	S.E.	Coefficient	S.E.
COMMUTE MODE CHOICE VARIABLES	**Income $<$ \$25,000 per HH member**		**Income $>$ \$25,000 per HH member**	
Commute Cost Not Including Parking Costs	-0.198***	0.032	-0.157***	0.036
Parking Cost At Work	-0.089***	0.014	-0.083***	0.013
Walking Time	-2.080***	0.166	-2.473***	0.222
Waiting Time	-7.671***	1.651	-7.604***	2.039
Riding Time	0.920***	0.212	0.472	0.313
Not Segregated By Income				
Subway Lines At Home for Bus	-0.327***	0.070		
Subway Lines At Home for Subway	-0.094	0.059		
Subway Lines At Home for Auto	-0.201***	0.066		
Subway Lines At Work for Bus	-0.165***	0.055		
Subway Lines At Work for Subway	0.262***	0.050		
Subway Lines At Work for Auto	-0.002	0.054		
One Car HH for Auto	2.981***	0.258		
Two Car HH for Auto	3.399***	0.305		
One Car HH for Transit	-0.059	0.182		
Two Car HH for Transit	-0.452*	0.257		
Staten Island HH for Auto	-0.204	0.591		
Staten Island HH for Transit	-0.987	0.608		
Manhattan HH for Auto	-0.647**	0.259		
Manhattan HH for Transit	-0.502**	0.213		
PLUS ALTERNATIVE-SPECIFIC CONSTANTS[a]				
ESTIMATION SUMMARY INFORMATION				
Observations	2728			
Alternatives[b]	7			
Pseudo R-squared	0.3897			

* significant at 10%; ** significant at 5%; *** significant at 1%

a. There are 6 alternative specific constants in this model, representing the commute mode alternatives.

b. There are 7 mode alternatives in this model.

Table A.5: Multinomial logit model of the Choice of Car Ownership Status

	Coefficient	**S.E.**	**Coefficient**	**S.E.**
CAR OWNERSHIP STATUS CHOICE VARIABLES	**Income < $25,000 per HH member**		**Income > $25,000 per HH member**	
Car Insurance Cost	0.040	0.102	0.481**	0.206
Income if One Car in HH	0.882***	0.111	0.169***	0.044
Income if Two or More Cars in HH	1.934***	0.152	0.271***	0.056
Subway Lines at Home if One Car in HH	-0.035	0.052	0.005	0.060
Subway Lines at Home if Two or More Cars in HH	0.202***	0.075	0.033	0.081
Miles to Midtown Manhattan if One Car in HH	0.035**	0.018	0.153***	0.034
Miles to Midtown Manhattan if Two or More Cars in HH			0.183***	0.036
Retail Density at Home if One Car in HH	0.090	1.079	0.647	0.617
Retail Density at Home if Two or More Cars in HH	-3.094	2.674	-2.184	1.338
Population Density at Home if One Car in HH			-0.061***	0.018
Population Density at Home if Two or More Cars in HH	-0.413***	0.036	-0.055**	0.023
Not Segregated By Income				
Household Size if One Car in HH	0.317***	0.048		
Household Size if Two or More Cars in HH	0.918***	0.061		
Auto Commuter if One Car in HH	2.245***	0.231		
Auto Commuter if Two or More Cars in HH	2.714***	0.311		
Transit Commuter if One Car in HH	-0.143	0.174		
Transit Commuter if Two or More Cars in HH	-0.595**	0.274		
Staten Island HH if One Car in HH	0.318	0.427		
Staten Island HH if Two or More Cars in HH	0.951**	0.421		
Manhattan HH if One Car in HH	-0.377*	0.212		
Manhattan HH if Two or More Cars in HH	-0.384	0.351		
PLUS ALTERNATIVE-SPECIFIC CONSTANTS[a]				

ESTIMATION SUMMARY INFORMATION	
Observations	2728
Alternatives[b]	3
Pseudo R-squared	0.2779

* significant at 10%; ** significant at 5%; *** significant at 1%

a. There are 2 alternative specific constants in this model, representing car ownership level alternatives.

b. There are 3 car ownership status alternatives in this model.

Table A.6: Elasticities of car ownership and car use for commuting in Joint Car Ownership and Mode Choice Model

	Car Use	**Zero Car**	**One Car**	**Two+ Car**
FIVE BOROUGHS OF NEW YORK CITY				
Population Density (home)	-0.15	0.20	0.15	-0.53
Subway Lines (home)	-0.04	0.02	-0.06	0.06
Subway Lines (home and work)	-0.24	0.07	-0.06	0.00
Car Commute Cost (w/o parking)	-0.20	0.05	0.00	-0.06
Car Commute Cost (incl. parking)	-0.32	0.08	-0.02	-0.10
Non-Car Commute Cost	0.13	-0.03	0.00	0.04
Car Commute Time	0.17	-0.05	0.02	0.05
Non-Car Commute Time	0.17	-0.04	0.00	0.04
Income	n/a	-0.63	0.13	0.79
MANHATTAN ONLY				
Population Density (home)	-0.30	0.17	-0.19	-0.94
Subway Lines (home)	-0.21	0.03	-0.08	0.08
Subway Lines (home and work)	-0.68	0.05	-0.10	-0.08
Car Commute Cost (w/o parking)	-0.21	0.00	-0.02	-0.05
Car Commute Cost (incl. parking)	-0.60	0.04	-0.06	-0.16
Non-Car Commute Cost	0.20	0.00	0.00	0.06
Car Commute Time	0.20	0.00	0.02	0.04
Non-Car Commute Time	0.48	-0.02	0.03	0.04
Income	n/a	-0.41	0.64	1.38
STATEN ISLAND ONLY				
Population Density (home)	-0.02	0.21	0.14	-0.08
Subway Lines (home)	0.00	0.00	0.00	0.00
Subway Lines (home and work)	-0.11	0.09	0.05	-0.03
Car Commute Cost (w/o parking)	-0.17	0.18	0.04	-0.03
Car Commute Cost (incl. parking)	-0.27	0.28	0.08	-0.06
Non-Car Commute Cost	0.08	-0.10	-0.02	0.02
Car Commute Time	0.11	-0.14	-0.02	0.02
Non-Car Commute Time	0.06	-0.05	0.00	0.00
Income	n/a	-1.56	-0.55	0.37

	Car Use	**Zero Car**	**One Car**	**Two+ Car**
REST OF NEW YORK CITY				
Population Density (home)	-0.16	0.21	0.19	-0.60
Subway Lines (home)	-0.03	0.02	-0.06	0.08
Subway Lines (home and work)	-0.23	0.07	-0.06	0.00
Car Commute Cost (w/o parking)	-0.20	0.06	-0.02	-0.06
Car Commute Cost (incl. parking)	-0.31	0.10	-0.02	-0.10
Non-Car Commute Cost	0.14	-0.04	0.00	0.04
Car Commute Time	0.18	-0.06	0.02	0.06
Non-Car Commute Time	0.17	-0.04	0.00	0.05
Income	n/a	-0.70	0.08	0.85

Table A.7: Elasticities of car ownership and car use for commuting in Joint Car Ownership and Mode Choice Model By Income Level

	Car Use		**Zero Car**		**One Car**		**Two+ Car**	
	Low	High	Low	High	Low	High	Low	High
FIVE BOROUGHS OF NEW YORK CITY								
Population Density (home)	-0.19	-0.06	0.18	0.24	0.28	-0.09	-0.75	-0.13
Subway Lines (home)	-0.02	-0.06	0.02	0.04	-0.09	-0.02	0.11	0.00
Subway Lines (home and work)	-0.20	-0.30	0.06	0.09	-0.09	0.00	0.04	-0.10
Car Commute Cost (w/o parking)	-0.20	-0.20	0.05	0.04	-0.02	0.00	-0.06	-0.05
Car Commute Cost (incl. parking)	-0.29	-0.37	0.08	0.09	-0.04	0.00	-0.09	-0.11
Non-Car Commute Cost	0.14	0.12	-0.04	-0.03	0.02	0.00	0.04	0.04
Car Commute Time	0.21	0.11	-0.06	-0.02	0.03	0.00	0.06	0.03
Non-Car Commute Time	0.13	0.25	-0.03	-0.06	0.02	0.00	0.03	0.06
Income	n/a	n/a	-0.66	-0.54	0.14	0.13	0.99	0.44
MANHATTAN ONLY								
Population Density (home)	-0.27	-0.32	0.07	0.23	0.10	-0.34	-2.11	-0.56
Subway Lines (home)	-0.19	-0.22	0.03	0.03	-0.13	-0.06	0.42	-0.03
Subway Lines (home and work)	-0.56	-0.74	0.04	0.06	-0.15	-0.07	0.28	-0.20
Car Commute Cost (w/o parking)	-0.30	-0.16	0.02	0.00	-0.04	0.00	-0.07	-0.04
Car Commute Cost (incl. parking)	-0.58	-0.62	0.04	0.04	-0.07	-0.06	-0.16	-0.17
Non-Car Commute Cost	0.21	0.18	0.00	0.00	0.02	0.00	0.06	0.06
Car Commute Time	0.30	0.15	-0.02	0.00	0.03	0.00	0.07	0.03
Non-Car Commute Time	0.33	0.57	0.00	-0.02	0.02	0.03	0.00	0.05
Income	n/a	n/a	-0.40	-0.41	0.74	0.58	2.04	1.17

	Car Use		Zero Car		One Car		Two+ Car	
	Low	High	Low	High	Low	High	Low	High
STATEN ISLAND ONLY								
Population Density (home)	-0.04	0.00	0.24	0.08	0.28	0.02	-0.13	0.00
Subway Lines (home)	0.00	0.00	0.00	-0.02	0.00	0.00	0.00	0.00
Subway Lines (home and work)	-0.10	-0.13	0.07	0.17	0.03	0.06	-0.02	-0.04
Car Commute Cost (w/o parking)	-0.16	-0.17	0.17	0.24	0.04	0.05	-0.03	-0.04
Car Commute Cost (incl. parking)	-0.26	-0.28	0.26	0.40	0.06	0.10	-0.05	-0.06
Non-Car Commute Cost	0.08	0.07	-0.10	-0.10	-0.02	-0.02	0.02	0.02
Car Commute Time	0.13	0.08	-0.15	-0.12	-0.03	-0.02	0.03	0.02
Non-Car Commute Time	0.03	0.09	-0.04	-0.12	0.00	-0.03	0.00	0.02
Income	n/a	n/a	-1.66	-1.03	-0.86	-0.27	0.50	0.18
REST OF NEW YORK CITY								
Population Density (home)	-0.21	-0.05	0.20	0.25	0.30	-0.03	-0.84	-0.11
Subway Lines (home)	-0.02	-0.05	0.00	0.05	-0.09	-0.02	0.12	0.00
Subway Lines (home and work)	-0.21	-0.29	0.06	0.14	-0.09	0.00	0.04	-0.10
Car Commute Cost (w/o parking)	-0.20	-0.20	0.06	0.09	-0.02	0.00	-0.06	-0.06
Car Commute Cost (incl. parking)	-0.28	-0.37	0.09	0.16	-0.04	0.00	-0.10	-0.11
Non-Car Commute Cost	0.14	0.12	-0.04	-0.05	0.02	0.00	0.05	0.04
Car Commute Time	0.21	0.12	-0.07	-0.05	0.03	0.00	0.07	0.03
Non-Car Commute Time	0.13	0.25	-0.03	-0.11	0.02	0.00	0.03	0.08
Income	n/a	n/a	-0.70	-0.74	0.11	0.02	1.05	0.42

Table A.8: Elasticities of car ownership and car use for commuting in Nested Car Ownership and Mode Choice Model (3 Nests, 3 IV Parameters Estimated)

	Car Use	**Zero Car**	**One Car**	**Two+ Car**
FIVE BOROUGHS OF NEW YORK CITY				
Population Density (home)	-0.15	0.20	0.15	-0.54
Subway Lines (home)	-0.03	0.03	-0.06	0.06
Subway Lines (home and work)	-0.23	0.07	-0.06	0.00
Car Commute Cost (w/o parking)	-0.18	0.04	0.00	-0.05
Car Commute Cost (incl. parking)	-0.31	0.07	-0.02	-0.09
Non-Car Commute Cost	0.12	-0.03	0.00	0.03
Car Commute Time	0.16	-0.04	0.00	0.04
Non-Car Commute Time	0.16	-0.02	0.00	0.02
Income	n/a	-0.63	0.13	0.79
MANHATTAN ONLY				
Population Density (home)	-0.32	0.17	-0.19	-0.97
Subway Lines (home)	-0.20	0.04	-0.09	0.05
Subway Lines (home and work)	-0.67	0.06	-0.11	-0.11
Car Commute Cost (w/o parking)	-0.20	0.00	-0.02	-0.04
Car Commute Cost (incl. parking)	-0.61	0.04	-0.05	-0.16
Non-Car Commute Cost	0.19	0.00	0.00	0.05
Car Commute Time	0.21	0.00	0.02	0.04
Non-Car Commute Time	0.45	0.00	0.02	0.04
Income	n/a	-0.40	0.63	1.37
STATEN ISLAND ONLY				
Population Density (home)	-0.02	0.21	0.14	-0.08
Subway Lines (home)	0.00	0.00	0.00	0.00
Subway Lines (home and work)	-0.10	0.09	0.04	-0.02
Car Commute Cost (w/o parking)	-0.16	0.15	0.04	-0.03
Car Commute Cost (incl. parking)	-0.26	0.24	0.08	-0.05
Non-Car Commute Cost	0.07	-0.08	-0.02	0.00
Car Commute Time	0.10	-0.12	-0.02	0.02
Non-Car Commute Time	0.05	-0.02	0.00	0.00
Income	n/a	-1.56	-0.55	0.37

	Car Use	**Zero Car**	**One Car**	**Two+ Car**
REST OF NEW YORK CITY				
Population Density (home)	-0.16	0.21	0.19	-0.61
Subway Lines (home)	-0.03	0.02	-0.06	0.07
Subway Lines (home and work)	-0.23	0.08	-0.06	0.00
Car Commute Cost (w/o parking)	-0.19	0.05	0.00	-0.05
Car Commute Cost (incl. parking)	-0.30	0.08	0.00	-0.10
Non-Car Commute Cost	0.13	-0.04	0.00	0.04
Car Commute Time	0.17	-0.05	0.02	0.05
Non-Car Commute Time	0.16	-0.03	0.00	0.03
Income	n/a	-0.70	0.08	0.85

Table A.9: Elasticities of car ownership and car use for commuting in Nested Car Ownership and Mode Choice Model (3 Nests, 3 IV Parameters Estimated) By Income Level

	Car Use		**Zero Car**		**One Car**		**Two+ Car**	
	Low	High	Low	High	Low	High	Low	High
FIVE BOROUGHS OF NEW YORK CITY								
Population Density (home)	-0.20	-0.07	0.19	0.24	0.28	-0.09	-0.76	-0.14
Subway Lines (home)	-0.02	-0.05	0.02	0.04	-0.09	-0.02	0.10	0.00
Subway Lines (home and work)	-0.20	-0.29	0.06	0.10	-0.10	0.00	0.04	-0.10
Car Commute Cost (w/o parking)	-0.18	-0.19	0.04	0.03	-0.02	0.00	-0.05	-0.05
Car Commute Cost (incl. parking)	-0.28	-0.37	0.07	0.08	-0.02	0.00	-0.08	-0.10
Non-Car Commute Cost	0.13	0.11	-0.03	-0.02	0.00	0.00	0.04	0.03
Car Commute Time	0.19	0.12	-0.05	-0.02	0.02	0.00	00.05	0.03
Non-Car Commute Time	0.12	0.22	-0.02	-0.04	0.00	0.00	0.02	0.04
Income	n/a	n/a	-0.66	-0.54	0.13	0.13	0.99	0.44
MANHATTAN ONLY								
Population Density (home)	-0.30	-0.34	0.07	0.23	0.10	-0.34	-2.12	-0.59
Subway Lines (home)	-0.18	-0.21	0.04	0.04	-0.14	-0.06	0.37	-0.05
Subway Lines (home and work)	-0.56	-0.74	0.05	0.06	-0.16	-0.08	0.24	-0.22
Car Commute Cost (w/o parking)	-0.29	-0.15	0.00	0.00	-0.03	0.00	-0.06	-0.04
Car Commute Cost (incl. parking)	-0.57	-0.63	0.03	0.04	-0.06	-0.04	-0.16	-0.17
Non-Car Commute Cost	0.21	0.18	0.00	0.00	0.00	0.00	0.05	0.05
Car Commute Time	0.29	0.16	0.00	0.00	0.02	0.00	0.06	0.04
Non-Car Commute Time	0.31	0.53	0.00	-0.02	0.00	0.02	0.00	0.05
Income	n/a	n/a	-0.40	-0.41	0.74	0.58	2.02	1.16

	Car Use		Zero Car		One Car		Two+ Car	
	Low	High	Low	High	Low	High	Low	High
STATEN ISLAND ONLY								
Population Density (home)	-0.03	0.00	0.24	0.09	0.28	0.02	-0.13	0.00
Subway Lines (home)	0.00	0.00	0.00	-0.02	0.00	0.00	0.00	0.00
Subway Lines (home and work)	-0.09	-0.12	0.07	0.16	0.02	0.06	-0.02	-0.04
Car Commute Cost (w/o parking)	-0.15	-0.16	0.14	0.20	0.03	0.04	-0.03	-0.03
Car Commute Cost (incl. parking)	-0.25	-0.27	0.22	0.34	0.06	0.09	-0.05	-0.06
Non-Car Commute Cost	0.08	0.07	-0.08	-0.09	0.00	-0.02	0.00	0.00
Car Commute Time	0.12	0.08	-0.12	-0.11	-0.02	-0.02	0.02	0.02
Non-Car Commute Time	0.03	0.08	0.00	-0.07	0.00	-0.02	0.00	0.00
Income	n/a	n/a	-1.66	-1.03	-0.86	-0.27	0.51	0.18
REST OF NEW YORK CITY								
Population Density (home)	-0.21	-0.06	0.21	0.26	0.30	-0.02	-0.84	-0.12
Subway Lines (home)	-0.02	-0.05	0.02	0.06	-0.09	0.00	0.11	0.00
Subway Lines (home and work)	-0.20	-0.28	0.06	0.15	-0.10	0.00	0.04	-0.10
Car Commute Cost (w/o parking)	-0.19	-0.20	0.05	0.07	-0.02	0.00	-0.05	-0.05
Car Commute Cost (incl. parking)	-0.27	-0.36	0.07	0.14	-0.02	0.00	-0.09	-0.11
Non-Car Commute Cost	0.14	0.12	-0.03	-0.04	0.00	0.00	0.04	0.03
Car Commute Time	0.20	0.12	-0.05	-0.05	0.02	0.00	0.06	0.03
Non-Car Commute Time	0.12	0.22	-0.02	-0.08	0.00	0.00	0.02	0.05
Income	n/a	n/a	-0.70	-0.74	0.11	0.02	1.06	0.42

Table A.10: Elasticities of car ownership and car use for commuting in Nested Mode and Car Ownership Choice Model (7 Nests, 7 IV Parameters Estimated)

	Car Use	**Zero Car**	**One Car**	**Two+ Car**
FIVE BOROUGHS OF NEW YORK CITY				
Population Density (home)	-0.16	0.20	0.15	-0.52
Subway Lines (home)	-0.03	0.02	-0.06	0.06
Subway Lines (home and work)	-0.23	0.06	-0.06	0.00
Car Commute Cost (w/o parking)	-0.20	0.05	-0.02	-0.06
Car Commute Cost (incl. parking)	-0.32	0.08	-0.02	-0.10
Non-Car Commute Cost	0.13	-0.03	0.00	0.04
Car Commute Time	0.17	-0.05	0.02	0.05
Non-Car Commute Time	0.16	-0.04	0.02	0.04
Income	n/a	-0.62	0.13	0.80
MANHATTAN ONLY				
Population Density (home)	-0.35	0.17	-0.20	-0.92
Subway Lines (home)	-0.20	0.03	-0.08	0.09
Subway Lines (home and work)	-0.66	0.05	-0.10	-0.07
Car Commute Cost (w/o parking)	-0.21	0.00	-0.02	-0.04
Car Commute Cost (incl. parking)	-0.59	0.04	-0.06	-0.13
Non-Car Commute Cost	0.20	0.00	0.00	0.06
Car Commute Time	0.22	0.00	0.02	0.04
Non-Car Commute Time	0.46	0.00	0.03	0.00
Income	n/a	-0.40	0.62	1.30
STATEN ISLAND ONLY				
Population Density (home)	-0.02	0.21	0.14	-0.08
Subway Lines (home)	0.00	0.00	0.00	0.00
Subway Lines (home and work)	-0.11	0.09	0.04	-0.03
Car Commute Cost (w/o parking)	-0.17	0.18	0.05	-0.04
Car Commute Cost (incl. parking)	-0.27	0.28	0.09	-0.06
Non-Car Commute Cost	0.08	-0.09	-0.02	0.02
Car Commute Time	0.11	-0.14	-0.03	0.02
Non-Car Commute Time	0.05	-0.10	-0.02	0.02
Income	n/a	-1.52	-0.58	0.37

	Car Use	**Zero Car**	**One Car**	**Two+ Car**
REST OF NEW YORK CITY				
Population Density (home)	-0.18	0.21	0.18	-0.59
Subway Lines (home)	-0.02	0.00	-0.06	0.08
Subway Lines (home and work)	-0.22	0.07	-0.05	0.00
Car Commute Cost (w/o parking)	-0.20	0.06	-0.02	-0.06
Car Commute Cost (incl. parking)	-0.31	0.10	-0.02	-0.10
Non-Car Commute Cost	0.14	-0.04	0.00	0.04
Car Commute Time	0.18	-0.06	0.02	0.06
Non-Car Commute Time	0.17	-0.05	0.02	0.05
Income	n/a	-0.70	0.08	0.87

Table A.11: Elasticities of car ownership in Joint Residential Location and Car Ownership Choice Model

	Zero-Car Ownership	One-Car Ownership	Two+ Car Ownership
FIVE BOROUGHS OF NEW YORK CITY			
Population Density (home)	0.19	0.10	-0.44
Subway Lines (home)	0.04	-0.06	0.04
Income	-0.45	0.06	0.62
MANHATTAN ONLY			
Population Density (home)	0.26	-0.14	-0.68
Subway Lines (home)	0.00	-0.04	0.06
Income	-0.40	0.28	0.85
STATEN ISLAND ONLY			
Population Density (home)	0.28	0.16	-0.22
Subway Lines (home)	0.04	-0.04	0.02
Income	-0.69	-0.24	0.42
REST OF NEW YORK CITY			
Population Density (home)	0.16	0.13	-0.45
Subway Lines (home)	0.04	-0.06	0.04
Income	-0.46	0.04	0.62

Table A.12: Elasticities of car ownership in Joint Residential Location and Car Ownership Choice Model By Income Level

	Zero Car Ownership		One Car Ownership		Two+ Car Ownership	
	Low	High	Low	High	Low	High
FIVE BOROUGHS OF NEW YORK CITY						
Population Density (home)	0.13	0.33	0.22	-0.12	-0.57	-0.20
Subway Lines (home)	0.06	-0.02	-0.09	0.00	0.04	0.04
Income	-0.48	-0.36	0.05	0.08	0.78	0.31
MANHATTAN ONLY						
Population Density (home)	0.13	0.35	0.14	-0.29	-0.95	-0.51
Subway Lines (home)	0.06	-0.02	-0.12	0.00	0.04	0.08
Income	-0.46	-0.36	0.30	0.27	1.22	0.62
STATEN ISLAND ONLY						
Population Density (home)	0.22	0.41	0.30	0.00	-0.31	-0.11
Subway Lines (home)	0.06	0.00	-0.06	0.00	0.02	0.00
Income	-0.82	-0.43	-0.39	-0.08	0.60	0.18
REST OF NEW YORK CITY						
Population Density (home)	0.13	0.31	0.23	-0.07	-0.58	-0.16
Subway Lines (home)	0.06	-0.02	-0.09	0.00	0.04	0.04
Income	-0.47	-0.37	0.05	0.02	0.78	0.27

Table A.13: Elasticities of car ownership in Nested Residential Location and Car Ownership Choice Model (11 Nests, 1 IV Parameter Estimated)

	Zero-Car Ownership	**One-Car Ownership**	**Two+ Car Ownership**
FIVE BOROUGHS OF NEW YORK CITY			
Population Density (home)	0.18	0.20	-0.58
Subway Lines (home)	0.00	-0.06	0.08
Income	-0.52	0.08	0.67
MANHATTAN ONLY			
Population Density (home)	0.19	-0.04	-0.62
Subway Lines (home)	0.02	-0.04	0.05
Income	-0.46	0.32	0.84
STATEN ISLAND ONLY			
Population Density (home)	0.27	0.28	-0.31
Subway Lines (home)	0.00	-0.05	0.04
Income	-0.73	-0.22	0.42
REST OF NEW YORK CITY			
Population Density (home)	0.17	0.24	-0.61
Subway Lines (home)	0.00	-0.06	0.10
Income	-0.54	0.05	0.69

Table A.14: Elasticities of car ownership in Nested Car Ownership and Residential Location Choice Model (3 Nests, 3 IV Parameters Estimated)

	Zero-Car Ownership	**One-Car Ownership**	**Two+ Car Ownership**
FIVE BOROUGHS OF NEW YORK CITY			
Population Density (home)	0.25	0.20	-0.71
Subway Lines (home)	0.09	-0.12	0.05
Income	-0.44	0.06	0.60
MANHATTAN ONLY			
Population Density (home)	0.28	-0.09	-1.10
Subway Lines (home)	0.08	-0.11	0.00
Income	-0.36	0.30	0.80
STATEN ISLAND ONLY			
Population Density (home)	0.42	0.33	-0.35
Subway Lines (home)	0.10	-0.07	0.02
Income	-0.70	-0.26	0.37
REST OF NEW YORK CITY			
Population Density (home)	0.23	0.25	-0.73
Subway Lines (home)	0.09	-0.12	0.06
Income	-0.46	0.04	0.61

Table A.15: Elasticities of car use for commuting in Joint Residential Location and Mode Choice Model By Income

	Car Use for Commuting		
	All Income	**Low Income**	**High Income**
FIVE BOROUGHS OF NEW YORK CITY			
Population Density (home)	-0.02	0.00	-0.02
Subway Lines (home)	-0.02	-0.02	-0.03
Subway Lines (home and work)	-0.17	-0.14	-0.21
Car Commute Cost (w/o parking)	-0.38	-0.37	-0.39
Car Commute Cost (incl. parking)	-0.44	-0.42	-0.48
Non-Car Commute Cost	0.23	0.24	0.22
Car Commute Time	-0.36	-0.32	-0.42
Non-Car Commute Time	0.84	0.81	0.89
MANHATTAN ONLY			
Population Density (home)	-0.03	-0.02	-0.03
Subway Lines (home)	-0.06	-0.04	-0.07
Subway Lines (home and work)	-0.40	-0.32	-0.45
Car Commute Cost (w/o parking)	-0.64	-0.74	-0.60
Car Commute Cost (incl. parking)	-0.86	-0.89	-0.85
Non-Car Commute Cost	0.36	0.39	0.34
Car Commute Time	-0.60	-0.55	-0.62
Non-Car Commute Time	1.29	1.28	1.30
STATEN ISLAND ONLY			
Population Density (home)	-0.02	0.00	-0.02
Subway Lines (home)	0.00	0.00	0.00
Subway Lines (home and work)	-0.08	-0.06	-0.09
Car Commute Cost (w/o parking)	-0.31	-0.32	-0.29
Car Commute Cost (incl. parking)	-0.36	-0.37	-0.35
Non-Car Commute Cost	0.16	0.18	0.15
Car Commute Time	-0.26	-0.24	-0.30
Non-Car Commute Time	0.61	0.60	0.63
REST OF NEW YORK CITY			
Population Density (home)	-0.02	0.00	-0.02
Subway Lines (home)	-0.02	-0.02	-0.03
Subway Lines (home and work)	-0.16	-0.15	-0.21
Car Commute Cost (w/o parking)	-0.37	-0.37	-0.38
Car Commute Cost (incl. parking)	-0.43	-0.41	-0.46
Non-Car Commute Cost	0.24	0.24	0.22
Car Commute Time	-0.35	-0.32	-0.42
Non-Car Commute Time	0.84	0.82	0.88

Table A.16: Elasticities of car use for commuting in Nested Residential Location and Mode Choice Model By Income (11 Nests, 1 IV Parameter estimated)

	Car Use for Commuting		
	All Income	Low Income	High Income
FIVE BOROUGHS OF NEW YORK CITY			
Population Density (home)	0.00	0.00	-0.02
Subway Lines (home)	-0.02	-0.02	-0.02
Subway Lines (home and work)	-0.16	-0.14	-0.20
Car Commute Cost (w/o parking)	-0.30	-0.30	-0.32
Car Commute Cost (incl. parking)	-0.38	-0.36	-0.43
Non-Car Commute Cost	0.20	0.20	0.19
Car Commute Time	-0.13	-0.12	-0.16
Non-Car Commute Time	0.52	0.50	0.55
MANHATTAN ONLY			
Population Density (home)	-0.02	-0.02	-0.03
Subway Lines (home)	-0.05	-0.04	-0.06
Subway Lines (home and work)	-0.36	-0.29	-0.40
Car Commute Cost (w/o parking)	-0.52	-0.58	-0.48
Car Commute Cost (incl. parking)	-0.76	-0.78	-0.76
Non-Car Commute Cost	0.29	0.39	0.28
Car Commute Time	-0.22	-0.20	-0.23
Non-Car Commute Time	0.77	0.76	0.78
STATEN ISLAND ONLY			
Population Density (home)	0.00	0.00	0.00
Subway Lines (home)	0.00	0.00	0.00
Subway Lines (home and work)	-0.08	-0.07	-0.10
Car Commute Cost (w/o parking)	-0.25	-0.25	-0.25
Car Commute Cost (incl. parking)	-0.31	-0.31	-0.31
Non-Car Commute Cost	0.15	0.16	0.14
Car Commute Time	-0.10	-0.08	-0.12
Non-Car Commute Time	0.39	0.38	0.40
REST OF NEW YORK CITY			
Population Density (home)	0.00	0.00	-0.02
Subway Lines (home)	-0.02	-0.02	-0.02
Subway Lines (home and work)	-0.16	-0.14	-0.20
Car Commute Cost (w/o parking)	-0.30	-0.29	-0.32
Car Commute Cost (incl. parking)	-0.37	-0.35	-0.41
Non-Car Commute Cost	0.20	0.20	0.18
Car Commute Time	-0.13	-0.12	-0.16
Non-Car Commute Time	0.52	0.51	0.55

Table A.17: Elasticities of car use for commuting in Nested Mode and Residential Location Choice Model By Income (7 Nests, 7 IV Parameters estimated)

	Car Use for Commuting		
	All Income	**Low Income**	**High Income**
FIVE BOROUGHS OF NEW YORK CITY			
Population Density (home)	0.00	0.00	0.00
Subway Lines (home)	0.00	0.00	0.00
Subway Lines (home and work)	-0.15	-0.13	-0.18
Car Commute Cost (w/o parking)	-0.34	-0.34	-0.35
Car Commute Cost (incl. parking)	-0.42	-0.39	-0.46
Non-Car Commute Cost	0.21	0.22	0.20
Car Commute Time	-0.32	-0.29	-0.38
Non-Car Commute Time	0.77	0.75	0.82
MANHATTAN ONLY			
Population Density (home)	0.00	-0.02	0.00
Subway Lines (home)	-0.04	-0.03	-0.04
Subway Lines (home and work)	-0.37	-0.30	-0.41
Car Commute Cost (w/o parking)	-0.58	-0.67	-0.53
Car Commute Cost (incl. parking)	-0.83	-0.84	-0.82
Non-Car Commute Cost	0.32	0.36	0.30
Car Commute Time	-0.54	-0.50	-0.56
Non-Car Commute Time	1.19	1.17	1.20
STATEN ISLAND ONLY			
Population Density (home)	0.00	-0.02	0.00
Subway Lines (home)	0.00	0.00	0.02
Subway Lines (home and work)	-0.06	-0.05	-0.06
Car Commute Cost (w/o parking)	-0.28	-0.30	-0.26
Car Commute Cost (incl. parking)	-0.34	-0.35	-0.33
Non-Car Commute Cost	0.15	0.16	0.14
Car Commute Time	-0.24	-0.22	-0.27
Non-Car Commute Time	0.57	0.56	0.59
REST OF NEW YORK CITY			
Population Density (home)	0.00	0.00	0.00
Subway Lines (home)	0.00	0.00	0.00
Subway Lines (home and work)	-0.15	-0.14	-0.18
Car Commute Cost (w/o parking)	-0.34	-0.34	-0.34
Car Commute Cost (incl. parking)	-0.40	-0.38	-0.44
Non-Car Commute Cost	0.22	0.22	0.19
Car Commute Time	-0.32	-0.29	-0.38
Non-Car Commute Time	0.78	0.76	0.82

Table A.18: Elasticities of car use for commuting in Mode Only Choice Model By Income

	Car Use for Commuting		
	All Income	**Low Income**	**High Income**
FIVE BOROUGHS OF NEW YORK CITY			
Subway Lines (home)	-0.04	-0.04	-0.05
Subway Lines (home and work)	-0.20	-0.17	-0.25
Car Commute Cost (w/o parking)	-0.20	-0.20	-0.20
Car Commute Cost (incl. parking)	-0.31	-0.28	-0.37
Non-Car Commute Cost	0.13	0.13	0.12
Car Commute Time	0.16	0.18	0.10
Non-Car Commute Time	0.15	0.11	0.23
MANHATTAN ONLY			
Subway Lines (home)	-0.16	-0.15	-0.17
Subway Lines (home and work)	-0.54	-0.47	-0.58
Car Commute Cost (w/o parking)	-0.20	-0.30	-0.14
Car Commute Cost (incl. parking)	-0.62	-0.63	-0.62
Non-Car Commute Cost	0.18	0.21	0.16
Car Commute Time	0.18	0.30	0.12
Non-Car Commute Time	0.40	0.24	0.48
STATEN ISLAND ONLY			
Subway Lines (home)	0.00	0.00	0.00
Subway Lines (home and work)	-0.10	-0.09	-0.11
Car Commute Cost (w/o parking)	-0.18	-0.19	-0.18
Car Commute Cost (incl. parking)	-0.28	-0.27	-0.28
Non-Car Commute Cost	0.08	0.09	0.08
Car Commute Time	0.11	0.14	0.08
Non-Car Commute Time	0.06	0.03	0.10
REST OF NEW YORK CITY			
Subway Lines (home)	-0.04	-0.04	-0.04
Subway Lines (home and work)	-0.19	-0.17	-0.24
Car Commute Cost (w/o parking)	-0.20	-0.19	-0.21
Car Commute Cost (incl. parking)	-0.30	-0.27	-0.36
Non-Car Commute Cost	0.13	0.14	0.12
Car Commute Time	0.16	0.19	0.11
Non-Car Commute Time	0.15	0.11	0.22

Table A.19: Elasticities of car ownership in Car Ownership Only Choice Model

	Zero-Car Ownership	One-Car Ownership	Two+ Car Ownership
FIVE BOROUGHS OF NEW YORK CITY			
Population Density (home)	0.16	0.19	-0.54
Subway Lines (home)	0.00	-0.05	0.08
Income	-0.50	0.08	0.67
MANHATTAN ONLY			
Population Density (home)	0.14	-0.15	-0.86
Subway Lines (home)	0.00	-0.03	0.16
Income	-0.32	0.50	1.06
STATEN ISLAND ONLY			
Population Density (home)	0.19	0.15	-0.08
Subway Lines (home)	0.00	0.00	0.00
Income	-1.26	-0.51	0.33
REST OF NEW YORK CITY			
Population Density (home)	0.17	0.23	-0.61
Subway Lines (home)	0.00	-0.05	0.10
Income	-0.57	0.04	0.72

Table A.20: Elasticities of car ownership in Car Ownership Only Choice Model By Income Level

	Zero Car Ownership		One Car Ownership		Two+ Car Ownership	
	Low	High	Low	High	Low	High
FIVE BOROUGHS OF NEW YORK CITY						
Population Density (home)	0.15	0.21	0.35	-0.10	-0.79	-0.08
Subway Lines (home)	0.00	0.00	-0.07	0.00	0.12	0.02
Income	-0.54	-0.40	0.08	0.09	0.86	0.33
MANHATTAN ONLY						
Population Density (home)	0.06	0.20	0.16	-0.32	-2.28	-0.39
Subway Lines (home)	0.00	0.00	-0.08	0.00	0.47	0.06
Income	-0.33	-0.31	0.64	0.43	1.71	0.85
STATEN ISLAND ONLY						
Population Density (home)	0.21	0.08	0.30	0.00	-0.13	0.00
Subway Lines (home)	0.00	0.00	-0.02	0.00	0.00	0.00
Income	-1.34	-0.82	-0.84	-0.22	0.46	0.14
REST OF NEW YORK CITY						
Population Density (home)	0.16	0.22	0.37	-0.04	-0.88	-0.06
Subway Lines (home)	0.00	0.00	-0.07	0.00	0.13	0.02
Income	-0.58	-0.52	0.06	0.00	0.92	0.32