ABSTRACT

Title of Dissertation: THE INTERACTION BETWEEN DISTANCE TO WORK AND VEHICLE MILES TRAVELED.

Hernán Mauricio González
PhD in Agricultural and Resource Economics
2007

Dissertation Directed By: Dr. John K. Horowitz

Economists have long been concerned with the externalities generated by automobiles, such as traffic congestion and air pollution. Since many of these externalities are closely bound up with the number of miles being driven, economists have been much interested in the behavior of what is known as *vehicle miles traveled* (VMT). Planners believe that land use can be manipulated to serve congestion management, air quality or related transport planning goals. The underlying idea is that household location may have a big impact on its transportation demand, including car ownership. In this context, I focus on distance to work (DTW) as the measure of household location. I chose a continuous measure of household location instead of a discrete one because, besides being easily measured, it matches better the data available for this study and it has a very straightforward interpretation—it allows me to calculate the contribution of commuting miles to total miles driven.

Despite the clear conceptual connection between DTW and VMT, and the constraining nature of household location, little is known about their joint behavior. City and household level attributes that may lead households to live close or far from their work may also lead them to drive few or many miles for non-commuting purposes. This effect must be accounted for when measuring the behavior of VMT conditional on DTW. I develop two models to analyze: (i) the role of city characteristics in explaining households' distance to work, (ii) the effect of distance to work on VMT and car ownership, (iii) the effect of city level attributes on VMT, conditional on DTW, (iv) the unobserved taste for driving, (v) differences between workers and non-workers. I find that: (i) City characteristics expected to affect commutes have a small effect on households' DTW, (ii) DTW provides an important effect on car ownership levels and VMT, (iii) City characteristics expected to influence non-commute miles have a small impact on VMT, (iv) taste for driving has a small but significant effect on VMT, and (v) non-workers are much less responsive to gas prices than workers.

THE INTERACTION BETWEEN DISTANCE TO WORK AND VEHICLE MILES TRAVELED

By

Hernán Mauricio González

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Dr. John K. Horowitz
Dr. Anna Alberini
Dr. Kenneth McConnell
Dr. Lori Lynch
Dr. Maureen Cropper

Dedication

To Kat and Dani, the two reasons I need to accomplish anything.

Acknowledgements

This dissertation has benefited from the input of many people over the last couple of years. Drs. Horowitz and Alberini provided great advice and insight in the development of the thesis prospectus and in the first period of the dissertation. I also benefited from comments provided by participants at two seminars held at the University of Maryland's Climate Change Center and during the 2005 meeting of the Association of Environmental and Resource Economists. Additionally, I would like to thank my dissertation committee for their time and comments that have improved the final version of the dissertation. But most of all, I'm deeply indebted to Professor John Horowitz, who never gave up on me, even though I had found many reasons to do so.

Table of Contents

List of Tables

List of Figures

1 Introduction

Economists have long been concerned with the externalities generated by automobiles. The externalities are exceptionally varied and are well detailed in the literature (Parry et al, 2007). Among these one can find environmental externalities (local and global air pollution, noise, indirect water pollution, and improper disposal of vehicles and parts), and non-environmental externalities such as traffic congestion, traffic accidents, highway maintenance costs, parking subsidies and urban sprawl. The wide variety of these externalities makes automobiles prime candidates for regulation and for analysis. Many of these externalities are closely bound up with the number of miles being driven, and economists have therefore been much interested in the behavior of what is known as *vehicle miles traveled* (VMT). VMT have also been fertile ground for study because household-level VMT are relatively easily observed, more so than most of the specific externality-causing activities.

Planners believe that there is a potential of reducing traffic and congestion problems (and indirectly air pollution) in modern cities by altering land use. In this context, theories such as “smart growth” and “livable communities” have been developed in order to reduce urban sprawl by promoting growth in city centers. Smart Growth also promotes transit oriented, bicycle friendly communities, including mixed-use development. Similar movements such as “New Urbanism” are also built on the idea that land use can be manipulated to serve congestion management, air quality or related transport planning goals. New Urbanist designs feature higher neighborhood densities, a mix of commercial and residential uses and street patterns along a grid. This has given rise to a great deal of literature dealing with how urban shape affects vehicle ownership and travel demand. Steiner (1994), Wilson (1998), and, Badoe and Miller (2000) present recent surveys of the literature on the interaction between land use and transportation.

1.1 Objective of the dissertation

In this context, this dissertation examines one particular aspect of transportation and land use interaction—it focuses on the interaction between vehicle miles traveled and household location. In particular, I use distance to work (DTW) as my measure of household location. I chose a continuous measure of household location instead of a discrete one because, besides being easily measured, it has a very straightforward interpretation—it allows me to calculate both the contribution of commuting miles to total miles driven and a measure of taste for driving.

Despite the clear conceptual connection between DTW and VMT, and the constraining nature of household location, little is known about their joint behavior. City- and household-level attributes that may lead households to live close or far from their work may also lead them to drive few or many miles for non-commuting purposes. This effect must be accounted for when measuring the behavior of VMT conditional on DTW. Previous literature has largely ignored the role of household location on car ownership and VMT.

Some recent papers such as Bento et al (2005, from now on BCMV) have highlighted the issue of endogeneity of location and addressed it by constructing city-wide measures of urban form, which are then taken to be exogenous to the household. They use variables such as density at the city level and other measures of urban form (jobs-housing balance, population centrality, city shape, land area and supply of rail and bus transit) to analyze the impact of urban form and transit supply on commute mode and vehicle miles traveled.

One case in which household location and car ownership/VMT were modeled simultaneously is that of Sermons and Seredich (2001). These authors estimate their model sequentially: First, a discrete model of household location (households choose between six clusters representing San Francisco) and car ownership is fitted using a conditional logit approach; and second, the VMT model is estimated including a regressor representing the predicted number of cars owned from the first part of the model. by the householdthe predicted number of cars owned are obtained from the multinomial logit model and is included in the VMT equation as a regressor. A third paper addressing the issue of endogeneity of household location is that by Schimek (1996)—the author includes population density at the neighborhood level and uses an instrumental variable approach to deal with the endogenous nature of this variable. The author uses a serried of dummy variables for city size to instrument for density.

Finally, in this dissertation I address the following issues: (i) the role of city characteristics in explaining distance to work, (ii) the effect of distance to work on VMT, (iii) the taste for driving, (iv) the effect of city characteristics on VMT conditional on DTW, and (v) the differences in VMT between workers and non workers.

1.2 Contribution to the literature on transportation and land use

The first contribution of this dissertation is that it expands the literature on household location and the literature on travel demand. In particular, in Chapters 4 and 5 of the dissertation present two models of the relationship between vehicle miles traveled and distance to work. There are two main differences between the work of BCMV and this dissertation: First, I develop a model where distance to work and vehicle miles traveled are estimated simultaneously (Chapter 4). While BCMV address the issue of endogeneity of household location by building exogenous measures of land use, I deal with this issue by explicitly modeling the location decision. Second, I analyze the effects of different measures of city characteristics on VMT conditional on DTW. This dissertation differs from Sermons and Seredich (2001) in that I use a continuous measure for household location and therefore a different approach to estimate the household location/VMT system.

Additionally, I reach the following general conclusions:

(i) City characteristics that might be expected to affect commutes have remarkably little effect on households' distance to work. Only my measure of congestion (median speed at the city level) has a substantial effect on DTW. The elasticity of DTW with respect to median speed is 0.083. Variables like city shape, city area, population density, or the joint jobs-housing distribution have little apparent effect on city-average distance to work. Section 4.5.1 expands on this issue.

(ii) Distance to work has an important effect on overall household vehicle miles traveled. A one percent increase in distance to work implies a 0.18 percent increase in overall VMT. This elasticity is comparable in magnitude to both the income elasticity of VMT and the gas price elasticity of VMT (see table 4.8). This elasticity is best understood in terms of marginal effects: one additional mile of DTW leads to an additional 0.33 VMT. This number is considerably less than one, which implies that as DTW increases, total VMT increase, but non-commute miles decrease. It is easiest to think of this last result as a strong degree of task-sharing.

When VMT is estimated conditional on DTW and car ownership (see section 5.4.4), the implied marginal effect of DTW on VMT increases from 0.33 to 0.42. This in turn suggests a lower amount of task sharing. By allowing car ownership to be a choice variable in the model in chapter 5, households respond to changes in DTW by altering the number of cars they own. This decision translates into a change in the households' overall miles (and in general, a decrease in their non-commute miles when DTW increases).

(iii) By treating distance to work as endogenous, I am able to estimate a parameter that I interpret as a "taste for driving".[1] This dissertation presents, to my knowledge, one of the first attempts at estimating this unobservable characteristic. I show that when we do not take taste for driving into account, we overestimate the effect of DTW on VMT (the coefficient changes from 0.33 to 0.36 when taste for driving is not included). The estimate of taste for driving is positive, as expected, but quite small. Either this taste is unimportant or it is adequately captured by other, included variables.

(iv) City characteristics that were expected to influence non-commute miles have a small impact on VMT, conditional on DTW. Variables such as city-level population density, which represents the density of friends and other non-commuting destinations, has a statistically significant but extremely small negative coefficient.

---

[1] This parameter is the correlation between the equations representing distance to work and vehicle miles traveled (equations 2 and 3 in section 4.3.2).

These results hold for both models developed in chapters 4 and 5.

(v) Finally, I explore the travel behavior of workers vs. non-workers. The most striking difference between these two groups is on the coefficient on gas price—non-workers are much less responsive to gas price changes than workers. This result suggests that workers have a higher degree of task-sharing, an option may contribute the higher gas price elasticity for workers.

1.3 Description of subsequent chapters

This dissertation consists of a literature review and 2 models of the interaction between VMT and DTW. Chapter 2 reviews the literature on transportation and land use. Chapter 3 describes the data available for the study. In chapter 4 I develop a simultaneous model of DTW and VMT, conditional on the work status of the household. In chapter 5, I estimate a continuous discrete model of car ownership and VMT, conditional on DTW. Finally, chapter 6 presents concluding comments and directions for future research.

2 Literature review

Existing land use and transportation interaction models draw from three modeling traditions, namely urban economics, spatial interaction or gravity models, and discrete choice based on random utility theory (Eliasson and Mattsson, 2000). These three modeling traditions were developed based on the type of question they wanted to answer. Urban economics explains the functioning of a city from an analytical point of view. This branch of microeconomic theory accounts for spatial relationships between individuals and organizations in order to understand the economic reasons behind the formation, functioning and development of cities (O'Sullivan, 2006). While urban economic models were developed with great mathematical rigor, they initially led to virtually no operational models. On the other hand, spatial interaction models of the Lowry type (or gravity models), were developed to allow planners to make rough forecasts of flows between different locations. Spatial interaction models are used to study are the flow of goods (e.g. trade patterns, freight distribution), workers (e.g. journey to work, migration), and transmission of information or capital, among others. Spatial interaction models were typically applied models, placing little importance theoretical content. Finally, discrete choice models based on random utility theory were first introduced to the field of travel demand by McFadden (1974) and Domencich and McFadden (1975). Early applications dealt with transport problems such as mode choice and destination choice. Discrete choice models typically divide the decisions involved in a trip into four steps: whether to travel, destination choice, mode choice, and route choice (Eliasson and Mattsson, 2000).

More recently, and following the work of Mannering and Winston (1985) and Train (1986), continuous/discrete models were applied to the simultaneously analysis of car ownership and travel demand. This type of analysis was possible thanks to the work done by Heckman (1979) and Dubin and McFadden (1984), who developed methods for specifying and estimating models that describe continuous/discrete situations. Recent examples of papers following these methods are Schimek (1996), Kockelman (1997) and Bento et al (2005). This framework allowed researchers analyze the effects of different policy variables on car ownership and travel demand. Recent advances in GIS allowed researchers to model the effect of different measures of urban form on travel demand.

As described in Chapter 1, the central question of this dissertation is to study the effect that household location (proxied by distance to work) has on automobile ownership and travel demand. This dissertation uses a variation of the existing models of transportation and land use interactions described below (Section 2.2). In particular, in Chapter 4 I estimate a model in which household location and travel demand are chosen simultaneously. Chapter 4 presents an extension to the existing literature in that household location is typically assumed exogenous. It is a first attempt at bringing together the models described in section 2.1 and 2.2. Finally, Chapter 5 extends models described in section 2.3 by explicitly modeling the effect of household location on travel demand and car ownership.

The remainder of the chapter is organized as follows: the next section describes the basics of the theory of residential location, starting with the monocentric model of household location (Section 2.1.1) and continuing with discrete choice models (Section 2.1.2). Spatial interaction models are not analyzed separately because, as Anas (1983) and Mattsson (1984) suggest, the entropy and the logit approaches are identical for practical purposes. Section 2.2 reviews the literature that analyses the interaction between transportation and land use. Models used to study this interaction typically assume that household location is exogenous and focus mainly on the effects of different measures of urban form on vehicle ownership and demand for miles

2.1 The theory of residential location

There are two basic approaches when dealing with the theory of residential location. The first deals with inter-urban location and the second with intra-urban moves. The question of why households locate in a particular city is out of the reach of this dissertation, though one of the main reasons households move between cities is a change in employment opportunities (Sjaastad (1962), Greenwood and Hunt (1989)). See the literature on household mobility for the basics of inter-urban location. On the other hand, intra-urban relocation occurs not only because of changes in employment, but also because of changes in supplies and demands for residential site characteristics (Linneman and Graves (1983), Clark and van Lierop (2000)).

2.1.1 The monocentric model of household location

The theory of household location is based on the monocentric model of household location. Its origins can be traced back to the early-19th century with the work of von Thunen (1826). In his model of joint determination of land use and land rent, Von Thunen assumes manufacture was concentrated in the central city, and agriculture is grown in the surrounding countryside. The main result from this model is a series of concentric rings of cultivation, each with a different product.

Von Thunen's model was later extended and formalized into an urban context by Alonso (1964) and Muth (1969). This model assumes that most commercial activity occurs in the central part of the city, usually called the Central Business District (CBD). All city residents are located around the CBD and commute towards the city center. Households choose housing and non-housing consumption as well as residence location to maximize utility subject to a budget constraint. The basic result of the standard monocentric model of urban land use is that wage gradient will be negative (Muth (1969)).[2] This model is based on the tradeoff between housing costs and commuting costs (Herrin and Kern (1992)). Households are willing to move further away from their job location and accept longer commutes (i.e. higher costs in term of money and time) as long as they get better houses.

Several generalizations and extensions of this model exist. The first obvious one is to allow for firms to locate outside the CBD (Muth (1969), Henderson (1985), White (1988)). While most of the general results of the monocentric model hold when firms locate outside the CBD, in some cases these results fall apart. White

[2] See also Fujita (1989), chapter 1 for a history of the development of urban economic theory, and chapter 2 for a detailed description of the monocentric model of household location.

(1988) argues that positive wage gradient may result because of out-commuting. If workers are capable of finding a job outside the CBD and they live in the CBD, employees will be forced to pay a higher salary to get these people to commute to the outside of the CBD.

A second extension is to allow for workers to be employed locally, this is, outside the CBD (DeSalvo (1977), Turnbull (1992)). Turnbull (1992) shows the Muth's basic results do not hold under local employment. Heckman (1980) extended Muth's and Wheaton's analysis from a one-person household to a two-person household where the connection between husband's and wife's income is made. Once again conditions similar to those from Muth's original model are obtained. Finally, DeSalvo (1985) included a time constraint and treated leisure explicitly. He proves that the results of the basic model still hold under these conditions. Most of these extensions are exhaustively analyzed theoretically, but not much has been done empirically. The main explanation offered is the lack of available data.

Empirical support for the monocentric model includes Eberts (1980), Eberts and Gronberg (1982), Gabriel and Rosenthal (1982), Madden (1985), McMillen and Singel (1991) and Ihlandfelt (1992). Ihandfelt (1992) presents some evidence on the existence, shape, and slope of intraurban wage gradient. He uses the 1980 PUMS data for Philadelphia, Detroit and Boston. He divides each area into 4 parts: central city and inner, middle and outer rings. He uses an indicator called "import ratio", defined as the number of jobs in a particular area divided by the total number of workers in that area. If this ratio is greater than 1, then there are more jobs than workers in the area and it is called a net importer of workers.

But the monocentric model is losing validity. Some cities have several business district centers (granted, that the monocentric model has been extended to include several centers and commuting costs). This has changed land-value configurations. Income and density profiles have also been changing in some cities with some of the higher income household moving to particular inner city areas (Beaudet, 1988).

2.1.2 Models of individual choice

Models of individual choice are typically based on the random utility model. In this type of models, households or individuals select a location among a discrete number of (mutually exhaustible) choices that will maximize their utility subject to an income constraint. In this dissertation, I use the household as the basis for analysis. The utility to a household selecting an alternative is assumed to be a linear function of the characteristics of the household and the attributes of the alternative plus an error term. The probability that a certain household will choose a particular location is given by the probability that the utility of that location to the household is greater than the utility to that household of any other alternative.

When the error term is assumed to be independently and identically distributed as log Weibull distribution, the model is known as a multinomial logit if only individual or household specific characteristics are considered or a conditional logit if the attributes of the characteristics are included in the estimation. The disadvantage of these types of logit models is that they suffer from the "independence of irrelevant alternatives" property. This property states the odds ratio for any pair of choices is the same irrespective of the total number of choices considered. This odds ratio does not change even if the choice set is expanded.

These models can be traced to the work of McFadden (1978). There are numerous applications of both the multinomial and the conditional logit. Recent examples applied to household location include Gabriel and Rosenthal (1989), Waddell (1996) and Sermons and Koppelman (1998). Waddell analyses the interactions within single and dual-worker households between workplace location, residential mobility, housing tenure, and location choice. The basic hypothesis tested is that the presence of a second worker adds constraints on household choices that should lead to a combination of lower mobility rates and longer commutes. Residential mobility is defined as the decision to move, and once this choice is made, households choose a residential location and tenure.

Sermons and Koppelman (1998) use a factor analytic approach to incorporate systematic taste variations into models of residential location choice. They argue that when calibrating models of household selection, planners must select relevant variables from a large set of potentially useful variables. But the problem is that all these variables are very likely to be highly correlated, therefore making estimation not fully efficient. They use factor analysis to select a representative set of variables in a household location model for Portland using 1994 data.

Table 1 in the annex contains a list of the most commonly used regressors in discrete models of household location.

2.2 Land use and transportation interaction

Steiner (1994), Wilson (1998), and, Badoe and Miller (2000) present recent surveys of the literature on the interaction between land use and transportation.

Empirical studies on the interaction between transportation (i.e. travel and automobile demand) and land use usually include a model of the number of cars owned by a household and a model of the demand for miles (VMT). Different measures of land use are typically included as explanatory variables in these models. These equations are interrelated, and usually appear in a nested form. Estimation is done following methods for continuous/discrete models.[3] In this general framework, households first select how many cars to own. Second, conditional on vehicle ownership, the household decides how much to use each car. This framework is commonly used in studies on travel behavior (BCMV, Train 1986; Kockelman, 1997), car ownership (de Jong et al, 2005), and gasoline demand (Berkowitz et al, 1990; Kayser, 2000). Several extensions to this structure exist. For instance, the choice of the mode of transportation or the type of vehicle chosen may be embedded as intermediate steps (Mannering and Winston, 1985; Train, 1986; Kockelman, 1997).

The number of cars a household owns is explained by the household's income and number of workers, the costs of owning an automobile, and the availability of public transit (Train, 1986). Table 2-2 summarizes the main variables used in car ownership models do date.

[3] See Train (1986), chapter 5, for a detailed description of continuous/discrete models.

With respect to VMT, two variables that have been consistently included as explanatory variables in this type of models are income and a measure of operating costs (gas cost per mile, price of gasoline) Population density at the neighborhood level is also commonly included in the previous equations as a measure of land use. Table 2-3 includes a list of variables commonly used in VMT models.

Analysts have observed that people tend to use less private cars as population density increases. Schimek (1996) offers three explanations why this relation may exist: First, potential destinations are contained in a smaller area as density increases. Second, the availability of transit increases as density increases because of economies of scale. And third, auto use becomes more complicated because of congestion or parking problems as density increases.

In studies on the demand for transportation, VMT is usually modeled as a function of demographic characteristics (household size, income, number of workers or drivers in the household, among others), car characteristics (cost per mile or price of gasoline) and land use measures (population density, jobs-housing balance, residential density, accessibility, and dummies for city size). Two land use variables have been consistently used in transportation models to proxy for household location are density (either neighborhood or at the subregional level—i.e. county) and access to transit. The issue of endogeneity of density and access to transit is tackled by Schimek (1996) and Bento et al (2005). While Schimek uses dummy variables to instrument for neighborhood density, Bento et al (2005) create measures of urban form that are truly exogenous to the household. The latter authors use measures at the city level and take advantage of the intercity variation to measure the effects of the variables of interest on travel demand and vehicle ownership.

While many studies support the hypothesis that there exists an inverse relationship between density and private transport, a branch of the literature has shown that there is no statistical evidence of a negative relationship between density and VMT. Support for a strong negative impact of density is provided Newman and Kenworthy (1988, 1989), Dunphy and Fisher (1993), Holtzclaw (1991, 1994), Cervero (1989), and Smith (1984), among others described in the subsequent paragraphs.

Dunphy and Fisher (1996) used the 1991 Federal Highway Administration (FHWA) Statistics and found an inverse relationship between VMT and local population density. They also found evidence of a positive relationship between transit use and density and between income and miles traveled per household. The problems with this study are that, first, they derive their results from simple cross-tabulations without attempting multivariate regressions, and second, they use average density values for large regions.

Handy (1993) studied the relationship between shopping trips and land use, controlling for accessibility. She found that high levels of accessibility were associated with shorter shopping distances. Also, non-work travel was found significantly lower in areas with high levels of accessibility. Trip frequency was not affected by accessibility measures. These results were obtained by finding correlations between every pair of variables in question, and no other factors were controlled for simultaneously.

But at the same time, there is also a branch of literature that finds no statistical evidence of a negative relationship between density and VMT. Levinson and Kumar (1993) study commuting time, speed and distance to work by mode of travel. Even though they find out an inverse relationship density and both speed and distance, density's effect on travel time is ambiguous as speed and distance have offsetting effects on time. They reached the conclusion that density and urban design did not explain transit usage or distance traveled. Kockelman (1997), using the 1990 San Francisco Bay Area Travel Survey, she found that density had a negligible effect on travel behavior once accessibility of an area (based on the gravity model) was accounted for. The measures of land use balance (measured using an entropy index) and accessibility (a measure proportional to the "attactiveness"—i.e. number of opportunities—of the zone and inversely proportional to travel time between zones) proved to be more relevant for explaining travel behavior than other commonly used demographic variables.

Giuliano and Small (1993) argue that altering the job-housing balance will have a very small effect on vehicle use and commute times. They measure the job-housing balance in terms of the number workers per job in the same area or in terms of the required commute time ("excess" commuting). They show that its effect on commute time is statistically significant but very small. They also argue that journey to work plays only a limited role in residential location choice. Some possible explanations for this last result are that (i) most commuting times are relatively short, (ii) job heterogeneity may prevent two worker households to simultaneously locate close to both work places, and (iii) the increasing importance of non-work trips.

Finally, Schimek (1996) found evidence of an inverse relationship between households located in high-density areas and VMT but this relationship is not very large. In other words, increasing density will not reduce VMT very much. He addresses the fact that neighborhood density is a household choice variable by using an instrumental variables approach. He instruments for neighborhood density by using different dummies for city size. He used the 1990 NPTS. Even though he does not explain the method of estimation, it seems that he used OLS to estimate the vehicle ownership model (the dependent variable is a count variable).

Additionally, in a study on travel characteristics across people from different ethnicities, Giuliano (2003) finds that residence within a central city is not associated with less VMT (not only commute VMT).

2.3 Mixed models of travel demand, car ownership and land use

A separate strand of the literature of transportation and land use has taken into consideration the household's location choice. This strand is reviewed in this section. The papers described below present different approximations at analyzing the effect of household location on transportation. As such, they represent the attempts to bring together the literature of residential location and the literature on the interaction between transportation and land use. This is important because even though some papers (Waddell, 1996) acknowledge the importance of distance to work as a determinant of household location, the effect that household location may have on vehicle ownership levels or travel demand has not been explored before in the literature. There is one key difference between this dissertation and the papers described in this section: the methods used to estimate the models in chapters 4 and 5 are different than those described in this section. The difference stems from the fact that I use a continuous variable, distance to work, to model household location, while the papers in this section are use a discrete set of alternatives where the household can locate.

Lerman (1976) estimated a model where households select a joint mobility bundle. Mobility bundles are a combination of housing type, automobile ownership levels, and mode to work choices. A household has a choice of two modes (car and transit), three automobile ownership levels (0, 1, and 2 or more), four housing types (own house and three types of rental types). There are a total of 20 possible options the household can choose from. He estimates a multinomial logit. The variables that affect the choice of mobility bundle are divided into several categories: transportation level of service to work, automobile ownership attributes, locational attributes, housing attributes, spatial opportunities and socioeconomic characteristics.

Sermons and Seredich (2001) model the joint choice of household location and car ownership. Household location is modeled by defining 5 clusters based on San Francisco's traffic analysis zones (TAZ). Cluster analysis was performed on variables such as residential and employment densities, travel time to urban core, median home value, percent land developed, and average rooms per housing unit in order to reduce the number of TAZ to a more manageable number of alternatives. In addition, the authors model travel demand taking into consideration the joint location and car ownership decision. The model developed can be used to predict the potential impacts on household vehicle availability and trip-making of policies that make higher density residential locations more attractive. They make a simulation to show the impact of a change in residential attributes on vehicle availability and trip-making.

The main difference between the work of Sermons and Seredich (2001) and this dissertation is they model household location by using a discrete choice approach while I use the continuous variable distance-to-work. The rationale for choosing a continuous variable over a discrete one to represent household location was discussed in the introduction. This difference in choice variables implies that the method of estimation will be different. In principle, both their model and my model can be used to determine the effects on changes in city characteristics on car ownership and vehicle miles traveled. By construction, the effect of changes in city characteristics on VMT will come through changes in location and car ownership in Sermons and Seredich's model. I allow some of the city characteristics to have both direct and indirect effects on VMT (for example area of MSA and population density). A second difference between my research and Sermons and Seredich (2001) is that I use country level data and they use city level data (San Francisco). Additionally, their VMT data is estimated from a 2-day diary, while the data from NPTS was collected using odometer readings over a specified time period (usually around 4 weeks between readings). Finally, a third difference between my research and theirs is that in Chapter 4 I take into consideration both working and non-working households. This is important because working and non-working households have very different travel patterns.

Anas (1981) uses aggregated data from Chicago to estimate a multinomial logit of joint location and travel mode choice. In it, households simultaneously select a location to live and the mode they will use to get to work in the urban core. Explanatory variables include zone specific attributes such as mean housing price and rent, distance measures of a zone location, average housing attributes of the zone, and travel time and travel cost to the CBD. Rouwendal and Meijer (2001) report stated preference of Dutch workers for combinations of housing, employment, and commuting. Individuals participating in the experiment were given hypothetical groups of housing characteristics and mode choice to choose from. Housing characteristics include type of location, number of rooms, type of dwelling, and mode choice to work. They find that even though households dislike commuting, some are willing to accept longer commutes if the dwelling characteristics are strong enough. Also, the value of commuting time implied by the model is high compared to wage rates.

As a summary, this dissertation follows the spirit of the papers mentioned above. It analyzes the link between location choice and its interaction with car ownership levels and demand for miles but the approach is different. The main difference between Sermons and Seredich and my research is that I use a continuous measure to model household location. This measure allows me to calculate the contribution of commute miles to total miles driven and estimate the direct and indirect effects of changes in city characteristics on VMT, conditional on DTW. Finally, this dissertation is also different from Anas (1983) and Lerman (1976) in the sense that it models household location and transportation as different but related choices. These two papers assume that households will choose among different combinations of location and travel options.

2.4 Tables for Chapter 2

Table 2-1. Household Location Models

Category	Explanatory variable used	Study
1. Travel time to work	Distance to CBD	Anas (1981)
	- Distance to work	- Waddell (1996)
	- Distance to the urban core	- Sermons and Seredich (2001)
	Total time to work	Ben Akiva et al. (1980) Anas (1981), Rowendal and Meijer (2001)
	- Total in-vehicle time	- Lerner (1977)
	- Out-of-vehicle travel time	- Lerner (1977)
	Mode of commute	Anas (1981)
	Cost of travel	Anas (1981)
2. Accessibility	"Generalized shopping price" by transit	Lerner (1977)
	"Generalized shopping price" by car	Lerner (1977)
3. Location attributes	Percent non-white households in location	Lerner (1977) Anas (1981)
	Residential density	Lerner (1977)
	- Location in center/suburb or large/small city	- Rowendal and Meijer (2001)
	School	
	- per pupil school expenditures	- Lerner (1977)
	Square of difference between household income and average annual tract income	Lerner (1977)
	Crime rates	Ben Akiva et al. (1980)
	Proximity of industrial land	Ben Akiva et al. (1980)
	Property taxes	Ben Akiva et al. (1980)
	Rent costs	Anas (1981)
	Percent renter's in area	Anas (1981)
	Characteristics of the dwelling	Rowendal and Meijer (2001)
4. Household characteristics	Income	Ben-Akiva et al. (1980), Anas (1981), Gabriel and Rosenthal (1989), Rowendal and Meijer (2001)
	- Income after expenses (taxes, housing commuting, and car costs)	- Lerner (1977)
	Number cars owned	Anas (1981)
	Life cycle dummies	Gabriel-Rosenthal (1989), Pollakowski and Eduards (1986)
	Marital status	Gabriel-Rosenthal (1989)
	Household head	
	- Sex	- Gabriel-Rosenthal (1989)
	- Education	
	- Race	

Table 2-2. Car ownership models

Category	Explanatory variable used	Study
1. Location attributes	Density	Schimek (1996)
		Zhao and Kockelman (2000)
		Cropper et al. (2002)
		Kockelman (1997)
	-	- Sermons and Seredich (2001)
	- Population centrality	Cropper et al. (2002)
	Job-housing imbalance	Cropper et al. (2002)
	City shape	Cropper et al. (2002)
	Transit Availability	
	- stop <3 blocks away	- Schimek (1996)
	- distance to nearest transit stop	- Cropper et al. (2002)
	Annual rainfall/snowfall	Cropper et al. (2002)
	Accessibility	Kockelman (1997)
	Entropy index	Kockelman (1997)
	General Mix	Kockelman (1997)
2. Household characteristics	- Household size	Schimek (1996)
		Zhao and Kockelman (2000)
		Kockelman (1997)
		Train (1986)
	Income	Schimek (1996)
	- Income per hh member	Train (1986)
		- Zhao and Kockelman (2000)
	- Adjusted for fixed costs of owning a car	Kockelman (1997)
		- Cropper et al. (2002)
	Number of workers in household	Schimek (1996)
		Cropper et al. (2002)
		Train (1986)
	Age of head	Schimek (1996)
	Number of Children	Cropper et al. (2002)
	Race of household head	Cropper et al. (2002)
	Level of Education	Cropper et al. (2002)
3. Transit availability	Road density	Cropper et al. (2002)
	Presence and supply of transit	Cropper et al. (2002)
	Distance to nearest Transit stop	Cropper et al. (2002)
	Annual transit trips per capita in household’s area of residence	Train (1986)
4. Vehicle Attributes	Gas cost per mile	Cropper et al. (2002)

Table 2-3. Travel demand models

Category	Explanatory variable used	Study
1. household characteristics	Life cycle	Dieleman et al (2002)
	Income	Dieleman et al (2002)
		Cropper et al. (2002)
		Kockelman (1997)
		Schimek (1996)
		Train (1986)
	Education	Dieleman et al (2002)
		Cropper et al. (2002)
	Number of cars owned	Dieleman et al (2002)
		Kockelman (1997)
	Age	Levinson and Kumar (1997)
		Schimek (1996)
	Sex	Levinson and Kumar (1997)
	# of workers	Cropper et al. (2002)
		Schimek (1996)
		Train (1986)
	Race	Cropper et al. (2002)
	HH size	Kockelman (1997)
		Schimek (1996)
		Train (1986)
2. location attributes	Dummy for size of city	Dieleman et al (2002)
	- city larger/less than 1M people	- Train (1986)
	Population Density	Cropper et al. (2002)
		Levinson and Kumar (1997)
		Schimek (1996)
	Household in central city of MSA	Schimek (1996)
	Population centrality	Cropper et al. (2002)
	Job-housing imbalance	Cropper et al. (2002)
	City shape	Cropper et al. (2002)
	# of suburban activity centers	Levinson and Kumar (1997)
	Annual rainfall/snowfall	Cropper et al. (2002)
	Population growth rate	Levinson and Kumar (1997)
	Road density	Cropper et al. (2002)
	Urbanized area residential density	Levinson and Kumar (1997)
	Local land use patterns	
	- Accessibility	Kockelman (1997)
	- Mean entropy	Kockelman (1997)
	- General mix	Kockelman (1997)
	US region	Train (1986)
3. Cost of driving	Gas cost per mile	Cropper et al. (2002)
		Train (1986)
	-Price of gasoline	-
4. Transit availability	Presence and supply of transit	Cropper et al. (2002)
	- stop <3 blocks away	- Schimek (1996)
	Distance to nearest Transit stop	Cropper et al. (2002)
	Transit trips per capita in household's area of residence	Train (1986)

3 Data

The data for the dissertation were obtained from four sources: the core body of data comes from the 1995 Nationwide Personal Transportation Survey (NPTS), which was complemented with measures on land use (estimated by BCMV), gasoline prices (from the Department of Energy's Energy Information Administration) and city level data on infrastructure, area, and population obtained from the Census.

3.1 The 1995 NPTS

The NPTS is a survey carried out by the Bureau of Transportation Statistics between May 1995 and July 1996, with the objective of collecting data on travel and transportation patterns in the US. The NPTS serves as the nation's inventory of daily personal travel. National data are collected on daily trips including, among others, purpose of trips, means of transportation, travel time, vehicle occupancy, and vehicle attributes. The 1995 NPTS updates similar information gathered in studies carried out in 1969, 1977, 1983, and 1990.

The NPTS is a stratified sample. The sample was stratified by geography (census region), MSA size, subway /elevated rail presence, and two levels of phone number density. 42,015 completed household interviews were collected in total. Data for the 1995 NPTS was collected in three phases using Computer Assisted Telephone Interview (CATI) technology. First, a household interview takes place. After this interview, the travel diaries and odometer forms are mailed to participating households. Second, a "person interview" is done, and data on the travel day (from travel diaries) and other personal data are collected. Finally, odometer readings are collected through a final phone call.

NPTS data was complemented with Census Tract characteristics for the year 1995 obtained from CLARITAS, Inc. These additional data were imputed from the 1990 census data. These include household descriptors such as the median household income, the median housing unit value, the current population, population density, and the percent of population of different races, among others.

I analyze households that live in the 134 largest metropolitan statistical areas (MSA). I restrict the data to MSAs because several of the explanatory variables are applicable only for urban areas. I further restricted analysis to MSAs that had at least 20 observations. These 134 MSAs constitute 85 percent of the original sample.

Regarding income data, the NPTS solicited income by asking for an estimate of total household income in the past 12 months. Answers were solicited in categories of $5,000 (for medium scale incomes) or $10,000 (for the smallest and largest incomes). Because these are relatively narrow intervals (most are $5,000 increments), I use the middle of each category as the household's income.

When there is at least one worker in the household (see below), the NPTS asks, "What is the one-way distance from (your) home to (your) workplace?" When there is more than one worker, this question is posed for each worker. Answers were given in either miles or blocks; I converted blocks into miles at the rate of 8 blocks per mile. Less than 4 percent of workers reported their distance-to-work in blocks.

There are two situations in which answers to the distance-to-work question are coded as missing: (a) The worker works at or out of the home; or (b) the worker does not have a fixed workplace. Examples of the latter category are a construction worker who travels directly to varying job sites or a cab driver. Conditional on there being at least one worker in the household and on the distance-to-work response not being missing, distance-to-work should be positive. There are, however, a small number of individuals that report having a fixed workplace not at home but who report a distance-to-work of 0. When all individuals in a household have missing distance-to-work or work outside the home but report zero distance-to-work, the household is dropped from my sample.

For households that own or lease at least one car, the survey estimates the miles driven for all cars during a one year period based on odometer readings recorded at varying intervals. See Pickrell and Schimek (1999) for a discussion of this vehicle-based estimate of VMT. A very small number of households (28 in total) report owning at least one car but have zero VMT.

3.2 Other sources of data

I also use data from several other sources. Gasoline price data are state-level data from the Energy Information Administration. Unemployment data are from the Department of Labor.

I use the Gini coefficient of the jobs-housing distribution (GINIJOBS) to examine the balance between jobs and housing across a city. The Gini coefficient was constructed using the procedure described in BCMV, which in turn followed Massey and Denton (1988). I obtained the number of jobs at the zip code level (Zip Code Business Patterns, 1995) and arranged them from smallest to the largest. I then plotted the cumulative percent of jobs (y-axis) against the cumulative percent of population (x-axis; these data were in the NPTS) to form a Lorenz curve. I then defined the variable GINIJOBS as the area between the Lorenz curve and the 45-degree line, expressed as a proportion of the 45 degree line. A higher value means that jobs and houses are more spatially disparate; a lower value means that jobs and houses are more closely matched in space.

Because of my interest in the effects of city characteristics, and to make results comparable to other studies, I also include a measure of city shape. This measure, developed by BCMV, was calculated as follows: First, each city in the data set was circumscribed with an ellipse equal in area to the urbanized area of the city. Second, both the minor and major axis of this ellipse were obtained. Finally, the measure of city shape was defined as the ratio of the minor to the major axis. This measure ranges between 0 and 1, with 1 indicating a circular city. The idea behind this measure is that trip distances should be shorter in circular cities with radial road networks than in long, narrow cities. City shape data are available for 109 MSAs. These measures were obtained from BCMV and use MSAs as defined in 1990. Since the choice of distance to work represents a medium to long term decision, the choice or city characteristics prior to 1995 (the DTW data year) are apt for my analysis.

Data on area and population density for each MSA are from the 1990 U.S. Census, http://www.census.gov/population/censusdata/90den_ma.txt, using MSAs as defined on June 30, 1990.

Table 3-1 presents a summary of the main city level variables. Data on annual VMT and DTW is also presented in this table for comparison purposes. Results are sorted by city area (in square miles).

Table 3-2 contains information on the average distance to work and annual vehicle miles traveled per household. Note for instance, that individuals with an income of less than $20,000 own on average 1.45 cars, live an average of 3,130 (annualized) miles from work and drive approximately 10,540 miles per year. This implies that, on average, their commute miles should correspond to roughly one third of their total miles. This ratio needs to be interpreted with care, as these are sample averages. It does not take into consideration important variables such as mode of commute to work and occupancy rates of vehicles used for commute. Nonetheless, the table below shows that the ratio of the average distance to work to the average vmt by income category varies between 30 and 50 percent. This ratio is higher than expected. Table 3-2 also shows that as income increases, car ownership increases, going from 1.45 cars for the poorest segments of society to 2.41 cars for the richest members.

Table 3-3 shows average HHDTW, VMT and car ownership levels by number of workers in the household. As number of workers in the household increase, car ownership, distance to work, and VMT increase, though at a decreasing rate. The ratio of the average HHDTW and VMT ranges between 0.35 and 0.65

Finally, Table 3-4 shows HHDTW and VMT by car ownership levels. Households that own 1, 2, and 3 or more vehicles drive an average of 10, 20, and 28 thousand miles per year, while they live on average 3,950, 9560, and 12,970 (annualized) miles per year. The ratio of average HHDTW to average VMT varies between 0.39 and 0.47, which is higher than expected.

3.3 Tables for Chapter 3

Table 3-1. Descriptive statistics by MSA

	Metropolitan Area	Area (sq-mi)	Pop. Density (per sq-mi)	Jobs-housing imbalance	Median speed in city (mph)	City shape	Highway density (lanes per sq-mi)	Price of Gas (dollars)	Access to transit	Annual VMT (thousand miles)	Average DTW (thousand miles)
1	Riverside-San Bernardino CA	27.27	0.09	0.41	38.51	0.55	1.09	1.16	0.50	19.30	7.53
2	Phoenix-Mesa AZ	9.20	0.23	0.40	30.00	0.45	0.48	1.21	0.37	17.54	3.88
3	Tucson AZ	9.19	0.07	0.07	30.00	0.80	0.48	1.21	0.30	13.60	2.62
4	Las Vegas NV	7.91	0.09	0.45	28.00	0.73	0.42	1.17	0.58	18.02	3.24
5	Fresno CA	5.96	0.11	0.33	26.77	0.93	1.09	1.16	0.57	20.07	3.10
6	St Louis MO	5.33	0.46	0.33	32.33		1.98	1.09	0.57	17.76	4.71
7	Houston TX	5.32	0.62	0.27	33.17	0.80	1.13	1.16	0.42	21.39	6.21
8	Atlanta GA	5.12	0.55	0.42	32.11	0.26	1.92	0.98	0.44	21.86	5.61
9	Sacramento-Yolo CA	5.09	0.29	0.37	33.00	0.55	1.09	1.16	0.37	20.35	5.61
10	Minneapolis-St Paul MN	5.05	0.49	0.40	31.67	0.84	1.65	1.18	0.64	19.56	4.84
11	Tulsa OK	5.01	0.14	0.38	32.70	0.81	1.64	1.04	0.46	19.93	4.33
12	Kansas City KS	4.99	0.31	0.31	34.50		1.72	1.05	0.37	19.61	5.38
13	Eugene-Springfield OR	4.55	0.06	0.08	24.86		0.87	1.29	0.76	13.95	2.90
14	Dallas TX	4.47	0.57	0.36	30.67	0.52	1.13	1.16	0.58	20.57	5.74
15	Detroit MI	4.47	0.98	0.40	32.81	0.80	2.07	1.06	0.47	21.84	5.56
16	Portland-Salem OR	4.37	0.34	0.37	25.71	0.87	0.90	1.29	0.62	17.46	4.13
17	Oklahoma City OK	4.25	0.23	0.41	33.00	0.81	1.64	1.04	0.40	19.64	4.48
18	Seattle-Bellevue-Everett WA	4.22	0.47	0.45	28.66	0.35	1.20	1.26	0.61	18.52	5.21
19	San Diego CA	4.20	0.59	0.41	33.62	0.36	1.09	1.16	0.62	18.23	5.05
20	Nashville TN	4.07	0.24	0.30	33.00	0.83	2.08	1.13	0.19	21.00	5.05
21	Los Angeles-Long Beach CA	4.06	2.18	0.42	28.00	0.55	1.09	1.16	0.68	18.51	5.58
22	Birmingham AL	3.98	0.23	0.25	34.00	0.62	1.84	1.12	0.58	23.95	5.54

Table 3-1. Descriptive statistics by MSA

	Metropolitan Area	Area (sq-mi)	Pop. Density (per sq-mi)	Jobs-housing imbalance	Median speed in city (mph)	City shape	Highway density (lanes per sq-mi)	Price of Gas (dollars)	Access to transit	Annual VMT (thousand miles)	Average DTW (thousand miles)
23	Washington DC	3.97	0.99	0.43	26.40	0.82	4.54	1.24	0.61	18.68	5.92
24	Denver CO	3.76	0.43	0.36	30.00	0.82	0.81	1.22	0.61	19.37	5.29
25	Columbus OH	3.58	0.38	0.35	32.00	0.80	2.80	1.16	0.47	17.88	5.74
26	Philadelphia PA	3.52	1.38	0.43	25.71	0.85	3.18	1.20	0.52	16.66	4.76
27	Greensboro-Winston Salem-High Point NC	3.45	0.27	0.30	37.33		1.99	1.11	0.40	17.93	4.89
28	Pittsburgh PA	3.40	0.60	0.30	28.00	0.61	2.65	1.21	0.53	15.57	3.88
29	Charlotte-Gastonia-Rock Hill NC	3.38	0.34	0.23	29.29	0.79	2.01	1.10	0.36	19.73	3.97
30	Albany-Schenectady-Troy NY	3.25	0.27	0.39	30.00	0.62	2.38	1.20	0.52	18.72	4.17
31	Indianapolis IN	3.07	0.41	0.27	30.00	0.76	2.59	1.07	0.57	17.99	4.13
32	Wichita KS	2.97	0.16	0.25	34.50	0.96	1.63	1.07	0.60	19.74	3.94
33	Richmond-Petersburg VA	2.94	0.29	0.34	35.67	0.82	1.75	1.14	0.28	18.78	5.78
34	Rochester NY	2.93	0.34	0.38	30.00	0.77	2.38	1.20	0.46	17.81	4.18
35	Cleveland-Lorain-Elyria OH	2.91	0.95	0.41	27.00	0.56	2.80	1.16	0.60	18.82	4.45
36	Little Rock-North Little Rock AR	2.91	0.18	0.35	35.50	0.49	1.48	1.09	0.35	18.87	4.30
37	Johnson City-Kingsport-Bristol TN	2.87	0.15	0.20	34.71		2.01	1.13	0.00	22.13	7.40
38	Scranton-Wilkes-Barre-Hazelton PA	2.84	0.26	0.58	30.00	0.30	2.65	1.21	0.63	21.71	4.94
39	Mobile AL	2.83	0.17	0.23	36.25	0.05	1.84	1.12	0.20	27.29	6.41
40	Austin-San Marcos TX	2.79	0.28	0.23	30.00	0.71	1.13	1.16	0.65	18.21	5.38
41	Knoxville TN	2.77	0.22	0.28	34.07	0.67	2.08	1.13	0.33	18.80	4.58
42	Santa Barbara-Santa Maria-Lompoc CA	2.74	0.14	0.47	27.27		1.09	1.16	0.75	14.74	2.41
43	Jacksonville FL	2.64	0.34	0.09	25.00	0.76	2.11	1.07	0.53	16.83	3.95
44	Utica-Rome NY	2.62	0.12	0.36	30.50	0.34	2.38	1.20	0.58	18.84	3.95
45	Baltimore MD	2.61	0.91	0.34	29.20	0.75	3.04	1.31	0.51	22.11	6.09
46	Charleston-North Charleston SC	2.59	0.20	0.38	32.20	0.44	2.14	1.05	0.53	19.55	3.75
47	Tampa-St Petersburg-clearwater FL	2.55	0.81	0.36	30.00	0.99	2.11	1.07	0.44	18.75	4.06
48	Orlando FL	2.54	0.42	0.26	28.00	0.74	2.11	1.07	0.37	19.85	4.59
49	San Antonio TX	2.52	0.52	0.18	30.00	0.80	1.13	1.16	0.76	20.74	4.80

Table 3-1. Descriptive statistics by MSA

	Metropolitan Area	Area (sq-mi)	Pop. Density (per sq-mi)	Jobs-housing imbalance	Median speed in city (mph)	City shape	Highway density (lanes per sq-mi)	Price of Gas (dollars)	Access to transit	Annual VMT (thousand miles)	Average DTW (thousand miles)
50	Fort Worth-Arlington TX	2.50	0.53	0.43	35.14		1.13	1.16	0.50	23.92	5.40
51	Syracuse NY	2.39	0.28	0.36	32.00	0.65	2.38	1.20	0.58	18.77	4.70
52	Jackson MS	2.36	0.17	0.22	37.00	0.77	1.56	1.11	0.67	22.13	5.33
53	New Orleans LA	2.31	0.54	0.68	24.43	0.68	1.38	1.17	0.62	20.15	3.89
54	Memphis TN	2.30	0.43	0.34	31.25	0.84	1.99	1.13	0.64	21.93	5.02
55	Louisville KY	2.27	0.42	0.35	30.00	0.87	2.04	1.13	0.66	16.06	4.10
56	Huntington-Ashland WV	2.16	0.14	0.33	30.00		2.01	1.19	0.63	16.72	2.57
57	Colorado Springs CO	2.13	0.19	0.25	28.80	0.60	0.81	1.22	0.58	15.36	4.48
58	Cincinnatti OH	2.13	0.68	0.41	26.67	0.71	2.62	1.15	0.49	18.25	4.70
59	Greenville-Spartanburg SC	2.10	0.31	0.32	31.20	0.93	2.14	1.05	0.38	15.98	2.97
60	Chattanooga TN	2.09	0.21	0.22	32.20	0.69	2.01	1.07	0.00	22.30	3.21
61	West Palm Beach-Boca Raton FL	2.03	0.42	0.27	34.14	0.28	2.11	1.07	0.36	16.69	5.39
62	Raleigh-Durham-chapel Hill NC	2.02	0.36	0.25	36.00	0.90	1.99	1.11	0.39	22.41	6.03
63	Harrisburg-Lebanon-Carlisle PA	1.99	0.30	0.58	32.00	0.52	2.65	1.21	0.56	20.66	5.15
64	Augusta-Aiken GA	1.95	0.20	0.46	31.00	0.45	1.99	1.00	0.12	22.68	4.27
65	Miami FL	1.94	1.00	0.17	25.20	0.04	2.11	1.07	0.62	19.48	4.64
66	Salem OR	1.93	0.14	0.53	28.00		0.87	1.29	0.63	18.89	4.91
67	Omaha NE	1.92	0.32	0.17	28.25	0.75	1.33	1.14	0.62	19.06	2.77
68	Chicago IL	1.88	3.22	0.46	26.00	0.48	2.47	1.19	0.72	17.78	5.39
69	Lakeland-Winter Haven FL	1.87	0.22	0.14	32.70		2.11	1.07	0.50	10.95	3.33
70	Ventura CA	1.85	0.36	0.26	41.65	0.24	1.09	1.16	0.36	20.41	8.05
71	Peoria-Pekin IL	1.80	0.19	3.06	34.00	0.76	2.47	1.19	0.54	20.61	4.77
72	Saginaw-Bay City-Midland MI	1.77	0.23	0.26	31.75		2.07	1.06	0.55	23.04	3.16
73	Spokane WA	1.76	0.20	0.26	24.86	0.69	1.20	1.26	0.70	15.47	2.99
74	Boston MA	1.76	1.63	0.41	30.60	0.82	3.89	1.28	0.49	20.98	5.89
75	Des Moines IA	1.73	0.23	0.27	28.00	0.91	2.02	1.09	0.67	15.36	2.83
76	Davenport-Moline-Rock Island IA	1.71	0.21	0.19	24.00	0.69	2.35	1.16	0.80	19.46	2.87

Table 3-1. Descriptive statistics by MSA

	Metropolitan Area	Area (sq-mi)	Pop. Density (per sq-mi)	Jobs-housing imbalance	Median speed in city (mph)	City shape	Highway density (lanes per sq-mi)	Price of Gas (dollars)	Access to transit	Annual VMT (thousand miles)	Average DTW (thousand miles)
77	Lansing-East Lansing MI	1.71	0.25	0.10	30.00	0.63	2.07	1.06	0.74	22.88	4.48
78	Glens Falls NY	1.71	0.07	0.39	33.50		2.38	1.20	0.47	19.17	5.09
79	Norfolk-Virginia Beach-Newport News VA	1.69	0.83	0.02	31.67	0.51	1.75	1.14	0.51	20.42	4.60
80	Dayton-Springfield OH	1.68	0.57	0.25	32.00	0.78	2.80	1.16	0.51	20.31	4.34
81	Pensacola FL	1.68	0.21	0.45	21.73	0.77	2.11	1.07	0.56	17.56	4.89
82	Tacoma WA	1.68	0.35	0.31	30.00	0.79	1.20	1.26	0.64	18.34	4.72
83	Salt Lake City-Ogden UT	1.62	0.66	0.16	29.10	0.46	0.50	1.16	0.73	18.71	5.08
84	Baton Rouge LA	1.59	0.33	0.42	31.20	0.60	1.38	1.17	0.42	19.74	4.86
85	Vallejo-Fairfield-NAPA CA	1.58	0.29	2.07	31.43		1.09	1.16	0.65	16.42	5.65
86	Santa Rosa CA	1.58	0.25	0.24	31.00	0.59	1.09	1.16	0.50	23.44	4.09
87	Modesto CA	1.49	0.25	0.13	31.00	0.52	1.09	1.16	0.47	23.43	5.82
88	Allentown-Bethlehem-Easton PA	1.46	0.47	0.24	30.00	0.37	2.65	1.21	0.46	19.34	4.59
89	Milwaukee-Waukesha WI	1.46	0.98	0.45	32.00	0.59	2.05	1.20	0.82	18.60	5.00
90	Oakland CA	1.46	1.43	0.50	28.71		1.09	1.16	0.71	17.71	5.88
91	Columbia SC	1.46	0.31	1.68	31.50	0.71	2.14	1.05	0.40	18.67	3.94
92	Grand Rapids-Muskegon-Holland MI	1.42	0.48	0.31	33.00	0.69	2.07	1.06	0.60	19.28	3.71
93	Stockton-Lodi CA	1.40	0.34	0.23	40.00	0.67	1.09	1.16	0.46	19.53	5.44
94	Appleton-Oshkosh-Neenah WI	1.40	0.23	0.34	33.00		2.05	1.20	0.82	18.97	3.99
95	Toledo OH	1.36	0.45	0.37	28.48	0.81	2.80	1.16	0.44	15.86	4.38
96	Fort Wayne IN	1.36	0.27	0.16	36.02	0.69	2.59	1.07	0.64	20.42	6.06
97	San Jose CA	1.29	1.16	1.15	26.40	0.46	1.09	1.16	0.61	18.36	4.12
98	Binghampton NY	1.23	0.22	0.27	30.00		2.38	1.20	0.54	17.24	3.55
99	Newark NJ	1.22	1.50	0.49	26.10		4.81	1.16	0.57	17.12	4.63
100	Fort Lauderdale FL	1.21	1.04	0.18	27.07	0.62	2.11	1.07	0.58	17.07	4.29
101	Madison WI	1.20	0.31	0.28	27.50	0.86	2.05	1.20	0.82	20.18	3.16
102	Nassau-Suffolk NY	1.20	2.18	0.46	27.40		2.38	1.20	0.49	18.90	6.43
103	Hickory-Morgantown-Lenoir NC	1.17	0.19	0.29	31.00		1.99	1.11	0.25	13.73	2.29

Table 3-1. Descriptive statistics by MSA

	Metropolitan Area	Area (sq-mi)	Pop. Density (per sq-mi)	Jobs-housing imbalance	Median speed in city (mph)	City shape	Highway density (lanes per sq-mi)	Price of Gas (dollars)	Access to transit	Annual VMT (thousand miles)	Average DTW (thousand miles)
104	Albuquerque NM	1.17	0.41	0.24	28.36	0.73	0.51	1.19	0.59	19.07	3.24
105	Wilmington-Newark DE	1.11	0.52	0.26	27.25	0.44	2.90	1.27	0.48	17.43	4.47
106	Monmouth-Ocean NJ	1.11	0.89	0.43	33.60		4.81	1.16	0.37	22.67	6.30
107	Daytona Beach FL	1.11	0.34	0.47	39.00	0.24	2.11	1.07	0.41	22.04	6.23
108	Hartford CT	1.07	0.71	0.26	30.00	0.56	4.23	1.39	0.34	18.81	4.00
109	Jamestown NY	1.06	0.13	0.31	28.35		2.38	1.20	0.59	17.70	3.87
110	Middlesex-Somerset-Hunterdon NJ	1.05	0.98	0.19	30.00		4.81	1.16	0.32	17.42	6.69
111	Buffalo-Niagara Falls NY	1.04	0.93	0.32	28.00	0.54	2.38	1.20	0.51	15.44	3.51
112	Youngstown-Warren OH	1.03	0.48	0.36	30.00	0.44	2.80	1.16	0.20	15.82	3.37
113	San Francisco CA	1.02	1.58	0.25	26.33	0.74	1.09	1.16	0.72	16.35	5.35
114	El Paso YX	1.01	0.58	0.16	36.50	0.45	1.13	1.16	0.79	14.97	5.26
115	Canton-Masillon OH	0.97	0.41	0.34	34.57	0.67	2.80	1.16	0.33	19.93	5.45
116	Lancaster PA	0.95	0.45	0.22	22.00		2.65	1.21	0.35	13.62	3.49
117	Gary IN	0.92	0.66	0.31	32.64		2.59	1.07	0.67	19.42	5.96
118	Akron OH	0.91	0.73	0.19	23.00	0.70	2.80	1.16	0.63	18.48	3.19
119	Orange County NY	0.82	0.38	0.18	30.00		1.09	1.16	0.57	20.26	6.95
120	Dutchess County NY	0.80	0.32	0.41	33.82		2.38	1.20	0.29	21.06	6.61
121	Rockford IL	0.80	0.36	0.17	30.00	0.71	2.47	1.19	0.50	18.46	3.35
122	Ann Arbor MI	0.71	0.40	0.26	36.00		2.07	1.06	0.83	24.14	6.60
123	Asheville NC	0.66	0.27	0.06	27.75		1.99	1.11	0.63	12.00	2.56
124	Flint MI	0.64	0.67	0.21	35.25	0.87	2.07	1.06	0.73	19.76	5.38
125	Providence-Fall River-Warwick RI	0.61	1.07	0.30	36.00	0.55	5.63	1.31	0.38	19.34	5.99
126	Honolulu HI	0.60	1.39	0.58	24.50		0.64	1.44	0.82	14.53	3.80
127	Springfield MA	0.59	0.89	0.57	30.00	0.55	3.92	1.28	0.45	18.94	4.42
128	Sarasota-Brandenton FL	0.57	0.49	0.16	24.00	0.22	2.11	1.07	0.42	13.29	1.71
129	Kalamazoo-Battle Creek MI	0.56	0.40	0.24	35.00		2.07	1.06	0.45	21.19	4.82
130	New Haven-Meriden CT	0.43	1.23	0.30	30.33	0.51	4.23	1.39	0.56	15.92	4.84

Table 3-1. Descriptive statistics by MSA

	Metropolitan Area	Area (sq-mi)	Pop. Density (per sq-mi)	Jobs-housing imbalance	Median speed in city (mph)	City shape	Highway density (lanes per sq-mi)	Price of Gas (dollars)	Access to transit	Annual VMT (thousand miles)	Average DTW (thousand miles)
131	Bergen-Passaic NJ	0.42	3.05	0.37	24.76		4.81	1.16	0.56	18.02	5.61
132	Elmira NY	0.41	0.23	0.43	28.50		2.38	1.20	0.54	15.58	3.02
133	Bremerton WA	0.40	0.48	0.01	31.00		1.20	1.26	0.54	15.54	4.82
134	Pittsfield MA	0.23	0.35	0.42	28.00		3.92	1.28	0.47	16.46	3.74

Table 3-2. Distance to work and VMT by income category

Income (dollars)	Household distance to work (thousand miles)	VMT (thousand miles)	Number of cars	Ratio of HHDTW to VMT (columns 3 and 4)
$0 to $20k	3.13	10.55	1.45	0.30
$20k to $30k	5.57	14.53	1.70	0.38
$30k to $45k	8.20	18.31	1.95	0.45
$45k to $60k	10.85	21.96	2.17	0.49
$60k to $80k	12.76	23.90	2.28	0.53
More than $80k	13.39	25.72	2.41	0.52
Total	8.67	19.11	2.00	0.45

Table 3-3. Distance to work and VMT by number of workers

Number of workers in the household	Household distance to work (thousand miles)	VMT (thousand miles)	Number of cars	Ratio of HHDTW to VMT (columns 2 and 3)
No workers	0.00	10.34	1.48	0
1 worker	5.67	16.16	1.68	0.35
2 workers	12.33	22.58	2.19	0.54
3 workers	16.07	27.42	2.93	0.58
4 workers	21.44	32.54	3.53	0.65
Total	8.67	19.11	2.00	0.45

Table 3-4. HHDTW and VMT by car ownership

Number of cars owned	Household distance to work (thousand miles)	VMT (thousand miles)	Ratio of HHDTW to VMT (columns 2 and 3)
1	3.95	10.13	0.39
2	9.56	20.51	0.47
3 or more	12.97	27.97	0.46
Total	8.67	19.11	0.45

4 The case of Distance to Work

4.1 Introduction

Economists have long been concerned with the externalities generated by automobiles. The externalities are exceptionally varied (see Parry et al, 2007 for a detailed description); they include when and where a car is driven (*i.e.*, congestion); air, noise, and indirect water pollution; and greenhouse gas emissions. There are non-environmental externalities such as accidents and road depreciation, and environmental externalities from the road network itself, such as habitat fragmentation. There are frequently externalities from parking, since parking in many cases is an open access resource, often at capacity or with inefficient search costs. The wide variety and presumed size of these externalities make automobiles prime candidates for regulation and for analysis.

Many of these externalities are closely bound up with the number of miles being driven, and economists have therefore been much interested in the behavior of what is known as *vehicle miles traveled* (VMT). VMT have also been fertile ground for study because household-level VMT are relatively easily observed, more so than most of the specific externality-causing activities.

Much of the research on VMT has focused on the effect of car ownership on VMT (Bento *et al*., hereafter BCMV, 2005; Mannering and Winston, 1985; Train, 1986). In this paper, I focus instead on the effect of the household's location (within a city) on its VMT. In particular, I examine households' *distance-to-work* (DTW).

The key motivation for my focus is that household location provides a much more substantial constraint on VMT than does cars-owned. Individuals can change the number and types of cars they own in as little as a day, with a modest investment in time. By contrast, a change in distance-to-work involves much more time and potentially substantial transaction costs; the non-monetary psychic costs of changing jobs or neighborhoods are also large. In sum, location is much more costly to change than car ownership in both monetary and non-monetary terms.

Thus, the short-to-medium-term constraints imposed by household location are likely to be more important economically and for policy than the automobile stock. Accounting for household location should then give a much clearer picture of factors that influence VMT in all but the very-long-run. This is the main purpose of this research.

Such an approach requires, of course, that I treat household location as endogenous. City- and household-level attributes that may lead households to live close or far from their work may also lead them to drive few or many miles for non-commuting purposes, and this effect must be accounted for when measuring the behavior of VMT conditional on DTW. I adopt several strategies for identifying the household location (DTW) effect.

I chose a continuous measure of household location over a discrete measure for several reasons. First, my sample is representative at the national level, not the local level. I feel that the level of detail needed to characterize discrete choice models works better at the local. This argument is supported in the literature, as all discrete choice models are estimated for a particular city. Second, estimating a continuous/discrete model following the work of Train (1986) would require estimating a VMT equation for each household location available. In order to make this type of model tractable, one would have to aggregate the date into a few choice possibilities such as urban core, urban, and suburban areas. Since aggregation of the location choices would be done somewhat arbitrarily, estimation of the effects of the choice variables in the discrete model (for instance, population density) is not very easily carried over to the VMT model. Third, interpretation of the effect of DTW on VMT is very straightforward when using a continuous variable for household location, particularly from a policymaker point of view—this relationship allows me to calculate the contribution of commute miles to total miles driven.

Despite the clear conceptual connection between distance-to-work and vehicle miles traveled, and the constraining nature of household location, little is known about their joint behavior. To the extent that the VMT literature considers a role for household location, it has focused on variables such as the neighborhood density (Boarnet and Crane, 2001; Chatman, 2002; Sermons and Seredich, 2001) or access to public transportation (Kayser, 2000), with none of these being treated as an endogenous choice. As discussed in section 1.1, Sermons and Seredich (2001) is one of the few papers addressing discrete household location and VMT. The authors use data from San Francisco to jointly estimate a model household location (they divide San Francisco into 5 clusters); a car ownership model (they include the predicted utility of residing in each of the 5 clusters in the estimation); and two equations for vehicle use, defined as number of trips and VMT. In both these last equations they include the predicted number of vehicles the household owns—this is an approach different than that outlined in Train (1986).

As explained above, I believe that DTW is a more relevant and interesting measure of household location than those based on discrete choice or using household density. The literature to explain DTW has focused on neighborhood and household characteristics (*e.g.*, Khattak and Amerlynck, 1999) rather than city-level attributes. The one exception is Levinson and Kumar (1997), who argued that DTW is higher in larger population cities, although without rigorous statistical analysis.

The VMT literature is voluminous, but it has only recently begun to look at the role of city attributes; a presumed large role for these attributes is one of the motivations behind the Smart Growth movement. BCMV is an important recent contribution to this literature. I discuss this paper's contributions below. Our paper builds on this line of inquiry with attention to how city "layout" affects both DTW and VMT conditional on DTW. Note that different city characteristics may be important for DTW and VMT decisions. I discuss and estimate the separate roles for these characteristics.

I reach two broad conclusions. First, those city characteristics that might be expected to affect commutes (other than the city's physical size) have remarkably little effect on households' distance-to-work.[4] Variables like city shape and area, commuting speed, or the joint jobs-housing distribution have little apparent effect on city-average distance-to-work.

Second, I conclude that distance-to-work provides an important effect on overall household vehicle miles traveled. A one percent increase in distance-to-work leads to a 0.18 percent increase in overall VMT. This elasticity is comparable in

[4]Following the literature, I analyze metropolitan statistical areas (MSAs) but refer to them as cities.

magnitude both to the income elasticity of VMT and the gas price elasticity of BMT. This effect is easier understood in terms of marginal effects rather than as an elasticity—in level terms, a one mile increase in one-way distance-to-work for one worker, which translates into roughly 480 additional commuting miles per year, leads to an annual increase of about 158 vehicle miles.

There are two other results that I find noteworthy and that I feel have received insufficient attention from the literature. First, I find that conditional on distance-to-work, people do not drive (much) more in physically larger cities. This result may not be surprising, since non-commuting "chores" can mostly be done locally, regardless of a city's size, but the size and nature of this conclusion has not been estimated to my knowledge. Previous research has either not examined the city area effect (despite, I feel, its seemingly obvious role) or, in the few cases where it has been included, has not emphasized it (BCMV). One implication of my finding is that household migration – mostly from physically small to large cities – will likely have small effects on nationwide VMT. This effect has not been much remarked on.

I also find that non-working households have a considerably smaller VMT-gas-price elasticity. Previous literature has not focused on the work decision, despite the fact that non-working households drive approximately 10,000 miles less per year than working households and constitute roughly 12 percent of the population (based on the NPTS). I estimate separate VMT equations for workers and non-workers (i.e., no workers in the household.) This distinction also entails my recognizing that the work decision is endogenous. This approach is different than that found in the literature, where VMT models typically include the number of workers as a right hand side regressor.

4.2 Data

As described in Chapter 3, I use the 1995 Nationwide Personal Transportation Survey sponsored by the U.S. Department of Transportation. The survey provides data on the amount and nature of personal travel in the U.S., by all modes. The unit of analysis is the household.

I analyze households that live in the 134 largest metropolitan statistical areas (MSA). I restrict myself to MSAs because several of my explanatory variables are applicable only for urban areas. I further restricted analysis to MSAs that had at least 20 observations. These 134 MSAs constitute 85 percent of the original sample.

I further restrict my sample to households with income data. This restriction is necessary because of the importance of income as an explanatory variable. This restriction loses an additional 11.7 percent of the original sample.

I also restrict attention to households with at least one car. This restriction loses just an additional 4.5 percent of households. The reason for my restriction is that I work with a reduced form model in which car ownership in endogenous. The required reduced form demand for VMT would be much more complex if it had to apply to both zero- and nonzero-car households.

Finally, I eliminate from the final sample 1,735 observations with missing VMT data.

Table 4-1 shows my final sample. The largest chunks of attrition are due to missing VMT or income data. This attrition is both severe and unlikely to be random.

This fact does not appear to have been much remarked on in the VMT literature and I similarly do not deal with it here. Future research on the consequences of missing income and VMT data is clearly warranted.

4.3 Econometric Model

4.3.1 Dependent variables

I examine three choice variables: whether anyone in the household works, distance-to-work, and household vehicle miles traveled.

WORK is a dummy variable with WORK = 1 if any adult in the household works, either full or part-time, and 0 otherwise. This variable is defined for me by the NPTS. The NPTS asks whether the individual interviewed or any other of the household members work for pay or for profit at the time of the interview. There may be no workers in a household if all adults are either unemployed or retired. Distance-to-work is recorded only for households for which WORK = 1.

For distance-to-work, I transform the NPTS's measure in two ways. First, I must take into account multiple workers in a household. I construct the average distance-to-work over all workers in the household. An alternative would be to use the household sum, but this variable is not right for me because it is unduly influenced by the household's labor force participation decisions. The average distance-to-work over all workers in the household reflects the household's decision about where to live to accommodate all of its workers and is unaffected by their number. On the other hand, the sum of distances to work is strongly affected by the number of workers. I use average distance to focus on household location rather than labor force participation.

The second transformation is scale. I want to make my distance-to-work variable commensurate with vehicle miles traveled, which are measured on a yearly basis. Therefore I multiply the reported distance-to-work by 480, which is 2 trips per day for 240 work days per year. This operation merely changes the reported one-way distance-to-work to a convenient scale and is not meant to reflect the true number of workdays.[5] DTW is the household-average yearly distance-to-work.

In my analysis of VMT I account for the number of workers, for obvious reasons. I use the variable HHDTW to represent the sum of the yearly distance-to-work over all working members in the household.[6] HHDTW is more comparable to VMT than DTW, and it allows me to calculate the contribution of commute miles to total miles. Note that the coefficient relating HHDTW to VMT (see equation (4) below) has a straightforward interpretation—if this coefficient equals 1, then an increase in distance-to-work is directly translated into annual vehicle miles traveled. I expect this coefficient to be less than one because of task sharing.

For households that own or lease at least one car, the NPTS estimates total miles driven during a one year period. I sum these miles over all cars in the household to construct my dependent variable, the household's vehicle miles traveled, VMT. It is also common to find in the literature VMT models where the left hand side variable is VMT per vehicle.

---

[5] I attempted to construct a household specific count of workdays, but dropped this approach due to missing occupation data and occupations whose workdays were difficult to assess.

6 Note that HHDTW = DTW x (Number of workers)

4.3.2 Model and Discussion

Our general model is shown in equations (1)-(4). This is an endogenous switching regression in which one of the branches is a system of equations.

$$\text{WORK} = 1 \text{ if } X_1\alpha + u > 0\text{; else WORK} = 0 \tag{1}$$

If WORK = 1:

$$\ln(DTW) = X_2\beta + \varepsilon \tag{2}$$

$$VMT = X_3\gamma + \delta \cdot HHDTW + \nu \tag{3}$$

If WORK = 0:

$$VMT = X_4\theta + \omega \tag{4}$$

where X_i is a vector of exogenous variables. In the notation below, I separate X_i into two components, Z (city-specific variables) and X (household-level variables).

Each of these equations has specific considerations that led me to adopt a particular form.

Distance-to-work. Because distance-to-work has not been studied much, I devote some attention to the specification of this equation and the distribution of the error term.

I adopt a conditional log-normal specification, which is both tractable and consistent with the data. Although it would be desirable to derive the appropriate distribution of DTW from more fundamental assumptions about each city's size, shape, and job-housing distribution, such constructions are intractable. The implied distribution of DTW would also be dependent on assumptions about the distribution of household tastes. Suppose that jobs and households are independently and uniformly distributed over a square with side K. The distance between a randomly chosen household and job is then a random variable, albeit without a closed-form solution for the distribution. It can be shown, however, that under these assumptions the mean and variance will be increasing and linear in K and K^2, respectively, a result I believe is instructive. The assumption of independence would, in this context, be an assumption about tastes; for example, jobs and houses could be uniformly distributed but DTW could still range from being everywhere zero to being everywhere large, depending on households' choices to live near or far from their workplace.

Let DTW_{ij} represent the yearly distance-to-work for household i in city j. Then household i's location within city j is given by:

$$\ln(DTW_{ij}) = Z_j\phi + X_i\beta + \varepsilon_i$$

where Z_j is a vector of city characteristics and X_i a vector of household characteristics.

Vehicle Miles Traveled. The specification for the VMT equations is based on the need for both tractability and economically useful parameters. Unlike previous studies, I do not model number of cars owned. Instead, I implicitly use a reduced form in which number of cars owned is endogenous.

I estimate separate VMT equations for households based on WORK ((3) and (4)) because it seems likely that travel patterns will differ greatly based on whether someone in the household works.

Error structure. Equations (2) and (3) form the heart of my analysis. An important element of both equations, however, is the unobservable "taste for driving." Thus, I adopt an error structure that allows correlation between ε and ν. Any attempt to measure the effect of household location on VMT must take this taste into account.

I also allow correlation between the WORK and VMT errors, denoted σ_{WV}. Such correlation is likely due to unobserved personal characteristics such as health; a better health status leads households to be both more likely to work and more likely to undertake activities outside of the house. I do not expect health status (or other propensity-to-work variables) to affect the distance-to-work.

To accommodate the possibility that DTW is more dispersed in larger cities, I assume within-city heteroskedasticity of the form:

$$\Omega_\varepsilon = E(\varepsilon_{ij}\varepsilon_{kj}) = AREA_j^{\,2}\sigma_\varepsilon^2 \tag{5}$$

I therefore impose the following error structure:

$$\Omega = E([u,\varepsilon,\nu,\omega][u,\varepsilon,\nu,\omega]') = \begin{bmatrix} 1 & 0 & \sigma_{\mu\nu} & \sigma_{\mu\omega} \\ & \Omega_\varepsilon & \sigma_{\varepsilon\nu} & 0 \\ & & \sigma_\nu^2 & 0 \\ & & & \sigma_\omega^2 \end{bmatrix} \tag{6}$$

<u>Estimation</u>. I estimated the model using a combination of FIML and LIML, but as a whole is estimated as LIML. The model is a variation of a sample selection model where one of the branches is distributed bivariate normal. The estimation procedure consists of 3 steps:

Step 1: Run probit regression for the household's labor force participation, yielding estimates $\hat{\alpha}$. Construct the inverse mills ratio, $\lambda = \phi(X\hat{\alpha})/\Phi(X\hat{\alpha})$.

Step 2: Estimate equations (2) and (3) (i.e. when WORK=1) under the assumption that the error terms in these equations are distributed bivariate normal. The inverse mills ratio estimated in step 1 is included as a regressor in the VMT equation. Estimation of these two equations is then done using Full Information Maximum Likelihood. I assume heteroskedasticity as described in section 4.3.

Step 3: Estimate equation (4). The inverse mills ratio from step 1 is included in this equation.

Identification. There are two conditions that should be met when dealing with identification in a system of equations: the rank condition and the order condition. The order condition with exclusion restrictions is a necessary but not sufficient condition for identification. The order condition states that "the number of excluded exogenous variables from the equation must be at least as large as the number of included right hand side endogenous variables in the equation". [7] Note that the order condition is met by all equations in the system. It is clear that the order condition holds for equations (1), (2), and (4), as there are no right hand side endogenous variables. The only equation where I have an endogenous variable in the right hand side is equation (3), therefore at least one exogenous variable that is not included in equation (3) is needed for it to be identified. Since the variables representing the"Job-housing Gini Coefficient", the "Median speed in city", and the "relative income" are not included in equation (3), I can conclude that this equation meets the order condition (the equation is overidentified).

With regards to the rank condition, once again the equation of interest is equation (3) as it is the one with the endogenous variable in the right hand side. Temporarily rewrite, for ease of exposition, equations (2) and (3) as follows:

$$\ln(DTW) = \beta_{21}\, GINIJOBS + \beta_{22}\, MEDSPEED + \beta_{23}\, HHSIZE + X^{*} \beta_{2i} + \varepsilon \qquad (2')$$

$$VMT = \delta\, HHDTW + \gamma_{31}\, GASPRICE + \gamma_{32}\, HWYDENS + \gamma_{33}\, NUMDRVR + \gamma_{34}\, NUMADLT + \gamma_{35}\, MILLSRATIO + X^{*} \gamma_{3i} + v \qquad (3')$$

where β_{2i} and γ_{3i} are vectors of dimension (1 x 8). The exogenous variables represented by X^{*} are the same in both equations. Note that the exogenous variables explicitly included in equation (2') are those not included in equation (3') and vice

[7] Taken from Wooldridge (2002), page 215, Theorem 9.1.

versa. The system contains two endogenous variables: VMT and DTW. Define the vectors of coefficients for these two equations as:

$\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23}, \gamma_{21}, \gamma_{22}, \gamma_{23}, \gamma_{24}, \gamma_{25}, \boldsymbol{\beta}_{2i})$

$\beta_3 = (\beta_{31}, \beta_{32}, \beta_{33}, \gamma_{31}, \gamma_{32}, \gamma_{33}, \gamma_{34}, \gamma_{35}, \boldsymbol{\gamma}_{3i})$

where $\beta_{31}= \beta_{32}= \beta_{33}= \gamma_{21}= \gamma_{22}= \gamma_{23}= \gamma_{24}= \gamma_{25}=0$. Vectors β_2 and β_3 can be rewritten as:

$$B = \begin{pmatrix} \beta_{21} & \beta_{22} & \beta_{23} & 0 & 0 & 0 & 0 & 0 & \beta_{2i} \\ 0 & 0 & 0 & \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} & \gamma_{3i} \end{pmatrix}'$$

Define the exclusion restriction for β_3 as:

$$R_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Therefore, the rank condition is given by:

$$R_3 B = \begin{pmatrix} 0 & \beta_{21} \\ 0 & \beta_{22} \\ 0 & \beta_{23} \end{pmatrix}$$

The rank of matrix R_3B is 1, and given that there are two endogenous variables in equation (3), we can conclude that the system is identified.

4.3.3 Explanatory variables

4.3.3.1 <u>Work equation.</u>

I follow the existing literature in modeling the work decision, including household characteristics such as the age of the respondent (AGE and AGE-SQ), gender, his or her education, and the household's size and life cycle. I cannot directly account for the age of the spouse (or second adult) because not all households contain a spouse. Instead, I construct a dummy variable for whether the respondent has a spouse, and then interact it with the spouse's age and gender. Our data do not include a separate measure of non-work income.

I also include two locality-specific variables: the unemployment rate in the MSA and the percent of the population that is retired in the household's census tract.

4.3.3.2 Distance-to-work equation.

Analysis of distance-to-work should capture both (i) the DTW choice of household i within city j and (ii) the mean and variance of DTW across households for a city with characteristics Z_j. In essence, I am treating distance-to-work as a "demand" and thus examine analogs to its price (which is common to all households within an MSA) and income and taste (which are household specific).

Note that it is conceptually possible for the distance-to-work to be zero for all households regardless of city size – imagine even a very large city in which everyone lives right next to his or her work. There is a subtle reason why I must temper this claim: The definition of the MSA is itself endogenous and dependent on the work decisions of thousands of households. For example, while it is conceptually possible for everyone to have a close-to-zero distance-to-work, it seems likely that this outcome would then lead to this area being divided into multiple smaller MSAs. That is, MSA boundaries are effectively implied by there being enough high DTWs in an area. I do not speculate further here on the endogenous-MSA problem. My claim that small DTWs could occur under any size MSA provides useful intuition; any further exploration of this problem requires much more sophisticated modeling of the distribution of DTWs within an MSA. Thus, the key to understanding the distribution of DTW is to understand those factors that affect all households and lead them, in general, to live close to or far from their jobs.

City-specific variables. I follow Bento et al (2005) in the choice of city specific variables, though I complement their measures with other measures I constructed based on the data available. This allows me to use variables like commuting speed, which has To capture the "price" of living close to work, I measured the congestion of each city by calculating for all households i in city j the ratio of DTW_{ij} to the reported Time to Work (TTW_{ij}), which is elicited by the survey in a manner similar to DTW. The ratio DTW_{ij}/TTW_{ij} is a measure of the commuting speed experienced by household i.[8] I then calculated the median commuting speed in city j over all sampled households, labeled MEDSPEEDj. The higher is MEDSPEEDj, the quicker people get to work in city j for a given DTW. This speed lowers the cost of living farther from work.

A second desirable measure of the price of DTW is each city's housing price-distance gradient: a measure of how quickly housing (rental) prices fall as one moves away from major job centers. A steeper price-distance gradient would mean that it is relatively more expensive to live close to a job center; thus, households would, on average, choose to live farther away.

Note, however, that the price-distance gradient's effects should be captured by the variables I am able to include, MEDSPEED and GINIJOBS. The reason is that housing prices are endogenous and therefore how quickly they fall off as one moves

---

[8] The measure of commuting speed is calculated at the individual level. The NPTS collects information on each individual's TTW and each individual's DTW. I build the ration and then obtain the average commuting speed at the city level.

away from the center city should reflect a combination of the supply of close-in housing, which GINIJOBS should capture, and the difficulty of commuting, which MEDSPEED should capture.[9] Therefore, I do not include a separate measure of the housing price-distance gradient.

I also consider two variables that capture city characteristics but without an explicit connection to jobs or housing. These are AREA and CITYSHAPE.

BCMV argue that cities that are closer to circles (CITYSHAPE closer to 1) should have lower VMT. I believe this intuition is especially relevant for the distance-to-work, since households have fewer options for jobs-housing connections and are therefore more constrained by the city network. This claim is complex. It is perhaps easiest to see for a University professor. For a given residence a household has a choice of *n* places to work where *n* is the number of universities in town. The same household will have a larger choice of shopping or entertainment destinations. This reasoning holds as long as the individual is more restricted in terms of his occupation than in terms of his non-commuting destinations. Shifting the perspective to the number of houses available from a given workplace does not alter this conclusion: there will almost always be more flexibility for non-commuting miles than commuting miles. Therefore, I include CITYSHAPE as an explanatory variable in the DTW equation but not the VMT equations.

I also include population density as measured at the city (MSA) level. I discuss its interpretation below. The relationship between VMT and residential neighborhood density has been studied extensively in the literature (Steiner, 1994;

[9] Despite this assessment, I think it would be worthwhile to examine the role of city-specific housing price-distance gradients. I leave this for future research, since the construction of such a measure is complex, and the housing price-distance gradient would be worthy of analysis on its own.

Schimek, 1997; Wilson, 1998; Badoe and Miller, 2000). Neighborhood density represents too limited a measure of the community in which a most of the household's vehicle travel takes place, however. I focus instead on city-level density, which provides a better measure of the density of the area in which household VMT occurs.

Finally, for households that have at least one worker in the household, access to transit becomes an important variable to consider. I therefore create a variable called Access to Transit (TRANSIT) defined as the percentage of people in a given city that live within a 0.5 mile radius of a transit stop. A transit stop is defined as any source of public transport, and includes buses, metro, light rail, streetcar, or commuter train. This variable is constant within a particular city, but differs between cities. I do not make a distinction between rail and bus transit. Though data on each household's distance to a transit stop is available, I build a city wide measure that is truly exogenous to the household.

<u>*Household characteristics.*</u> For household characteristics I use a typical set of measures commonly found in the literature (see Table 2.3). In considering the role of income, I expect (negative) DTW to be a normal good; that is, people prefer to live closer rather than farther to work, ceteris paribus,. Thus, a higher household income should be associated with a shorter DTW. Because of differences in housing costs across MSAs, I measure household income *relative* to the city's median income. For consistency in income measures, I use the median income of NPTS respondents. Relative income is then calculated as $RELINCOME_{ij} = Income_{ij}/ Median\text{-}Income_j$. To my knowledge, this approach has not been used before in the literature. The main

reason being, in my opinion, that residential location models typically focus on a household's choice within a single city (Bayoh et al., 2006) therefore there is no need to correct for the relative income between MSAs.

4.3.3.3 Vehicle Miles Traveled equation

The VMT equation is a reduced form equation that reflects factors that affect both the number of cars owned and the number of miles driven conditional on that set of cars. Thus, I include variables that may enter into either of these choices.

Income is a key variable in explaining vehicle ownership and thus widely used in explaining VMT (see Table 2.3 and Train (1986)). For the VMT regressions I use absolute, not relative, income since most of the purposes of non-commute VMT involve purchase of items that are tradable across cities, unlike houses. The costs of VMT (maintenance, insurance, and operating costs) are also more likely to involve nominal (rather than relative) prices.

The VMT literature typically focuses on the "driving price" per mile, which is the state-level gas price adjusted for the miles-per-gallon of the household's cars. I focus instead on the truly-exogenous state-level gas prices.

Other important household-level variables are the number of drivers, number of children, and the household's life cycle.

4.3.4 Taste for driving

I cannot observe individuals' taste for driving. But this taste, if unaccounted for, could seriously affect my estimates and inferences. The reason is that people with an affinity for driving will choose to live relatively far from work and drive more than a low-taste-for-driving household that lives that same distance from work. This taste manifests itself as a correlation between the error terms in equations (2) and (3), this is $\sigma_{\varepsilon\upsilon}$. Failure to take this correlation into account would then lead me to overestimate the effects of DTW on VMT.

Note that in the estimation procedure described in section 4.2 and in the appendix to the dissertation, we are able to recover the parameters σ_{ε} and, σ_{υ} and ρ, where $\sigma_{\varepsilon\upsilon} = \rho\ \sigma_{\varepsilon}\ \sigma_{\upsilon}$.

4.4 Research Questions and Hypotheses

Our main questions concern (i) the role of city characteristics in explaining DTW; (ii) the effect of DTW on VMT; (iii) the taste for driving; (iv) the behavior of VMT conditional on DTW and the implied taste for driving, with special attention to the role of city characteristics; and (v) differences in VMT between workers and non-workers. I do not focus on the results of the WORK equation in this paper.

4.5 Results

I focus on the specification shown in (1)-(4). Table 4-2 shows the summary statistics for the samples used. Our main regression is labeled Regression #1 in Tables 4, 5 and 6. Regressions #2 to #6 are further variations on my sample or model and are discussed in further detail in Section 4.6. Finally, in section 4.6 I also include regression results for a different subset of the data. I estimate the model in chapter 4.3 only for households in which every member commutes by car.

To gauge the economic significance of the estimated coefficients, I calculate the implied elasticities for a representative household in six cities (MSAs), the approach adopted by BCMV.[10] Since I exclude New York City from my sample, I substitute Phoenix, one of the fastest growing MSAs. These calculations are shown in Table 4-7.

4.5.1 Distance-To-Work equation

<u>*City characteristics*</u>. For DTW, I use five variables to characterize cities. These are AREA, CITYSHAPE, GINIJOBS, MEDSPEED, and POPDENSITY.

There are three related findings (see Table 4-7): (i) MEDSPEED has the largest effect on DTW of all variables used to characterize cities. This effect is positive, as expected: a higher commuting speed lowers the cost of living farther from work and therefore induces greater DTW. For the other four effects the implied elasticities are quite small, except for POPDENSITY in Chicago.

Our measure of congestion (MEDSPEED) suggests that people will locate around 8.8 percent further from their work when commute speed increases by 10 percent. These percentages imply that if commute speed were to increase from a national average of 29.5 mph to 32.5 mph, I expect people to live on average 1.24

---

[10]Elasticities are estimated for each equation separately. In the DTW equation, the elasticity is $\varepsilon_{Xi} = [\partial \ln DTW/\partial X_i]\cdot X_i = \beta_i \cdot X_i$. In the VMT equation, the elasticity is $\varepsilon_{\ln X1} = [\partial VMT/\partial \ln X_1]\cdot[1/VMT] = \gamma_1 \cdot [1/VMT]$ or $\varepsilon_{X2} = [\partial VMT/\partial X_2]\cdot[X_2/VMT] = \gamma_2 \cdot [X_2/VMT]$. Elasticities are calculated at the mean value of the variable for that city.

miles further their jobs than they do now. That is, the mean one-way distance-to-work would increase from 9.7 to 10.9 miles.

An alternative explanation of the MEDSPEED result is that it reflects the distribution of the housing stock rather than the cost-of-commuting. This effect could occur if a more dispersed housing stock is an important component of lower commuting speeds (a claim that I do not verify here). Under such a circumstance, households could be relatively unconcerned about commuting speeds and instead simply selecting among a wider array of housing choices. This claim is related to my previous one: A greater housing stock does not imply greater DTW. This explanation would be consistent with my overall interpretation of HHDTW behavior; see below.

(ii)Two variables related to city size but not necessarily to commuting (AREA and POPDENSITY) have larger estimated elasticities than the two variables most directly related to potential commuting patterns (GINIJOBS and CITYSHAPE).

(iii) POPDENSITY has a positive effect on DTW, an unexpected result. The likely explanation is straightforward, however: A higher population density, conditional on MSA area, indicates a higher population. A higher population indicates a greater stock of housing, roughly speaking. This in turn means a greater variety of available household locations. One way to test this explanation is to run regression #1 including POPDENSITY and excluding AREA. After running this regression, I find that the coefficient on POPDENSITY is still positive. Therefore a more likely explanation is that the data is merely reflecting the fact that larger more denser cities have, on average, greater DTW (recall that our measure of POPDENSITY is at the city level).

There is a simple and compelling explanation for this pattern of results. In the U.S., the range of jobs-houses-commute combinations in any given city is large. Even very small cities or ones with unusual shapes or a high jobs-housing imbalance offer a wide range of available housing and commuting patterns. Therefore, GINIJOBS or CITYSHAPE put little structure on the choice set. Commuting speed may matter because it reflects the distribution of the housing stock.

In other words, the general picture of city-level variables is that in cities with more housing options, people tend to live farther from work. Other city characteristics just do not matter much. People appear to spread themselves out based on a social or economic dynamic that is simply not much affected by the city's "smart-growth" attributes.

Household characteristics. Households with higher relative income live farther from their work. This is an unexpected result but it again likely reflects a housing stock effect rather than a pure income effect. The reason is that newer and larger houses tend to be in less dense suburbs, requiring greater distances to work on average.

More established households, as measured by respondent's age, tend to locate closer to their work. They were there first, so to speak.

4.5.2 Vehicle miles traveled equation

4.5.2.1 <u>*The effect of DTW on Vehicle Miles Traveled*</u>

The regression of VMT on HHDTW (Table 4-5) allows me to calculate the contribution of commuting miles to total miles driven. Recall that HHDTW equals DTW times the number of workers in the household. Note that the term "commuting mile" is an accounting construct used for expository purposes. I do not claim that DTW or HHDTW are measures of actual annual commuting miles.

Since VMT and HHDTW are measured in the same units, It is easiest to see the effect of DTW in terms of a marginal effect rather than as an elasticity. I estimate that one additional mile of HHDTW for one worker leads only to 0.33 additional VMT; that is, substantially less than one. It is easiest to think of this as a strong degree of task-sharing, although I cannot confirm this claim here.

Consider a household with 2 workers, each of whom would normally drive 15,000 miles, of which 5,000 are commute miles and 10,000 are non-commute miles. VMT is 30,000 and the one-way distance-to-work is 10.4 miles for each of the workers. Increasing the commute miles by 2,400 per worker, or 5 one-way miles each, would increase overall miles only by 1,584 miles (using δHHDTW = 0.33)[11], to 31,584 miles. Non-commute miles would actually decrease by 3,220, from 20,000 to 16,784.

[11] The expected contribution of commuting miles to VMT prior to the change is $0.33 \times 5{,}000 = 1650$ miles. The expected contribution after the change is $0.33 \times (5000 + 4800) = 3234$. The change in commute miles is $3234 - 1650 = 1584$ miles

4.5.2.2 Behavior of VMT conditional on distance to work

City characteristics. In regression #3 in Table 4-5, I adopt a specification that focuses on the role of what I perceive ex ante to be "non-commuting" city characteristics. I include population density because it represents, roughly speaking, the density of friends and (in most cases) other non-commuting destinations. I expect that this characteristic is the key determinant of non-work miles. The higher is the density, the lower should be VMT.

I exclude access to transit, city area, city shape, the jobs-housing balance, and commuting speed because these variables are expected to be more relevant to commuting and HHDTW. Because non-commuting driving would seem to have greater flexibility than driving for commuting, it seems more likely that VMT would reflect a demand for driving that would be unaffected by city size, for example.

POPDENSITY has a statistically significant, negative (as expected), but extremely small coefficient. A one percent increase in population density leads to around a 0.2 percent decrease in VMT.

In regression #1, I examine roles for AREA and CITYSHAPE in explaining VMT, acknowledging that the underlying reasons for any effects (and therefore the interpretation of any results) are less compelling than for they are for the DTW equation. The calculated elasticities for AREA and CITYSHAPE are, -0.01 and 0.05, respectively; in other words, quite small, as hypothesized. Access to transit does have a higher effect on VMT: the calculated elasticity is -0.11. In other words, increasing the percent of the population located within a half mile of a transit stop by 10 percent decreases VMT by 1.1 percent. This effect is important from a policy perspective, as it indicates that a higher population living close to transit stops will reduce travel demand, presumably through mode change.

Price and income effects. I expect households to increase their VMT as their income increases. I estimate the income elasticity to be around 0.18. This is a short-run elasticity conditioned on HHDTW.

Our estimate of income elasticity is at the low end of estimates in the literature (see Tables 4-8 and 4-9). Though in principle this result could be due to failure of other studies to condition on distance-to-work, this does not seem to be the case. I estimated the model without including HHDTW in the VMT equation. In this case I obtained an income demand elasticity for VMT equal to 0.21—this is, 20% larger than originally estimated but still in the lower end of estimated elasticities. Note that many other studies, not included in the table, calculate an income elasticity of VMT per vehicle. Since vehicles-owned is strongly influenced by income, these estimates are not directly comparable to ours and would be far below the income elasticity of total household VMT.

With respect to gas price elasticity of VMT, table 4-9b presents a summary of the elasticities found in the literature. Several authors have provided reviews of the price elasticity of vehicle miles traveled, including Litman (2007), Goodwin et al (2003), and de Jong and Gunn (2001). Estimates of gas price elasticity of VMT in the literature are estimated around -0.1 in the short run and -0.3 in the long run. I estimate the gas price elasticity of VMT to be around -0.15. Again, this is a short run estimate as we are conditioning my VMT equation on distance to work. Our estimates of gas price elasticity range from -0.10 (in regression #6) to -0.41 (in regression #5). When

I run the same model without conditioning on HHDTW, my estimated parameter for gas price is not significantly different than zero.

Selection effects. The coefficient on λ, β_λ, captures the unobserved correlation between VMT and WORK. I expect workers to be likely to drive more because on average they are healthier (physically or mentally), even conditional on DTW and income. Therefore I expect $\beta_\lambda > 0$.

The coefficient on λ is statistically significant and negative, which implies that unobserved influences on the decision to work are inversely correlated with unobserved components of VMT. I hypothesized that health status would be an important component of both equations, with a positive effect in each case; that is, a healthier individual would be more likely to work and also likely to drive more. Therefore, this result is unexpected. One possibility is that the main component of the error in WORK is wealth, since a higher wealth would lead a household to be less likely to work but to drive more. [What if income is endogenous? Would that have an effect?]

Other household characteristics. Other household characteristics will also influence VMT. The greatest effect comes from the number of drivers, as would be expected. Adding one new driver to the household will increase annual VMT by 21 percent on average. This is a substantial increase of roughly 4,750 miles, but still below the national average VMT per driver.

4.5.2.3 *<u>Behavior of Workers vs. Non-workers</u>*

Differences in the mileage of workers and non-workers have not been much explored. To the extent that any such differences are due solely to differences in household composition, life cycle, or income then the coefficients in (3) and (4) should be similar. Of course, some relevant variables, such as hours worked or wealth, are unobserved.

The most striking difference is the coefficient on gas price, which suggests that non-workers are much less responsive to gas price changes than workers. The computed gas price elasticity for non-workers is -0.07, which may be compared with -0.15 for workers (see Table 4-8.) Our VMT results for workers suggested that workers have a high degree of task-sharing; that is, they appear to combine work and non-work trips. This option may contribute to the higher gas price elasticity for workers.

Note that except for the gas price results, non-workers behave similar to workers except that they drive less. The difference in the amount driven between these two groups can be explained based on the difference in the variable means (the Xs) and the estimated coefficients (the βs). My model predicts a difference between workers and non-workers of 13.9 thousand miles per year. This difference can be attributed roughly 73% to differences in the Xs and 27% due to differences in the βs.[12]

---

[12] The difference in VMT between workers (w) and non-workers (nw) is given by

$$\begin{aligned}\Delta VMT &= VMTw - VMTnw = Xw\beta w - Xnw\beta nw \\ &= (Xw\text{-}Xnw)\,\beta w + (\beta w\text{-}\,\beta nw)\,Xnw = \Delta X\,\beta w + \Delta\beta\,Xnw \\ &= 10.15 + 3.74 = 13.9 \text{ (thousand miles)}\end{aligned}$$

4.5.3 Taste for driving

I interpret the ε-ν correlation as the unobserved taste for driving after other observed taste effects (through household demographic variables) have been accounted for. I estimate a value for taste for driving equal to 0.32. The estimated effect is relatively large, which means that this taste important and needs to be taken into account in the estimation of process.

To see the effects of this correlation, consider the estimated parameters of VMT when I do not account for this correlation. This is shown as Regression 2 in Table 4-5. I am most interested in the change in the DTW coefficient, which changes from 0.33 to 0.36. That is, if I ignored the so-called taste for driving, I would overestimate the effect of DTW as hypothesized. Furthermore, the parameter on DTW is overestimated, in this particular case, by approximately 10%.

4.5.4 Predicted DTW and VMT for 6 cities

In this section, I conduct an exercise similar to BCMV to analyze the effects of city characteristics on DTW and VMT. The main idea is to predict the net effect on DTW and VMT of moving a young household from each of my six representative cities to each of the others.

To model a young household, I first selected households whose reference person is under 30 years old to obtain an "average young household" for each city. I then predicted its DTW and VMT for each of the six cities, using Regression 4. The difference between the predicted VMT in the home city and the other 5 cities is thus the predicted net effect of moving this household across the U.S. Table 4-10 contains descriptive statistics of the "average young household" in each city. The predicted net effects on DTW and VMT are shown in Table 4-11.

Note that this is a slightly different interpretation for an exercise that is very similar to BCMV. BCMV compared predicted VMT in different cities and described the results as showing the effect of the differences in city characteristics. In other words, they implicitly pose the question as: What if Houston became more like Boston?

I frame the question differently. Because there is a fair bit of household mobility across cities in the U.S. and because city characteristics (as I have characterized them) are so slow to change, the greatest effect on VMTs will likely come from changes in where people live across the U.S. Thus, I frame the question as: What happens when a household moves from city A to city B? The fact that such moves are a common feature in the U.S. economy may make such a prediction particularly informative.

The results of this exercise can be seen in Table 4-11. Consider the first line of the table. It tells me how much closer or further a representative household that moves from Atlanta to each of the other cities will likely locate from its workplace, measured in yearly distance to work (DTW). Note that a movement to newer cities like Phoenix or Houston will result in a substantially longer commute. Movement to older cities such as Chicago or Boston will result in a shorter commute.

I then predict this household's VMT. Note that the preceding results (Section 4.5.4) suggest that commuting miles have relatively important marginal effects on VMT. Thus I find that a young household moving from Atlanta to Boston, Chicago, or San Diego will drive 792, 2163, and 675 miles less per year, respectively (Table 4-11). It is interesting to note that even though a household that moves from Atlanta to Boston or Chicago is expected to increase its DTW, its annual VMT are expected to decrease.

Our results for the VMT equation differ from those obtained by BCMV. Going from a city like Atlanta to another like Boston implies a reduction in VMT of about 792 miles per year. BCMV found that going from Atlanta to Boston would reduce miles by a larger amount, approximately 4,100 miles per year. Boston is about one-third the area of Atlanta, but its average VMT is just 7 percent less. Boston is approximately 3 times denser than Atlanta. These two forces basically offset each other, making the difference in VMT not as large in my model.

The predictions in Table 4-11 must be interpreted carefully, of course. Households that move between cities are not a random sample of households, not even of young households. Predicted DTW reflects average effects and does not necessarily capture the marginal DTW of a new household in a particular city, even if I were able to predict the characteristics of the marginal (i.e., moving) household.

4.6 Robustness of Estimated Effects

I next analyze how robust my results are to different subsets of the data and to a different functional form specification. Results are shown in Tables 4-4 to 4-6 and are numbered regressions #3-#6. In regression #3, the only variable describing city form included as an explanatory variable for DTW is POPDENSITY. The variables TRANSIT, CITYSHAPE, GINIJOBS, and HWYDENS are excluded. Sample size increases because I have data on CITYSHAPE for only 110 MSAs. In regressions #4 and #5 I include all explanatory variables describing city form except CITYSHAPE. In #6 I exclude the 25 metropolitan areas that have rail transit.[13]

The results obtained in regression #1 are highly robust. Coefficients and calculated elasticities are nearly identical between regressions 1, 3, 4, 5, and 6.

In regression 6, I estimate a log-log specification for VMT using the sample of cities in regression 1. Because coefficients are not directly comparable across regressions 1 and 6, I report the calculated elasticities in Table 4-8. The results, except for gas price and number of drivers, are not much affected by this change in functional form. Therefore, I did not pursue further the question of functional form.

Finally, I also estimate the model using a sample of households that commute by car. A household that commutes by car is defined as that where all its working members commute by private vehicle. With this restriction I lose 7 percent of the sample. The NPTS data shows that over 90 percent of the households in the U.S. commute by private vehicle. The results are shown in tables 4-12 to 4-14. I included in these results those obtained for regression #4 in tables 4-4 and 4-5 in order to compare how the estimated parameters differ.

The general result for distance to work is that elasticities for individuals that commute by private vehicle are higher than when transit commuters are included in the sample. In particular, note that the effect of MEDSPEED on DTW has increased fivefold (Table 4.14a). The elasticity of DTW relative to MEDSPEED is 9.3, which

[13]Atlanta, Baltimore, Boston, Buffalo-Niagara Falls, Chattanooga, Chicago, Cleveland, Denver, Detroit, Hartford, Jacksonville, Los Angeles-Long Beach, Memphis, Miami, New Orleans, New York, Philadelphia, Pittsburgh, Sacramento, San Diego, San Francisco, San Jose, Seattle, Tampa-St Petersburg, Washington, DC.

implies that a 1 percent increase in commute speed will increase distance to work by 9.8 percent. Using the average speed in section 4.5.1, I find that increasing commute speed from 29.5 mph to 32.4 mph, implies that the average household one-way distance to work will increase from 9.7 to 19.2 miles. This increase means that households that commute by private vehicle are far more sensitive to changes in urban form—this may be a result of the greater mobility households gain from owning a vehicle. An increase in median speed or income allows households to locate further from work.

Finally, with respect to the VMT equation, only the gas price elasticity of vehicle miles traveled changes when only private vehicle commuters are included in the estimation (Table 4-14b). The higher gas price elasticity (in absolute terms) implies that households will be more affected by a price increase, and therefore will reduce their demand for miles more than when transit commuters are included in the sample.

4.7 Concluding Comments

Summary and interpretation of results. I find, contrary to my expectations, that city characteristics related to jobs, housing, and commuting – attributes that might seem key to household location decisions – have little effect on households' distances-to-work.

The reason, I believe, is that in U.S. cities, households have plenty of choices of where to work and where to live. This array of choices reflects the economic diversity and dynamism of these cities. Given this dynamism and array of choices, the layout of the city has little additional influence on household location choice. The value provided by this wide range of housing choices appears to outweigh many of the city characteristics that might more directly affect households' commutes.

This result suggests that while it may be possible for governments to influence distance-to-work through subsidies or policies explicitly targeting distance-to-work, it will likely not be possible to influence it through changes in city design. Our results suggest that cities may already be just too complex and varied.

The consequences of this conclusion, however, are weakened by my conclusions about VMT. Our main finding, put bluntly, is that distance-to-work does not constrain VMT very much. In retrospect, this conclusion too should have been anticipated. The economic and social dynamism that underlie the distance-to-work results similarly make it possible for households to widely adjust VMT, regardless of distance-to-work. The fact that commuting miles are a relatively minor part of a household's VMT also contributes to the small effect of distance-to-work. Based on these findings, policies to reduce commuting should not be expected to have much influence on VMT.

Directions for future research. I propose two directions for future research, one following and building on the existing literature. A natural extension is to model vehicle ownership (see Chapter 5) or commute mode. Another extension would be to use actual days-worked to compute HHDTW. I also feel that sample attrition due to missing income and VMT data has received insufficient attention in this literature.

A second direction is to tackle a different set of questions. Two such questions in particular are suggested by my research. First, I wonder how these conclusions would change if I characterized household location using Time-to-Work (TTW) rather than DTW. TTW may be a more accurate measure of the costs of living far from one's work than is DTW. It has two potential drawbacks, however: (i) its role in contributing to externalities is weaker, and (ii) it cannot be naturally measured on the same scale as VMT and DTW allow.

Second, the high degree of household mobility in the U.S. suggests a future research agenda that explicitly studies the DTW or TTW decisions of movers and the implications of such mobility for VMT.

4.8 Tables for Chapter 4

Table 4-1. Number of observations (households)

Description	Number of observations
Initial sample (total observations in 1995 NPTS)	**42,033**
Observations dropped	**23,910**
Not in MSA or MSA with less than 20 observations	10,629
Households with no vehicles	1,887
No income data	4,938
No VMT data	1,735
Missing values in other explanatory variables	4,721
Total usable observations (134 MSAs)	**18,123**
No workers in household (WORK = 0)	**3,189**
At least one retired individual in household	2,331
No retired individuals in household	858
At least one worker in household (WORK = 1)	**14,934**

Table 4-2. Variable means

Variables	No workers in HH	At least one worker in HH	Entire sample
Dependent variables			
WORK	0.00	1.00	0.79
DTW	-	5.93	4.66
VMT	8.17	19.61	17.16
Independent variables			
City (MSA) level variables			
Area	2.61	2.65	2.64
Pop. Density	1.42	1.34	1.36
Jobs-Housing Gini	0.40	0.40	0.40
Median commute speed	29.33	29.62	29.56
City Shape	0.69	0.69	0.69
Unemployment rate	0.05	0.05	0.05
Percent population over 65	15.15	12.72	13.24
Log of gas price (state-level)	0.17	0.17	0.17
Household level variables			
Log of relative income	-0.78	0.06	-0.12
Log of income	2.87	3.74	3.55
Household size	1.80	2.87	2.64
Number of children <6	0.03	0.10	0.08
Number of children 6-18	0.12	0.51	0.43
Number of adults over 65	0.89	0.10	0.27
Number of drivers	1.19	1.95	1.79
Number of workers	0.00	1.75	1.39
Age of reference person	63.66	43.16	47.55
Age of spouse (Spouse dummy × Spouse age)	26.40	28.43	27.99
Sex of spouse (Spouse dummy × Spouse = Female)	0.34	0.55	0.50
Household is adults only	0.21	0.45	0.40
Number of observations	**3,189**	**14,934**	**18,123**

Table 4-3. Probit regression for WORK=1

Variable	#1 Coeff. *(Std. err.)*
Age of reference person	0.08** *(0.01)*
Age of reference person squared	-0.001** *(0.00)*
Household size	0.43** *(0.06)*
Life cycle - adults, no children	1.69** *(0.06)*
Life cycle - adults with children	0.69** *(0.11)*
Life cycle – retired	1.53** *(0.17)*
Unemployment rate	-2.07 *(1.53)*
Percent population over 65	-0.01** *(0.00)*
Age of spouse	0.01** *(0.005)*
Age of spouse squared	$1.6x10^{-4}$** *($6.7x10^{-5}$)*
Sex of spouse	0.03 *(0.07)*
Constant	-2.03** *(0.3)*
Number of observations	20,242

Standard errors in parenthesis. **Significant at 5% level

Table 4-4. DTW Results for WORK=1. (Dependent variable: Ln of DTW)

Dependent variable	#1 Coeff. *(s.e.)* [a]	#2 Coeff. *(s.e.)*	#3 Coeff. *(s.e.)*	#4 Coeff. *(s.e.)*	#5 Coeff. *(s.e.)*	#6 Coeff. *(s.e.)*
Area of MSA (1000 square miles)	0.06	0.02	0.06	0.08	0.09	0.08
	(0.012)	*(0.004)*	*(0.012)*	*(0.014)*	*(0.020)*	*(0.014)*
Pop. Density in MSA	0.13	0.13	0.13	0.26	0.29	0.26
	(0.020)	*(0.015)*	*(0.020)*	*(0.016)*	*(0.023)*	*(0.016)*
Jobs-Housing Gini	0.18	0.03	0.18	-	-0.05	-0.04
	(0.058)	*(0.065)*	*(0.058)*		*(0.095)*	*(0.074)*
Median speed in city (miles per hour)	0.03	0.02	0.03	0.06	0.06	0.06
	(0.005)	*(0.004)*	*(0.005)*	*(0.005)*	*(0.006)*	*(0.005)*
Cityshape	0.21	0.05	0.20	-	-	-
	(0.081)	*(0.058)*	*(0.081)*			
Log of relative income	0.38	0.33	0.38	0.35	0.35	0.35
	(0.017)	*(0.017)*	*(0.017)*	*(0.017)*	*(0.021)*	*(0.017)*
Household size	0.03	0.05	0.03	0.10	0.11	0.10
	(0.012)	*(0.011)*	*(0.012)*	*(0.011)*	*(0.014)*	*(0.011)*
Number of children <6	-0.12	0.01	-0.12	-0.19	-0.20	-0.19
	(0.037)	*(0.035)*	*(0.037)*	*(0.035)*	*(0.044)*	*(0.035)*
Number of children 6-18	-0.07	-0.08	-0.07	-0.17	-0.18	-0.17
	(0.016)	*(0.015)*	*(0.016)*	*(0.015)*	*(0.019)*	*(0.015)*
Age of reference person	-0.01	-0.01	-0.01	0.00	0.00	0.00
	(0.001)	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*
Constant	7.24	7.62	7.24	5.98	5.86	6.00
	(0.156)	*(0.145)*	*(0.156)*	*(0.150)*	*(0.195)*	*(0.154)*
Number of observations	**12,239**	**12,239**	**12,239**	**14,934**	**9,524**	**14,934**
MSAs	**109**	**109**	**109**	**134** [b]	**110** [c]	**134** [b]

Notes:
[a]) Standard errors in parenthesis; [b] Includes MSAs w/o CITY SHAPE data); [c] Excludes cities for which have rail transit

Each regression in Table 4-4 was estimated simultaneously with its corresponding regression in Table 4-5. For instance, in order to understand why coefficients differ between regressions #1 and #2 in Table 4-4, one must also look at regressions #1 and #2 in Table 4-5 in order to see which regressors were included.

Table 4-5. VMT Results for WORK=1.

Dependent variable: VMT

Independent variables	#1 Coeff. *(s.e.)* [a]	#2 Coeff. *(s.e.)*	#3 Coeff. *(s.e.)*	#4 Coeff. *(s.e.)*	#5 Coeff. *(s.e.)*	#6 Coeff. *(s.e.)*
Log of income	4.09	4.02	4.07	4.09	4.25	0.39
	(0.206)	*(0.206)*	*(0.207)*	*(0.185)*	*(0.224)*	*(0.013)*
Log of gas price	-3.16	-3.22	-4.38	-3.45	-9.17	-0.10
	(1.876)	*(1.877)*	*(1.655)*	*(1.714)*	*(2.211)*	*(0.125)*
Population density (	-0.23	-0.26	-0.46	-0.51	-1.66	-0.04
	(0.205)	*(0.205)*	*(0.187)*	*(0.170)*	*(0.285)*	*(0.002)*
Access to transit	-4.91	-4.86	-	-3.78	-0.47	-0.29
	(1.302)	*(1.303)*		*(1.080)*	*(1.282)*	*(0.079)*
Area of MSA (1000 square miles)	-0.05	-0.06	-	0.01	-0.03	0.00
	(0.046)	*(0.046)*		*(0.043)*	*(0.047)*	*(0.003)*
Cityshape	1.50	1.46	-	-	-	-
	(0.714)	*(0.714)*				
Highway density	-0.12	-0.12	-	-0.03	-0.05	-0.02
	(0.104)	*(0.104)*		*(0.095)*	*(0.187)*	*(0.007)*
Number of drivers	4.61	4.53	4.63	(4.47	4.30	0.24
	(0.180)	*(0.178)*	*(0.180)*	*(0.161)*	*(0.202)*	*(0.012)*
Number of children < 6	0.51	0.54	0.52	0.24	-0.06	0.04
	(0.385)	*(0.386)*	*(0.386)*	*(0.344)*	*(0.425)*	*(0.025)*
Number of children 6-18	-0.20	-0.19	-0.20	-0.11	-0.06	-0.02
	(0.139)	*(0.139)*	*(0.140)*	*(0.125)*	*(0.154)*	*(0.009)*
Number of adults over 65	-1.09	-1.09	-1.10	-1.06	-1.26	-0.05
	(0.415)	*(0.416)*	*(0.416)*	*(0.371)*	*(0.468)*	*(0.027)*
DTW x Number of workers	0.33	0.36	0.33	0.33	0.33	0.01
	(0.015)	*(0.013)*	*(0.015)*	*(0.012)*	*(0.015)*	*(0.001)*
Inverse mills ratio	-0.01	0.07	-0.01	0.15	0.65	-0.03
	(0.628)	*(0.628)*	*(0.629)*	*(0.558)*	*(0.689)*	*(0.041)*
Constant	-4.36	-4.19	-5.86	-4.04	-4.66	0.88
	(1.199)	*(1.199)*	*(0.798)*	*(0.924)*	*(1.151)*	*(0.067)*
rho	0.04	-	0.04	0.03	0.03	0.05
	(0.011)		*(0.011)*	*(0.009)*	*(0.011)*	*(0.009)*
Number of observations	**12,239**	**12,239**	**14,934**	**14,934**	**9,524**	**14,934**
MSAs	**109**	**109**	**134** [b]	**134** [b]	**110** [c]	**134** [b]

Notes:
[a] Standard errors in parenthesis; [b] Includes MSAs w/o CITY SHAPE data); [c] Excludes cities which have rail transit; [d] Dependent variable for equation #6 is **ln(VMT)**

Table 4-6. VMT Results for WORK=0.

Dependent variable: VMT

Independent variables	#1 Coeff. *(s.e.)* [a]	#2 Coeff. *(s.e.)*	#3 Coeff. *(s.e.)*	#4 Coeff. *(s.e.)*	#5 Coeff. *(s.e.)*	#6 Coeff. *(s.e.)*
Log of income	2.75	2.75	2.71	2.72	2.75	0.40
	(0.244)	*(0.244)*	*(0.218)*	*(0.218)*	*(0.274)*	*(0.031)*
Log of gas price	-0.81	-0.81	0.00	-1.40	-1.64	-0.11
	(2.770)	*(2.770)*	*(2.216)*	*(2.494)*	*(3.133)*	*(0.352)*
Population density	-0.72	-0.72	-0.53	-0.62	-1.34	-0.10
	(0.316)	*(0.316)*	*(0.242)*	*(0.259)*	*(0.459)*	*(0.037)*
Access to transit	2.08	2.08	-	0.63	2.28	-0.06
	(1.861)	*(1.861)*		*(1.588)*	*(1.892)*	*(0.224)*
Area of MSA (1000 square miles)	-0.04	-0.04	-	0.00	0.03	-0.01
	(0.062)	*(0.062)*		*(0.058)*	*(0.064)*	*(0.008)*
Cityshape	0.64	0.64	-	-	-	-
	(1.016)	*(1.016)*				
Highway density	0.12	0.12	-	0.19	0.67	0.05
	(0.171)	*(0.171)*		*(0.155)*	*(0.279)*	*(0.022)*
Number of drivers	4.58	4.58	4.49	4.50	4.74	0.53
	(0.363)	*(0.363)*	*(0.326)*	*(0.326)*	*(0.407)*	*(0.046)*
Number of children < 6	2.82	2.82	2.29	2.24	2.88	0.07
	(1.198)	*(1.198)*	*(1.067)*	*(1.069)*	*(1.324)*	*(0.151)*
Number of children 6-18	0.28	0.28	-0.26	-0.25	-0.27	-0.08
	(0.466)	*(0.466)*	*(0.412)*	*(0.413)*	*(0.525)*	*(0.058)*
Number of adults over 65	-0.63	-0.63	-0.52	-0.53	-0.48	-0.05
	(0.306)	*(0.306)*	*(0.271)*	*(0.271)*	*(0.340)*	*(0.038)*
Inverse mills ratio	-1.92	-1.92	(-2.04	-2.03	-1.88	-0.15
	(0.347)	*(0.347)*	*(0.312)*	*(0.312)*	*(0.391)*	*(0.044)*
Constant	-5.95	-5.95	-4.58	-5.09	-7.16	-0.15
	(1.545)	*(1.545)*	*(0.792)*	*(1.158)*	*(1.492)*	*(0.163)*
Number of observations	**12,239**	**12,239**	**14,934**	**14,934**	**9,524**	**14,934**
MSAs	**109**	**109**	**134** [b]	**134** [b]	**110** [c]	**134** [b]

Notes:
[a] Standard errors in parenthesis; [b] Includes MSAs w/o CITY SHAPE data); [c] Excludes cities which have rail transit; [d] Dependent variable for equation #6 is **ln(VMT)**

Table 4-7. Calculated DTW elasticities (from regression 1)

Variable	Atlanta	Boston	Chicago	Houston	Phoenix	San Diego	Entire Sample
City Area	0.30	0.10	0.11	0.31	0.54	0.24	0.13
Pop. Density	0.07	0.21	0.42	0.08	0.03	0.08	0.22
Job-Housing Gini	0.08	0.08	0.08	0.05	0.07	0.08	0.08
Median commute speed	0.88	0.84	0.71	0.91	0.82	0.92	0.83
City shape	0.06	0.17	0.10	0.17	0.09	0.08	0.15
Relative income	0.38	0.38	0.38	0.38	0.38	0.38	0.38
Age of reference person	-0.33	-0.35	-0.35	-0.34	-0.35	-0.34	-0.35

Table 4-8. Estimated VMT elasticities

Variable	Atlanta	Boston	Chicago	Houston	Phoenix	San Diego	Entire Sample
Log of income	0.18	0.18	0.22	0.18	0.20	0.19	0.18
Log of gas price	-0.14	-0.14	-0.17	-0.14	-0.15	-0.15	-0.15
Population density	-0.01	-0.02	-0.04	-0.01	0.00	-0.01	-0.02
Access to transit	-0.09	-0.11	-0.19	-0.09	-0.09	-0.14	-0.11
Area of MSA	-0.01	0.00	-0.01	-0.01	-0.02	-0.01	-0.01
Cityshape	0.02	0.05	0.04	0.05	0.03	0.03	0.05
Highway density	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Distance to work	0.16	0.18	0.20	0.17	0.14	0.19	0.18

Table 4-9. Review of elasticities of VMT with respect to income and fuel prices

Part (a) – Elasticities with respect to income	
Authors / study	**Estimated Elasticity**
Dahl (Survey of literature)	0.23 to 0.60
Schimek	1.2 to 1.4
Goodwin, Dargay and Hanly (U.K)	0.2
Kayser	0.48 to 0.26

Part (b) – Elasticities with respect to fuel price			
Authors / study	**Travel type**	**Short Run**	**Long Run**
Elasticity with respect to gas price			
Agras and Chapman	Total travel	-0.15	-0.32
Goodwin, Dargay and Hanly	Total travel	-0.10	-0.29
	Total travel (per vehicle)	-0.10	-0.30
Johansson and Schipper	Total travel		-0.3
	Total travel (per vehicle)		-0.20
Puller and Greening	Total travel		-0.7
Schimek	Total travel	-0.26	
De Jong and Gunn (Europe)	Commuting only	-0.12	-0.23
	Total travel	-0.16	-.026
INFRAS (Europe)	Total travel	-0.1 to -0.2	-0.25 to -0.5
Mayeres (Europe)	Essential trips	-0.16	-0.43
	Optional trips	-0.43	-0.36
Luk and Hepburn (Australia)	Total travel	-0.10	
Elasticity with respect to travel cost			
Parry and Small (Cost)	Total travel	-0.22	
Small and Winston	One vehicle households	-0.228	-0.279
	Two-vehicle households	-0.059	-0.099

Source: Based on Littman (2007)

Table 4-10. Mean city-level and young-household variables for six cities

Variable	Atlanta	Boston	Chicago	Houston	Phoenix	San Diego
City Characteristics						
Area	5.12	1.76	1.88	5.32	9.20	4.20
Pop. Density	0.55	1.63	3.22	0.62	0.23	0.59
Jobs-Housing Gini	0.42	0.41	0.46	0.27	0.40	0.41
Median commute speed	32.11	30.60	26.00	33.17	30.00	33.62
City shape	0.26	0.82	0.48	0.80	0.45	0.36
Ln(Gas Price)	-0.02	0.25	0.17	0.15	0.19	0.15
Mean young-household characteristics						
Household size	2.22	2.47	2.47	2.57	2.82	2.85
Number of children < 6	0.00	0.08	0.09	0.14	0.18	0.23
Age of reference person	25.30	26.06	25.79	25.86	27.18	24.23
Ln(Income)	3.66	3.63	3.73	3.58	3.62	3.03
Number of drivers	1.93	1.88	1.81	1.81	1.91	1.85
Number of adults > 65	0.00	0.00	0.00	0.00	0.00	0.00
Number of workers	1.67	1.81	1.84	1.71	1.55	1.85
DTW	6.05	7.01	8.48	7.19	5.45	8.60

Table 4-11. Predicted effect on *DTW* and *VMT* of moving a mean young-household from its city of origin to six other cities.

Part (a) – Yearly Distance-to-work						
A household that moves from the city below:	… to each of the following cities will locate *X* miles closer to work than in their city of origin. (Based on regression 1)					
	Atlanta	Boston	Chicago	Houston	Phoenix	San Diego
Atlanta	0	81	206	770	1037	61
Boston	-76	0	118	652	905	-19
Chicago	-193	-117	0	531	783	-136
Houston	-540	-485	-398	0	191	-498
Phoenix	-739	-683	-596	-194	0	-697
San Diego	-49	16	117	575	792	0

Part (b) – VMT						
A household that moves from the city below:	… to each of the following cities will drive *X* miles more than in their city of origin. (Based on regression 1)					
	Atlanta	Boston	Chicago	Houston	Phoenix	San Diego
Atlanta	0	-792	-2163	62	510	-675
Boston	863	0	-1489	930	1420	127
Chicago	2846	1870	0	2896	3462	1920
Houston	-37	-763	-2107	0	427	-726
Phoenix	-440	-1119	-2331	-405	0	-1084
San Diego	762	-41	-1511	803	1277	0

Table 4-12. **Distance to work, only households that commute by Car**

Variable	**#4** [(b)] **Coeff.** *(s.e.)* [(a)]	**#7 Coeff.** *(s.e.)*	**#8** [(c)] **Coeff.** *(s.e.)*	**#9 Coeff.** *(s.e.)*
Area of MSA (1000 square miles)	0.08	0.12	0.07	0.12
	(0.014)	*(0.066)*	*(0.018)*	*(0.066)*
Pop. Density in MSA	0.26	1.02	1.00	1.02
	(0.016)	*(0.079)*	*(0.069)*	*(0.079)*
Median speed in city	0.06	0.30	0.20	0.30
	(0.005)	*(0.024)*	*(0.019)*	*(0.024)*
Log of relative income	0.35	0.98	1.09	0.98
	(0.017)	*(0.082)*	*(0.077)*	*(0.082)*
Household size	0.10	0.18	0.14	0.18
	(0.011)	*(0.052)*	*(0.049)*	*(0.052)*
Number of children <6	-0.19	-0.52	0.06	-0.52
	(0.035)	*(0.167)*	*(0.159)*	*(0.167)*
Number of children 6-18	-0.17	-0.43	-0.34	-0.43
	(0.015)	*(0.072)*	*(0.069)*	*(0.072)*
Age of reference person	0.00	0.00	-0.03	0.00
	(0.001)	*(0.004)*	*(0.004)*	*(0.004)*
Constant	5.98	-4.57	-0.10	-4.55
	(0.150)	*(0.717)*	*(0.634)*	*(0.717)*
Number of observations	**14,934**	**13,967**	**13,967**	**13,967**

Notes:
[(a)] Standard errors in parenthesis; [(b)] Regression #4 from table 4-4; [(c)] Assumes no correlation between DTW and VMT equations

Table 4-13. VMT regression, only households that commute by car

Variable	#4 [b] Coeff. *(s.e.)* [a]	#7 Coeff. *(s.e.)* [a]	#8 Coeff. *(s.e.)*	#9 Coeff. *(s.e.)*
Log of income	4.09	4.28	4.25	4.29
	(0.185)	*(0.196)*	*(0.195)*	*(0.196)*
Log of gas price	-3.45	-4.36	-4.39	-5.05
	(1.714)	*(1.793)*	*(1.794)*	*(1.553)*
Population density	-0.51	-0.29	-0.31	-0.44
	(0.170)	*(0.180)*	*(0.179)*	*(0.165)*
Transit	-3.78	-3.62	-3.59	-
	(1.080)	*(1.128)*	*(1.129)*	
Area of MSA (1000 square miles)	0.01	0.00	-0.01	-
	(0.043)	*(0.045)*	*(0.045)*	
Highway density	-0.03	-0.02	-0.02	-
	(0.095)	*(0.112)*	*(0.112)*	
Number of drivers	(4.47	4.55	4.48	4.55
	(0.161)	*(0.170)*	*(0.166)*	*(0.170)*
Number of children < 6	0.24	0.44	0.46	0.45
	(0.344)	*(0.355)*	*(0.355)*	*(0.355)*
Number of children 6-18	-0.11	-0.09	-0.08	-0.09
	(0.125)	*(0.129)*	*(0.129)*	*(0.129)*
Number of adults over 65	-1.06	-0.97	-0.96	-0.96
	(0.371)	*(0.394)*	*(0.394)*	*(0.394)*
DTW x Number of workers	0.33	0.34	0.35	0.34
	(0.012)	*(0.014)*	*(0.011)*	*(0.014)*
Inverse mills ratio	0.15	0.56	0.61	0.57
	(0.558)	*(0.598)*	*(0.598)*	*(0.599)*
Constant	-4.04	-4.84	-4.78	-6.53
	(0.924)	*(0.976)*	*(0.976)*	*(0.758)*
rho	0.03	0.02	-	0.02
	(0.009)	*(0.011)*		*(0.011)*
Number of observations	**14,934**	**13,967**	**13,967**	**13,967**

Notes:
[a] Standard errors in parenthesis; [b] Regression #4 from table 4-4; [c] Assumes no correlation between DTW and VMT equations

Table 4-14. Calculated DTW and VMT elasticities, only households that commute by car

Part (a) – Elasticities for DTW equation					
Variable	**Variable means**	**Elasticities**			
		Reg. #4	**Reg. #7**	**Reg. #8**	**Reg. #9**
City Area	2.79	0.22	0.33	0.20	0.33
Pop. Density	0.81	0.21	0.82	0.81	0.82
Median commute speed	30.50	1.83	9.30	6.19	9.29
Relative income	0.10	0.35	0.98	1.09	0.98
Age of reference person	43.06	-0.35	-0.14	-1.45	-0.14

Part (b) – Elasticities for VMT equation					
Variable	**Variable means**	**Elasticities**			
		Reg. #4	**Reg. #7**	**Reg. #8**	**Reg. #9**
Income	51.02	0.19	0.201	0.199	0.201
Gas price	1.18	-0.16	-0.204	-0.206	-0.236
Population density	0.81	-0.02	-0.011	-0.012	-0.017
Access to transit	0.50	-0.09	-0.085	-0.085	-
Area of MSA	2.79	0.001	0.000	-0.001	-
Highway density	2.45	-0.003	-0.002	-0.002	-
Distance to work	10.69	0.17	0.168	0.178	0.169

5 The case of car ownership

5.1 Introduction

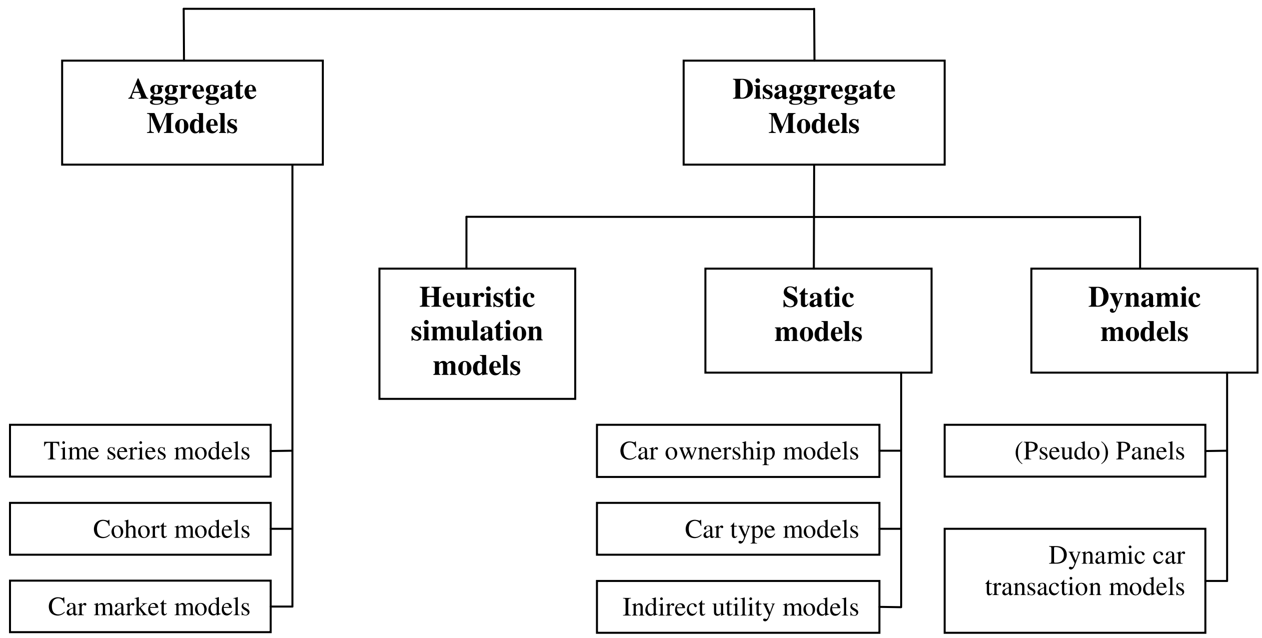
Chapter 4 of this dissertation analyzed the interaction between distance to work and travel demand. In particular, it looked at the effect of a household's distance-to-work (DTW) on its total annual vehicle miles traveled (VMT). Vehicle ownership was not modeled explicitly and the VMT equation was a reduced form equation reflecting factors affecting both cars owned and number of miles driven. In this Chapter 5 I extend the model developed in the previous chapter to analyze vehicle ownership together with distance to work and travel demand. In this sense, the objective of Chapter 5 is to analyze the effect that urban form and distance to work have on demand for miles once vehicle ownership is made endogenous.

The regression analysis in this section is, like in the previous chapter, a short to medium term analysis, as demand functions are estimated for given levels of distance to work. Note also that individuals can change the number and types of cars they own in a short period of time, therefore they can adapt relatively quickly to changes in exogenous shocks. Consider the example of rising gasoline prices and travel demand—consumers will eventually find ways to conserve their fuel use. However, some alternatives to reducing gasoline consumption, like finding a more fuel-efficient car, take time. In the short run, individual may adapt to price shocks by switching away from private vehicles into less expensive forms of transport (i.,e. public transport). But in the long term, individuals will adjust their job or housing location or the efficiency of their vehicles (i.e. change vehicle type) to adapt to these changes. This example is supported by empirical results, as researchers have concluded that the short term travel demand (measured by annual VMT) elasticity with respect to gasoline price is around -0.1, while the long-run elasticity is about -0.3 (see section 5.4.3 and Table 4-9).

Vehicle ownership and use (mileage, number of trips) has been studied significantly in the past. Bhat and Sen (2006) outline two important reasons for the amount of research in this area: first, car manufacturers are interested in understanding household's preferences for different kinds of vehicle types in order to obtain information that can allow them to target the market in a more strategic way. Second, understanding vehicle holdings and use is key for policy makers, as it has serious implications from a congestion and pollution perspective.

De Jong et al (2004) present a recent survey of the literature of vehicle ownership. In general, studies on vehicle ownership can be classified according to the data used. Models using aggregate data are usually at the national level and are based on time series (sometimes disaggregation is done by cohorts). These studies are typically used to estimate gasoline consumption demand (and pollution) and traffic demand. Examples of these types of studies include Ingram and Liu (1997), Whelan et al. (2000), Whelan (2001) and Dargay and Cately (1999). Models based on disaggregate data (household data) can be static (Train, 1985, Whelan, 2001, and Rich and Nielsen, 2001, Bento et al., 2005) or dynamic (Mannering and Winston 1985).

Figure 5-1. Classification of Car ownership models by type of data



Source: Based on classification proposed in de Jong et al (2004)

Static and dynamic models of car ownership typically focus on analyzing the determinants of car ownership. In these types of models, a household's or individual's inclination to own vehicles is linked to its socio-economic characteristics, the costs of owning and driving a car, the availability of alternative means of transport, and more recently due to advances in GIS, its locational characteristics (for example Bento et al. 2005). Early studies in the field tended to be cross-sectional (static) and focused mostly on car ownership levels (Lerman and Ben-Akiva, 1976 and Train, 1980, Bhat and Pulugurta, 1998). As discrete choice methods in econometrics advanced (i.e. nested logits), researchers started exploring the choice of car type of the household given car ownership levels (Berkovec, 1985; Mannering and Winston, 1985; Train, 1986; de Jong, 1990; and Hensher et al, 1992).

Later, as Whelan (2005) points out, increases in traffic congestion and advances in fuel efficiency led researchers to explore the interrelated choice of car ownership and use, largely based on the framework of continuous-discrete models. This framework is typically used to analyze the interaction between transportation and land use, and includes an equation to model of the number of cars owned by a household and an equation to model of the annual demand for miles (VMT). Different measures of land use enter the model as explanatory variables and allow researches to measure their impacts both on car ownership and travel demand. Steiner (1994), Wilson (1998), and Badoe and Miller (2000) present recent surveys of the literature on the interaction between land use and transportation. In this literature, car ownership and VMT equations are jointly estimated, using continuous discrete methods, as described in Train (1986). First, households select how many cars to own. Second, conditional on vehicle ownership, the household decides how much to use each car. Though decisions are modeled sequentially for estimation purposes, the choices are made simultaneously as they come from the same utility maximization problem (see Train, 1986; chapter 5). Examples of papers analyzing the interaction between car ownership and VMT include Mannering and Winston (1985), Train (1986), de Jong (1990), Berkowitz et al. (1990), Hensher et al (1992), Kockelman (1997), Linciano (1997), Choo and Mokhtarian (2004), West (2004), and Bento et al (2005).

Finally, dynamic models follow a similar structure as static models, except that instead of having a cross-section of data, researchers have either a panel or pseudo panel of data. Models incorporate this additional source of information using standard estimation techniques (i.e. fixed or random effects). Examples include Kitamura (1987), who used 10 waves of the Dutch National Mobility Panel to determine simultaneously car ownership and number of trips per week. Nobile et al (1996) used panel data to estimate a random effects multinomial probit of car ownership levels. More recently, Hanly and Dargay (2000) use a panel to analyze vehicle ownership in Great Britain

A common feature in the studies on demand for transportation is that role of household location is typically not recognized and not included in the estimation. In general, VMT is modeled as a function of demographic characteristics (household size, income, number of workers in the household, among others), car characteristics (cost per mile) and land use measures. Among the land use measures, population density of the individual's neighborhood is sometimes included as an explanatory variable, although without considering its correlation to unobservable variables affecting VMT (e.g. Schimek, 1997). Some papers like Bento et al. (2005) highlight this potential problem of endogeneity and use measures of density that are truly exogenous to the household—density at the city level instead of the local level. I follow this approach and include variables such as density and access to transit at the city level, as discussed in Section 4.3.3. This approach allows me to compare the effect of these important variables by studying the differences in density and access to transit among different MSAs. Though this is not the best approach, there is still a good deal of information that can be learned from following it.[14]

[14] Ideally, one should model each of these endogenous decisions explicitly, but models would become untractable due to the large amount of dependent variables.

5.2 Data

I use a slightly different sample of the NPTS than that used in Chapter 4. The two main differences are: (i) I include households that own no vehicles in the CARS equation, but these households are dropped in the VMT equation because their VMT=0; and (ii) I exclude households with no workers. Households with no workers are excluded because I am interested in studying the effect that distance to work has on travel demand conditional on car ownership. Table 5-1a shows the sample size used in the estimation.

The original sample size of the 1995 NPTS is 42,033 households. As shown in Table 5-1a, a total of 23,446 households were dropped from the sample because (i) there were no workers in the household, (ii) household is not in an MSA or is located in an MSA with less than 20 observations, and (iii) there are missing values in the explanatory variables, including no VMT data. The CARS equation was estimated using a total of 18,587 observations. This includes households in the 134 largest metropolitan statistical areas (MSA) that had at least 20 observations. The threshold of 20 observations was chosen randomly but I feel it is a large enough number to permit enough variation in the calculation of the MSA wide measures.

The NPTS asks households to indicate the number of cars in the household, and then proceeds to collect detailed information on each of the vehicles (make, model, and year, odometer readings, principal driver). As a result, households reported car ownership levels ranging from no vehicles to 10 cars. Four categories of car ownership were created as follows (see Table 5-1b): households with no vehicles

(6.65 percent of the sample), 1-car households (27.45 percent), 2-car households (46.61 percent), and households owning 3 or more vehicles (19.27 percent).

To estimate the VMT equation an additional 1,237 observations are lost because the household owns no cars (1,237 observations). As a result, a total of 17,350 observations are used to estimate the three VMT equations. See Table 1a for a description of number of observations used in each of the equations. As is described later in this paper (section 3.2), three VMT equations are estimated separately depending on whether the household owns 1 car (4,191 observations), 2 cars (7,042 observations), or 3 cars (2,917 observations).

The NPTS is merged with data from other sources, as described in Chapter 3.

5.3 Econometric Model

5.3.1 Dependent Variables

Car ownership (CARS). The NPTS collected data on the number of cars owned by each household. I define a new variable, car ownership (CARS), as a categorical variable based on the number of cars owned by each household. The categories are 0 (No cars), 1 car, 2 cars, and 3 or more cars in the household. Over 90 percent of the households in the sample own at least one car, with the biggest portion of households owning two cars. Table 5-1b shows the distribution of the categorical variable—6.7 percent of the households in the sample own no vehicles, while 27.4, 46.5, and 19.3 percent of the households own 1-, 2, and 3 or more vehicles, respectively.

Vehicle Miles Traveled (VMT). The variable representing annual vehicle miles traveled (VMT) is built as described in the Chapter 4: as the sum of miles over all cars in the household.

5.3.2 Model and Discussion

Our general model is a system of 2 equations, one continuous and one discrete. The two equations are:

$$Prob\ (CARS = i) = Prob\ (U_i > U_j) \text{ for } i \neq j \text{ and } i = 0,1,2, \text{ or } 3 \quad (1)$$

where $U_i = \beta_{1i} Z_1 + \alpha_i DTW + \varepsilon_i$

$$VMT_i = \beta_{2i} X_{2i} + \delta\ HHDTW + v_i \qquad \text{for } i = 1,2, \text{ or } 3 \quad (2)$$

Equation 1 represents the vehicle ownership equation. The variable CARS is categorical in nature, represents the number of cars owned by each household, and ranges from 0 to 3 (a value of 3 stands for households owning 3 or more vehicles). Equation 2 represents household travel demand. As discussed before, I use an annualized measure of travel demand equal to the sum of vehicle miles traveled in all vehicles in the household. Since I am interested in exploring the effect of distance to work (DTW), this variable has been made explicit in the above model. HHDTW is measured as the sum of the annualized distance to work for each of the members in the household. The annualized distance to work is obtained by multiplying the one-way distance to work by 480.[15]

Note that equation (2) is really 3 different equations. The system above can be rewritten as follows:

$$Prob\ (CARS = i) = Prob\ (U_i > U_j) \quad \text{for } i \neq j \text{ and } i=0,1,2,\ or\ 3 \tag{1a}$$

where $U_i = \beta_i Z_1 + \alpha_i\ HHDTW + \varepsilon_1$

1-car households:

$$VMT_1 = \beta_{21} X_1 + \delta_1\ HHDTW + \gamma_1\ \ scf_1(P_i) + \upsilon_2 \tag{2a}$$

2-car households:

$$VMT_2 = \beta_{22} X_2 + \delta_2\ HHDTW + \gamma_2\ \ scf_2(P_i) + \upsilon_3 \tag{3a}$$

3-car households:

$$VMT_3 = \beta_{23} X_3 + \delta_3\ HHDTW + \gamma_3\ \ scf_3(P_i) + \upsilon_4 \tag{4a}$$

The error structure is given by:

$$\Omega = E([\varepsilon_1, \upsilon_2, \upsilon_3, \upsilon_4][\varepsilon_1, \upsilon_2, \upsilon_3, \upsilon_4]')$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ & \sigma_2^2 & 0 & 0 \\ & & \sigma_3^2 & 0 \\ & & & \sigma_4^2 \end{bmatrix} \tag{5a}$$

[15] It is assumed that households work a total of 240 days a year (5 days per week for 48 weeks). The one-way distance to work is multiplied by 2 to simulate the total daily commute in miles.

The model above is described in the literature as a *continuous-discrete model.* Its estimation procedure is also well documented in the literature (see Train, 1986, Chapter 5). Ideally, since the household decisions of how many cars to own and how much to drive each vehicle are made simultaneously, estimation of the system of equations (1) and (2a)-(2c) would be done using full information maximum likelihood methods. But in practice this approach is not used because it is difficult to make such a complicated model converge. Instead the model is estimated in three steps, as follows: .

i. First, estimate the car ownership equation and obtain the predicted choice probabilities, $\hat{P}_i$

ii. Second, estimate the selectivity-correction term, $scf(\hat{P}_i)$. Dubin and McFadden (1984) have shown that when the discrete choice probability is distributed log-weibull and the continuous variable is distributed normal, then the selectivity correction factor in equations 2(a) to 4(a) takes the form:

$$scf_i(\hat{P}_i) = \sum_{j \neq i} \left[\frac{\hat{P}_j \cdot \ln \hat{P}_j}{1 - \hat{P}_j} + \ln \hat{P}_i \right]$$

iii. Finally, estimate equations (2a) to (2c) including the term $scf(\hat{P}_i)$ is included. Note that an equation of VMT on its regressors is estimated for each category of car ownership. In my particular case, I estimate three different equations. Parameter estimates obtained following this procedure are consistent (Train, 1986).

Exogenous vs. endogenous DTW. Estimation of the model described in (1) to (2c) assumes that DTW is exogenous. As basic econometric textbooks show, the inclusion of an endogenous variable as an explanatory variable in an equation leads to biased estimates. In this case I face a tradeoff between obtaining biased estimates and the tractability of the model. In this chapter, I make the choice of using DTW as an explanatory variable because chapter 4 shows it is an important variable explaining travel demand. Not only is this variable important (income elasticity of VMT increases by 10% when DTW is omitted), but it also allows us to estimate the contribution of commute miles to total miles. Finally, by assuming that household location is exogenous, I can focus my attention on the relationship between car ownership and travel demand, and the econometric complications of estimating this kind of system.

5.3.3 Explanatory Variables

5.3.3.1 CARS equation

I focus on static car ownership models, as classified by Jong et al. (2004) and described in Section 2. Recent studies analyzing the interrelated choice of vehicle ownership and use include Mannering and Winston (1985) Train (1986), Goldberg (1998), West (2004), Bento et al. (2005), and Bhat and Sen (2006). In these models, and following the work by Dubin and McFadden (1980) and Heckman (1984), the car ownership and VMT equations are interrelated, and usually appear in nested form.

A review of the literature shows that the main variables explaining the number of cars owned are: the household's income, the number of workers, the cost of owning an automobile, and recently, different measures of land use, including density and the availability of public transit. I use these variables and others, as described below.

Income. Income has been consistently found to be significant and positively correlated with car ownership in the literature. It is one of the most important variables explaining car ownership as buying a car requires a significant investment.

Number of workers. A second important variable explaining car ownership is the number of workers. Given a household location (i.e. DTW as well as transit availability and access), as the number of workers (NUMWORK) in the household increases, more cars will be needed for the commuters. Therefore car ownership will increase as number of workers increase in the household.

Price of gas. The literature often includes a measure of the costs of owning a vehicle, typically the purchase price or an annualized user cost. Annualized user costs include costs of insurance and fuel expenses. All these costs depend on the type of car (make and model) that the household owns. Bigger and more "powerful" cars typically have higher purchase prices, insurance costs, and fuel expenses. As in the chapter 4, I use state-level gas prices (GASPRICE). Gas price is selected over annualized user costs in order to compare my estimates to the literature.

Distance to work (HHDTW). The relationship between car ownership and household location has been taken into consideration by Waddell (1996) and Sermons (2000), but in a context quite different from the one I am currently using: these authors used number of cars as an explanatory variable in their (discrete) household location models.

City characteristics. Car ownership will be affected by measures of city form that describe access to jobs or other facilities (schools, doctor, shopping, transit). I proxy accessibility to jobs and other facilities with population density (POPDENSITY) at the city level. I cannot use population density at the zipcode or census tract level because this would be an endogenous choice. Population density at the city level is a very rough measure of accessibility but is closer to being truly exogenous.

I complement population density and access to transit with other measures developed by Bento et al (2005) and described in Chapter 4. These include city shape (CITYSHAPE), and city area (AREA). In addition, a variable describing highway density at the city level (HWY_DENS) is also included in the regression. Households living in cities with higher highway density are expected to own more cars, ceteris paribus.

Household characteristics. Other variables that affect car ownership and have been frequently included in previous studies are age and education level of the reference person as well as variables that represent the life cycle. Households that have children, for instance, may tend to own more cars to take them to extracurricular activities or parks over the weekends. I include variables NUM_KID6 to represent the number of children under 6 in the household, and NUM_OLD to represent the number of household members over 65.

Finally, number of drivers (DRVRCNT) in the household is also an important variable that affects car ownership. By number of drivers the NPTS refers to individuals that know how to drive. It does not necessarily imply having a drivers license.

5.3.3.2 *VMT equation*

Variables used in the VMT equation are discussed in section 4.3.3. I mention them here for clarity, but the rationale for their inclusion is not described in this section.

Income is a key variable in explaining vehicle ownership and thus widely used in explaining VMT. For the VMT regressions I use absolute, not relative, income since most of the purposes of non-commute VMT involve purchase of items that are tradable across cities, unlike houses. Regarding a measure of cost of travel, I use the truly-exogenous state-level gas prices. Other important household-level variables are the number of drivers, number of children, and the household's life cycle.

Table 5-2 contains the means of the variables used in the above-mentioned models.

5.3.4 Research Questions and Hypotheses

My main questions concern (i) the effect of distance to work on car ownership (ii) the effect of distance to work on VMT, conditional on car ownership, including an analysis of the task sharing implied by this relationship, (iii) the effect of city characteristics on car ownership levels, conditional on household location, (iv) the effect of city characteristics on VMT, conditional on household location and vehicle ownership, and finally

(v) does the model of car ownership change the relationship between VMT and DTW?.

5.4 Results

5.4.1 Car ownership model

Table 5-3 shows the results for the multinomial logit regression. The dependent variable is the number of vehicles owned by the household. The base category for the regression is "household does not own any cars". This means that the coefficients are all relative to this category.

Regressions #1 to #4 in Table 5-3 show the effect of different variables on car ownership, in order to analyze the robustness of the estimated parameters. Regression #1 is the main equation analyzed, and the results described below are based on this regression.

The results from regression #1 show that an increase in household income by 10 percent will increase the log-odds of owning three cars over no cars by 3.49, the log-odds of owning two cars over none by 3.21, and the log-odds of owning one vehicle over none by 1.66. The variables "number of drivers" and "age of household head" also have positive effects on number of cars: increasing the number of workers and the age of the household head increases the log-odds of owning one, two, or three vehicles over owning no vehicles. Furthermore, the log-odds of owning more vehicles increase as the number of workers in the household increase. The presence of an additional worker in the household increases the log-odds of owning 2 and 3 vehicles by 1.01 and 1.94, respectively, but reduces the log-odds of owning one vehicle by 0.04.

In general, when dealing with multinomial logit models it not straightforward to interpret the coefficients arising from the regression since these represent the log of the "odds-ratio". It is easier to analyze the results in terms of the partial effects or the differences in probabilities. The partial effects can be calculated (following Greene, 2005) according to the following formula:

$$\frac{\partial P(y=i\,|\,x)}{\partial x_j} = P(y=i\,|\,x)\left\{\beta_{ij} - \frac{\sum_{h=1}^{H}\beta_{hj}\exp(x\beta_h)}{1+\sum_{h=1}^{H}\beta_{hj}\exp(x\beta_h)}\right\}$$

Even though estimation of the partial effects is not difficult, it is easier to interpret the results if these are described in terms of the difference in probabilities. Therefore I follow this latter approach in the ensuing discussion. Following Wooldridge (2002, pg. 499), define the predicted probability in situation *t*, where *t* equals 0 or 1 depending on whether it is before or after the change in the variable of interest, is given by:

$$P_t(car_i) = \frac{\exp(X_t\hat{b}_i)}{1+\sum_{j=1}^{3}\exp(X_t\hat{b}_j)} \qquad \text{, for } i = 1, 2 \text{ or } 3; \text{ and } t = 0 \text{ or } 1$$

where i denotes the number of cars owned by the household, t represents the "with" ($t=0$) and "without" ($t=1$) cases, and $\hat{b}_i$ are the estimated coefficients from the multinomial logit regression. The predicted probability of owning no cars is given by

$$P_t(car_0) = \frac{1}{1+\sum_{j=1}^{3}\exp(X_t\hat{b}_j)}$$

The difference in the probability of owning *i* cars is as:

$\Delta P(car_i) = P_1(car_i) - P_0(car_i)$.

For instance, given an average household income of $40,000, the predicted probability of owning 2 cars, $P_{t=0}(car_2)$, equals 61.8 percent (see Table 4a). If the average household's income were to increase to $60,000, the new predicted probability, $P_{t=1}(car_2)$, increases to 67.3 percent. Therefore by increasing income from $40,000 to $60,000, the probability of owning 2 cars is increased by 5.49 percentage points. Following this same logic, one finds that the predicted probability of owning 3 cars increases by 4.67 percentage points, and that the probability of owning 0 or 1 car decreases by 0.6 and 9.56 percentage points after this increase in income.

The analysis in the remainder of this section is based on the logic described in the paragraph immediately above. Table 4b shows that increasing the number of workers in the household from 2 to 3 increases the probability of owning 3 cars by 18.6 percentage points, while the probability of owning 1 car decreases by 18.0 percentage points. At the same time, the probability of owning 0 and 2 cars is practically unchanged. The probability of owning no cars decreases by 0.8 percentage points, while the probability of owning 2 cars increases by 0.2 percentage points.

Since income, number of workers, and distance to work are the main variables of interest for explaining vehicle ownership, the marginal effect analysis is extended for these three descriptors. The change in probabilities due to a change in income was calculated starting from a base income of $40,000 (the average household income in my sample) and varying income from $10,000 to $100,000 using a step of $5,000. The results are shown in Graph 1 below. For example, if income decreases from $40,000 to $20,000, the probability of owning 0 and 1 cars increases by 4 and 15 percentage points, respectively, while the probability of owning 2 and 3 cars decreases by 12 and 6percentage points, respectively.

Figure 5-2. Car ownership – marginal effects from changes in income

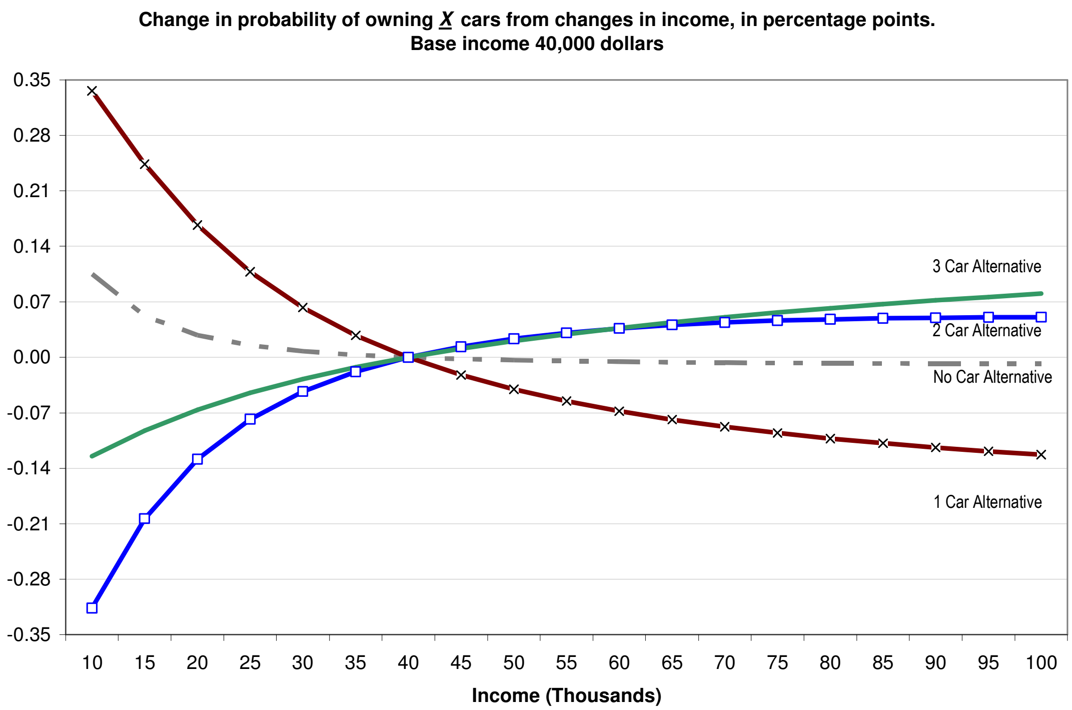
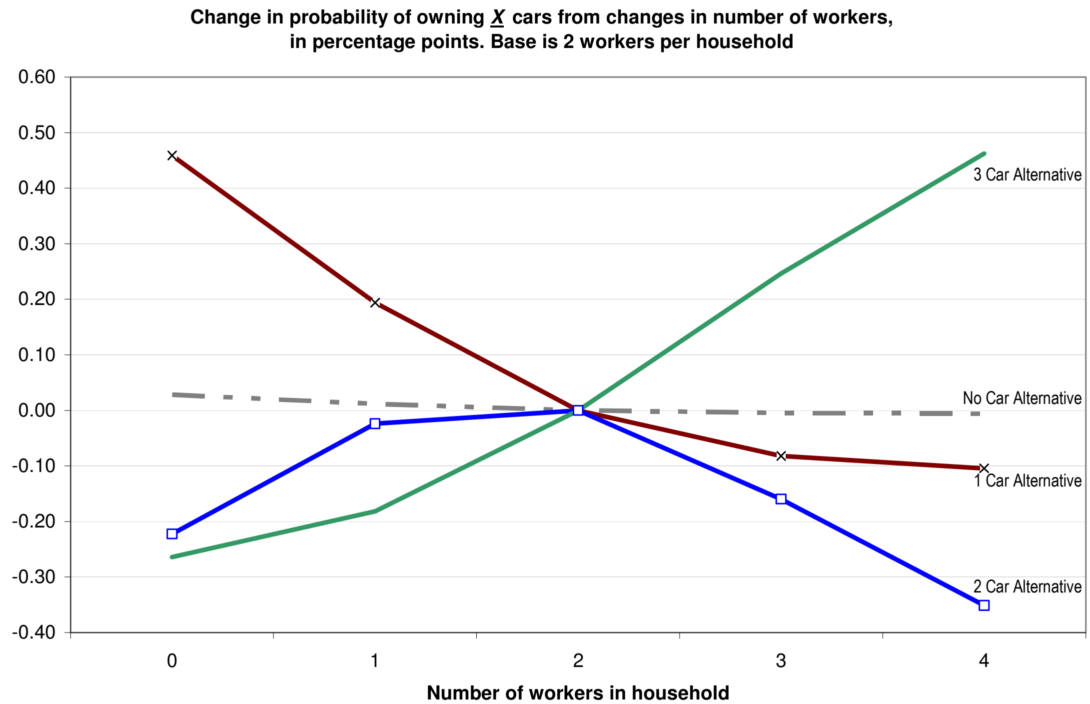
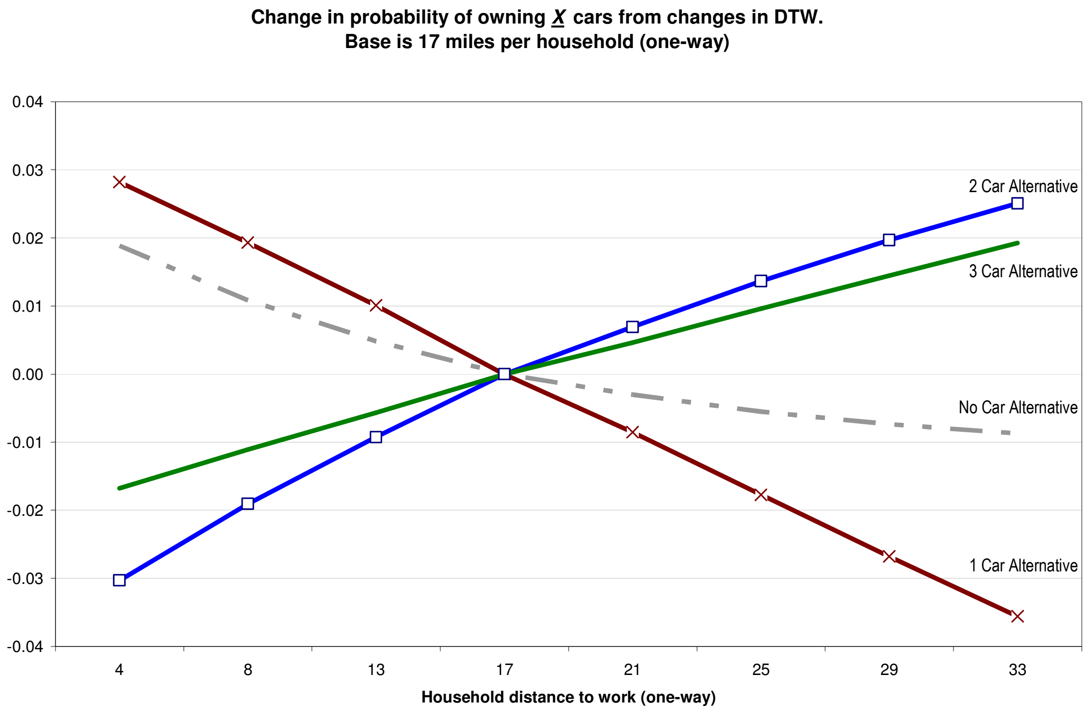


Figure 5-3. Car ownership – marginal effects from changes in workers



Graph 2 above shows the relationship based on variation in the number of workers in the household. Starting from a base of 1 worker, the probability of owning 3 cars increases substantially as number of workers in the household increase. When the number of workers in a household changes from 2 to 4, the probability of owning 3 vehicles increases by a dramatic 45 percentage points. On the other hand, if the number of workers decreases from 2 to 1, the probability of owning 1 vehicle increases by almost 18 percentage points, while the probability of owning 3 decreases by the same 18 percentage points.

Figure 5-4. Car ownership – marginal effects from changes in DTW



Distance to work has a small effect on car ownership (see Graph 3, above). Recall that the household's total one-way distance to work is estimated as the sum of the one-way distances of each of the working member. For instance, if there are 2 working members in particular household, each living 5 miles from their work, the total one-way distance to work for this household is 10 miles. Increasing the average household's one way distance to work from 17 to 25 miles per household, increases the probability of owning 2 and 3 cars by 2 percentage points each.

<u>City characteristics.</u> My results show that increasing population density by 10 percent decreases the probability of owning 2 and 3 vehicles by 7.1 and 8.8 percentage points, respectively. At the same time, the probability of owning 0 and 1 vehicles increases by 2.2 and 13.7 percentage points after the above-mentioned increase in density. Similarly, increasing the number of households living within half

a mile of a transit stop (i.e. providing better access to transit) by 10 percent decreases the probability of owning 2 and 3 vehicles by 2.93 and 0.6 percentage points, respectively, while it increases the probability of owning 0 or 1 car by 0.3 and 3.2 percentage points. Finally, and surprisingly, increasing highway density in a city decreases the probability of owning 3 cars by 0.6 percentage points, while increasing probability of owning 1 and 2 cars by 0.4 and 0.2 percentage points, respectively. The effect of increasing highway density on car ownership is very small.

5.4.2 Predicted car ownership levels for 6 cities

To further understand the effect of the explanatory variables on car ownership, I carried out the following analysis: I predicted the effect on car ownership of moving a household from its current city of residence to other cities in the US. For presentation purposes I use the following 6 cities: Atlanta, Boston, Chicago, Houston, Phoenix, and San Diego. These represent cities from most of the census regions and have different growth histories and patterns. Table 3.b presents a summary of the results. These results are based on regression #1.

To understand the results in Table 3.b, suppose that a household moves from Atlanta (column 1) to Chicago (column 5). The table below describes the average household in both of these cities, as well as the city measures used in the regressions. The average household in Atlanta has an annual income of $46,500, has 1.4 workers, 0.6 children below the age of 6, 0.36 kids 6 to 18 years old, and 0.27 adults over the age of 65. The reference person is 45.7 years old and has 13.5 years of education (see Table 5-2b).

If the average Atlanta household moves to Chicago, it will tend to own fewer vehicles than it does now. The probability of owning 2 and 3 vehicles is predicted to decline by 12.8 and 9.6 percentage points, respectively, while the probability of owning 0 or 1 cars increases by 2.1 and 20.3 percentage points. To put this another way, an average household in Atlanta has a 20, 62, and 18 percent probability of owning 1, 2, and 3 cars, respectively. If this same household were to move to Chicago, its probability of owning 1 vehicle would jump to 40 percent, while its probability of owning 2 and 3 cars would decrease to 49 and 8 percent, respectively. Table 3.b shows that if an average household from Chicago or Boston moves to any of the other cities, it will have a greater probability of owning 2 or 3 cars. The opposite happens when an average household from Atlanta, Houston, Phoenix or San Diego moves to Boston or Chicago: the probability of it owning 0 or 1 cars increases substantially.

The general conclusion of this exercise is that households that move to higher density cities, with better transit access, and higher road density will tend to own less cars. In the particular example of a household moving from Atlanta to Chicago, the difference in the population density between the cities is the main force leading the results. This result is important from a policy perspective, as households with less cars drive less miles per year. Section 5.4.4 expand on this finding be extending the effect of changes in city characteristics to VMT.

5.4.3 Demand for miles conditional on Car Ownership

Tables 5-5 to 5-8 show the results for the VMT regression. As described in the estimation procedure, the sample is broken down by number of vehicles owned. A separate OLS regression is run for households owning 1, 2 and 3 vehicles. Equations 1 to 3 in Table 5-5 are estimated using the true value of distance to work in the DTW variable. In other words, equations 1 to 3 assume that DTW is exogenous or at least uncorrelated with other right-hand-side variables.

Income. Income elasticity varies between 0.19 and 0.13 for households owning 1 and 3 vehicles, respectively, holding number of vehicles constant.[16] In other words, increasing household income by 10 percent would increase VMT by 13 to 19 percent conditional on current vehicle ownership. Since the average income of households owning one vehicle is approximately $24,000 and that they drive 10,800 miles per year, increasing household income by $2,400 dollars would imply that one-vehicle households would drive an average of 2,052 miles more per year. On the other end, the average income of households owning 3 is approximately $51,400 dollars, so increasing household income by $5,100 would increase travel demand by three-vehicle households by 3,890 miles per year.

Two issues should be highlighted from these results. First, my results are in the lower end of those found in the literature (see Table 4-9)—BCMV find that income elasticity for 1-car households is around 0.3, while for 2-car households the elasticity is about 0.15—I find that my estimated elasticities for 1-car households is almost half theirs. My elasticity for 2-car households is 0.15, very similar to what

---

[16] For the VMT equation, the elasticity is $\varepsilon_{\ln X1} = [\partial VMT/\partial \ln X_1]\cdot[1/VMT] = \gamma_1 \cdot [1/VMT]$ or $\varepsilon_{X2} = [\partial VMT/\partial X_2]\cdot[X_2/VMT] = \gamma_2 \cdot [X_2/VMT]$. Elasticities are calculated at the mean value of the variable for each level of car ownership.

they found. Second, income elasticity is higher for households owning 1 car than those owning 2 or 3 cars. The latter is an expected result, as households with more cars are closer to a point of satiation (there's so many cars one person can use at a time).

Effect of DTW on VMT. The regression of VMT on DTW (Table 5-5) allows me to calculate the contribution of commuting miles to total miles driven. It is easier to analyze the effect of DTW on VMT by looking at the marginal effect instead of the elasticity. The estimated marginal effects are 0.32, 0.35, and 0.36 for 1-, 2- and 3-vehicle households, respectively.[17] Given that the marginal effects are less than one, it is easiest to think of this as a strong degree of task-sharing, as discussed in section 4.5.4. Section 5.4.4 analyses the overall effect of HHDTW on VMT, conditional on car ownership.

Gas price. Several authors review the price elasticity of vehicle miles traveled, including Litman (2007), Goodwin, Dargay and Hanly (2003), and de Jong and Gunn (2001). Gasoline price elasticity of travel demand is relatively inelastic in the short run, with values ranging between -0.1 and -0.26. In the long run, as households account for higher prices in their decisions such as the type of car to own and where to live, gas price elasticities are higher. In the literature, estimates for the long run price elasticities range from -0.29 to -0.32. See Table 4-9 for a summary of elasticities found in these reviews and the literature in general.

[17] The elasticity of VMT with respect to DTW is estimated as follows:

$$E_{DTW} = [\partial VMT/\partial DTW]\cdot[DTW/VMT]\ [NUMWORKERS]$$
$$= [\delta\ NUMWORKER\ ] \times [DTW/VMT]$$

Recall that the household's distance to work is estimated as (average) DTW x NUMWORKERS.

The gas price elasticities obtained in my analysis are -0.041, -0.078, and -0.18 for 1-, 2- and 3-vehicle households, respectively. These elasticities are around the short term elasticity found in the literature of -0.1. Note also that elasticities increase as vehicle ownership increases. Since VMT increases as car ownership increases, the effect of the elasticity is magnified. Households owning 1 vehicle travel on average 10,000 miles per year, while 3-vehicle households travel around 28,000 miles per year. The elasticity of -0.04 for one-car households implies that their annual miles will decrease by 40 miles per year if gasoline prices increase by 10%. On the other hand, 3-vehicle households will reduce their travel demand by approximately 504 miles per year as the price of gas increases 10%. This is an expected result as households with more vehicles can adapt easier to price increases, for instance by car sharing—this is, households may decide to travel together (in one vehicle) instead of having two individuals making discretionary trips in response to the increase in the price of gas.

Finally, recall that the gasoline price elasticity of travel demand obtained in Chapter 4 was around -0.28. Note that my current estimates for this elasticity conditional on number of cars owned range from -0.04 to -0.18. Both of these results agree with the existing literature, where long the term elasticity is around -0.3 while the short term elasticity is approximately -0.1. As discussed above, by taking car ownership into consideration in the analysis of travel demand, I am looking at a shorter term relative to the case in which I explore only household location. Therefore it is expected that the elasticities in this chapter will be smaller in magnitude than those obtained in Chapter 4.

City characteristics. The results show that Population Density and City Area have very small effects on VMT conditional on number of cars. I include population density because it represents, roughly speaking, the density of friends and (in most cases) other non-commuting destinations. I expect that this characteristic is the key determinant of non-work miles. The higher is the density, the lower should be VMT. The marginal effect for population density ranges from 0.1 to -0.3 percent for 2- and 3- vehicle households. In other words, increasing density will not produce significant reduction in travel demand in the short term. Similarly, for city area, the marginal effect ranges from -11% for 1 vehicle households to 4.7% for 3- vehicle households.

Household characteristics. The remaining variables are in line with what I expected (see Table 5.5). (i) An increase in the number of drivers leads to an increase in miles for all levels of vehicle ownership. (ii) an increase in the number of adults over 65 reduces miles for all levels of vehicle ownership, and (iii) the marginal effects associated to increases in the number of children under 6 and the number of kids between 6 and 18drivers is very small.

5.4.4 Overall effects.

Sections 5.4.1 and 5.4.2 show the marginal effects for the car ownership and travel demand models separately. By presenting results in the previous fashion, one can understand more clearly the effects of the explanatory variables. But many variables such as income, distance to work, population density, and number of workers affect both the number of cars a particular household will own and its travel demand. This section looks at the marginal effect of some key independent variable as a whole—this is, the total of their direct and indirect (through number of cars) effects on travel demand. The results described in the next paragraphs are shown in Table 5-8.

Consider a young household living in Atlanta. Assume this household is composed of four individuals: the reference household head, a spouse and two children. The household head is the only worker in the household. This individual earns an annual income of $40,000, and lives 11 miles away from work. The household head is 45 years old and has 16 years of education. Based on regression #1 in Table 5-3, the predicted probabilities of owning 0, 1, 2, and 3 households are 1, 20, 67, and 12 percent, respectively. Thehousehold is estimated to travel 18,076 miles per year.[18]

First, suppose that the household's income were to increase by 50 percent. As a result, the probabilities of owning 2, and 3 vehicles would increase by 5 and 2 percentage points, respectively, while the probability of owning 1 car decreases by 7 percentage points. The average number of cars for this household would increase from 1.88 to 2.00, and total VMT are expected to increase by 2,102 miles to 20,179 miles.

Second, suppose that the reference person's distance to work increases by 50 percent, from 11 miles to 16.5 miles. As a result, the probabilities of owning 2 and 3 vehicles increase by 1.3 and 0.5 percentage points, while the probability of owning 0 and 1 vehicles decreases by 0.7 and 1.1 percentage points. The average number of cars for this household would increase from 1.88 to 1.91, and total VMTs are predicted to increase by 1,120 miles per year to 19,197 miles per year. These results

[18] The average total VMT over all households for Atlanta is 20,320 miles per year.

imply a marginal (overall) effect of HHDTW on VMT equal to 0.42,[19] which turn suggests a significant amount of task sharing. Recall that in section 4.5.2.1 I had estimated this marginal effect to be around 0.33. By simultaneously estimating VMT and car ownership, we are obtaining a higher marginal effect—an expected result given the longer term analysis that the model in chapter 5 represents. By allowing DTW to change, households' respond by changing the number of cars they own. Households need additional mobility to be able to meet their needs. As a result, the level of task sharing is reduced compared to chapter 4.

Third, suppose that a second household member gets a job, but the household's (average) distance to work does not change. This can occur if the spouse gets a job very close to home or if the family moves to a different home so that, even though both workers will be closer to their jobs, the household's overall distance to work does not change. In this case, the probability of owning 3 cars increases by 16.8 percentage points, while the probabilities of owning 0-, 1- and 2 cars decreases (?) by 1.4, 13.1 and 2.4 percentage points, respectively. Even though the probability of owning 3 vehicles increases by such a large amount, the household's VMT are predicted to increase only by 1,744 miles to 19,821 miles. This relatively small change is driven by the fact that I have artificially kept the household's overall distance to work constant. If, for example, the second household member got the job close to that of the reference person's, the predicted VMT would increase by 3,750 miles to 21,826.

[19] An increase of 5.5 one-way miles is equivalent to increasing 2640 annual miles. Since the overall effect was to increase VMT by 1120 miles, the implied ratio of VMT to HHDTW is 1120/2640=0.42.

Finally, suppose that the reference household in Atlanta moves to Chicago. Assume also that the reference person's income does not change and that he locates also 11 miles from work. This household's probability of owning 2 and 3 cars decreases by 17 and 6.7 percentage points, while the probability of owning 0 and 1 vehicles increases by 8.4 and 15.3 percentage points. These changes are due mostly to changes in population density. Note that the reference household's probability of owning 2 or more vehicles in Atlanta is 79 percent. This probability reduces to 55 percent if the household was to move to Chicago. As a result, the household is expected to demand 4,012 miles less per year.

Households adapt to the new city structure by changing their car ownership levels, as shown by the data (the probability of owning 0 or 1 cars increases by 22 percentage points, while the probability of owning 2 or 3 cars decreases by the same amount). Therefore, the change in travel demand is mostly due to changes in vehicle ownership. In the long run, households adjust their household location and therefore the impact on travel demand is smaller.

5.5 Comparison of results to existing literature

In this section I calculate the marginal effects of income and number of drivers using Train's (1985) results[20]. This was done in order to compare my results with the ones he obtained. The focus is on income and number of drivers as these are the main variables explaining car ownership in the literature. Unfortunately (for my purposes) Train did not calculate the effects of population density or other city-level

[20] I use Table 8.1 in page 147 for this exercise.

variables. Table 5-9b shows the marginal effects using Train's parameter estimates. Note that Train used only three alternatives for the vehicle quantity submodel (i.e. household owns zero, one, or two vehicles), while my model has 4 alternatives.

Nonetheless, my results are comparable to Train—The marginal effects for Train show that increasing household income from $40,000 to $60,000 would increase the probability of owning 2 cars by roughly 5%, while the probability of owning 0 or 1 car is decreased by a similar amount. Our results are along these lines, as I found that increasing income from 40,000 to 60,000 increases the probability of owning 2 and 3 cars by 2.7 and 3.6 respectively. The probability of owning less than 2 cars decreases by 6.3%.

With respect to number of drivers, the results obtained by Train (1986) show that increasing the number of drivers in the household from 1 to 2, increases the probability of owning 2 cars by 10.5%. Our results show that increasing the number of drivers from 1 to 2, increases the probability of owning 2 and 3 cars by 3.4% and 18.3%, respectively, meanwhile, the probability of owning 0 or 1 cars decreases by 2.1% and 19.6%, respectively.

BCMV focus their study on analyzing the impacts of urban form on car ownership and travel demand. Their results show that their measures of urban form have little impacts on the odds of car ownership. Only the population centrality measure has a significant impact on car ownership: households in less sprawled cities are less likely to own one, two, or three or more vehicles. A 10% increase in population centrality reduces the probability of owning 2 vehicles by 1.5% and the probability of owning 3 vehicles by 2.1%. Similarly, Train uses the variable "annual transit trips per capita in area" to capture the quality of transit in the households area. This variable also has the expected sign and magnitudes: an increase in the quality of transit in a household's area increases the probability of choosing one vehicle over two and the probability of owning no vehicles over none.

Regarding urban form, my results show that density does have an important effect on car ownership. A 10% increase in the average city level density reduces the probability of owning 2 and 3 cars by 1.7 and 8.9 percentage points, while increasing the probability of owning 0 and 1 cars by 10.6 percentage points. The remaining variables I used to describe urban shape have very small effects on car ownership.

Regarding travel demand, Bento et al show that adding an additional worker to the household raises VMT by approximately 5,000 miles per year, though most of these miles are due to an increase in car ownership (4,000 miles per year). I find that, for the specific example of a young household in Atlanta, the additional worker increases travel demand also by roughly 4000 miles per year.

5.6 Robustness of Estimated Effects

I next analyze how robust my results are to different subsets of the data and to a different functional form specification. The analysis was done following the procedure described in Chapter 4 (previous paper). Elasticities were estimated for regressions #1 to #4 in table 2 (car ownership model), and for different sample sizes (i.e. removing some cities and therefore altering sample size), and by adding and removing variables. The results obtained in regressions table 2 are highly robust. Coefficients and calculated elasticities are nearly identical between all estimated regressions. A similar approach was followed for the VMT model, with estimated coefficients being robust. Finally, a log-log specification was also estimated for the VMT equation. Results were not much affected by this change in function form. These results are reported in Table 5-10.

5.7 Concluding comments

This paper analyzes the effect that household location has on travel demand conditional on car ownership. It also analyzes the effect of household location on travel demand and car ownership.

Effect of household location on travel demand. Household location affects travel demand in two separate ways: first, it affects the number of vehicles owned by a household, and second, it affects the total number of miles driven by each car ownership level. As households move further away from their jobs, their travel demand is expected to increase. But the increase in travel demand will be less than proportional to the increase in distance to work. This is due to the fact that "commute miles" are only a small proportion of total travel demand. According to my data, approximately 23% of the average household's annual vehicle miles traveled are due to its commute.

Our results show that increasing distance to work by 10 percent will only increase travel demand by 0.8, 1.4 and 1.6 percent for households owning 1, 2, and 3 vehicles respectively. In the particular case of a young household from Atlanta, and once I consider the overall effect (i.e. change in car ownership and change in VMT) of an increase in travel demand, a 10 percent increase in distance to work will increase VMT by 1.4 percent. Additionally, most of the change (77 percent) can be attributed to a change in VMT, rather than a change in the car ownership levels.

From a policy perspective, this result is important because it allows planners to analyze the impact of programs such as the "Live Near Your Work" program in Baltimore. For instance, suppose that a family that used to live 11 miles from their work takes advantage of the incentives in this program and purchases a home only 3 miles from their work. It is estimated that this household will travel approximately 1600 miles less per year. By driving 1600 miles less per year, a household would avoid consuming roughly 80 gallons of gasoline and save around $240 per year.[21]

<u>Price elasticity of travel demand.</u> As described in Section 4.2 and Table 9, the gasoline price elasticity of travel demand is relatively inelastic in the short run, with values ranging between -0.1 and -0.26. In the long run, as households account for higher prices in their decisions such as the type of car to own and where to live, gas price elasticities are higher. In this case, estimates for the long run price elasticities range from -0.29 to -0.32. Since I am modeling travel demand taking into consideration car ownership, my estimates represent short term impacts of the independent variables in my model. Our estimated gas price elasticities are -0.041, -0.078, and -0.18 for 1-, 2- and 3-vehicle households, respectively. These elasticities are around the short term elasticity found in the literature of -0.1.

Additionally, note that my elasticities increase as vehicle ownership increases. Since VMT increases as car ownership increases, the effect of the elasticity is magnified. Households owning 1 vehicle travel on average 10,000 miles per year, while 3-vehicle households travel around 28,000 miles per year. The elasticity of -

[21] Assuming a vehicle efficiency of 20 miles per gallon and a price of gasoline of $3 dollars per gallon.

0.04 for one-car households implies that their annual miles will decrease by 40 miles per year if gasoline prices increase by 10%. On the other hand, 3-vehicle households will reduce their travel demand by approximately 504 miles per year is the price of gas increases 10%. This is an expected result as households with more vehicles can adapt easier to price increases, for instance by car sharing—this is, households may decide to travel together (in one vehicle) instead of having two individuals making discretionary trips in response to the increase in the price of gas.

Finally, recall that the gasoline price elasticity of travel demand obtained in chapter IV was around -0.28. Note that my current estimates for this elasticity once car ownership is taken into account range from -0.04 to -0.18. Both of these results agree with the existing literature, where long the term elasticity is around -0.3 while the short term elasticity is approximately -0.1. As discussed above, by taking car ownership into consideration in the analysis of travel demand, I am looking at a shorter term relative to the case in which I explore only household location. Therefore it is expected that the elasticities in this chapter are smaller in magnitude than those obtained in Chapter IV.

<u>Measures of land use.</u> Our measures of land use show that density is the most important variable affecting travel demand. In my example in section 4.3, I assumed that a household moved from Atlanta to Chicago. This household was expected to travel roughly 4,000 miles less per year. Additionally, most of the savings in annual miles are due to changes in car ownership (80%) instead of actual reductions in VMT. This means that policies affecting land use could in principle reduce travel demand, mainly by inducing changes in car ownership levels.

Endogeneity of household location. Finally, the issue of endogenous household location is an important one from the point of view of biased estimators. This is left for future research. But since the effect on distance to work on car ownership and travel demand has not been analyzed in the literature before, even treating location exogenous gives me an initial approach at the relationship between these variables.

5.8 Tables for Chapter 5

Table 5-1. Number of observations and distribution of CARS variable

Part (a) – Number of observations	
Description	**Number of observations**
Initial sample (total observations in 1995 NPTS)	**42,033**
Observations dropped from original sample	**23,446**
Not in MSA or MSA with less than 20 observations	10,629
No workers in household	2,859
No income data	3,885
No VMT data	1,371
Missing values in other explanatory variables	4,702
CARS equation [(a)]	
Total usable observations (134 MSAs)	**18,587**
Households owning no vehicles	1,237
Households owning 1 vehicle only	5,102
Households owning 2 vehicle only	8,667
Households owning 3 or more vehicles	3,581
VMT equation [(a)]	
Total usable observations (134 MSAs)	**17,350**
Households owning 1 vehicle only	5,102
Households owning 2 vehicle only	8,667
Households owning 3 or more vehicles	3,581

Notes:
[(a)] The difference in sample between the CARS and VMT equations are the 1,237 households that do not own any vehicles.

Part (b) – Distribution of CARS variable		
Category	**Observations**	**Frequency**
Households owning no vehicles	1,237	6.7%
Households owning 1 vehicle only	5,102	27.4%
Households owning 2 vehicle only	8,664	46.6%
Households owning 3 or more vehicles	3,581	19.3%
Total usable observations (134 MSAs)	**18,587**	100.0%

Table 5-2. Descriptive Statistics

Part (a) – Descriptive statistics for CARS equation

Variable	1-car household	2-car household	3-car household	Overall sample
Income, in thousands	23.97	44.46	51.03	35.04
Log gas price, in dollars	1.18	1.18	1.17	1.18
Population density	0.83	0.81	0.75	0.81
City area	0.54	0.53	0.53	0.53
City shape	2.76	2.76	2.86	2.77
Highway density	0.69	0.69	0.69	0.69
Number of drivers	2.54	2.46	2.31	2.47
Number of children	1.25	2.01	2.64	1.82
Number of kids 6-18	0.04	0.11	0.07	0.08
Number people over 65	0.24	0.49	0.62	0.43
Number of workers	0.40	0.21	0.17	0.27
Household yearly distance to work	0.80	1.57	2.14	1.39
Annual VMT, in thousands	10.09	20.47	27.99	17.16

Part (b) – Descriptive statistics for VMT equation

Variable	Atlanta	Boston	Chicago	Houston	Phoenix	San Diego
Income, in thousands	46.5	37.43	39.15	42.92	32.30	34.44
Number of workers	1.41	1.50	1.47	1.48	1.12	1.22
DTW, in thousand miles	9.31	9.85	9.13	10.17	5.74	8.25
Population density	0.55	1.63	3.22	0.62	0.23	0.59
Access to transit	0.48	0.53	0.74	0.45	0.41	0.62
Highway density	1.92	3.89	2.47	1.13	0.48	1.09
Number of kids under 6	0.06	0.10	0.07	0.09	0.09	0.09
Number of children 6-18 years old	0.36	0.45	0.48	0.29	0.40	0.28
Number of members 65 or older	0.22	0.23	0.27	0.16	0.39	0.41
Age of reference person	45.77	46.54	47.28	44.46	51.48	49.10
Years of education of reference person	13.59	12.69	12.91	13.85	12.12	13.05
Annual VMT, in thousands	20.32	19.25	15.95	20.52	16.00	17.57

Table 5-3. Car ownership model – regression results

Part (a) – Parameter estimates for 1 car option				
1 car option	**Eq. #1 Coef.** *(std. err)*	**Eq. #2 Coef.** *(std. err)*	**Eq. #3 Coef.** *(std. err)*	**Eq. #4 Coef.** *(std. err)*
Log of income	1.05	0.15	1.14	0.90
	0.0488	*0.2356*	*0.0455*	*0.0352*
Log of income squared	--	0.21	--	--
		0.0487		
Number of workers	0.02	-0.04	-0.08	-0.32
	0.0764	*0.0766*	*0.0647*	*0.0461*
Yearly distance to work	0.13	0.12	0.11	0.09
	0.0161	*0.0160*	*0.0146*	*0.0089*
Population density	-0.32	-0.31	-0.30	--
	0.0512	*0.0515*	*0.0492*	
Access to transit	-0.70	-0.68	-0.85	--
	0.3811	*0.3814*	*0.3701*	
Highway density	-0.05	-0.05	-0.05	--
	0.0192	*0.0198*	*0.0180*	
Number of children under 6	-0.49	-0.49	--	--
	0.1491	*0.1486*		
Number of children 6 to 18	-0.11	-0.11	--	--
	0.0578	*0.0577*		
Number of old	0.39	0.38	--	--
	0.0828	*0.0835*		
Reference person: Age	-0.01	-0.01	--	--
	0.0030	*0.0029*		
Reference person: Education	0.05	0.05	--	--
	0.0068	*0.0068*		
Constant	-1.59	-0.76	-1.15	-1.57
	0.2918	*0.3933*	*0.2216*	*0.0971*
Log likelihood	-16,917.8	-21,262.9	-16,784.9	-14,531.9
Number of observations	18,587	20,598	18,587	16,252
LR chi2(3)	11,079.7	7,031.88	11,345.5	6,935.9
Prob > chi2	0	0	0	0
Pseudo R2	0.2467	0.1419	0.2526	0.1927

Table 5-3 (cont.)

Part (b) – Parameter estimates for 2 car option				
2 car option	**Eq. #1 Coef.** *(std. err)*	**Eq. #2 Coef.** *(std. err)*	**Eq. #3 Coef.** *(std. err)*	**Eq. #4 Coef.** *(std. err)*
Log of income	2.31	-1.29	2.26	1.85
	0.0593	*0.2764*	*0.0537*	*0.0417*
Log of income squared	--	0.63	--	--
		0.0536		
Number of workers	1.08	1.01	0.72	0.44
	0.0793	*0.0790*	*0.0659*	*0.0461*
Yearly distance to work	0.16	0.15	0.13	0.11
	0.0161	*0.0160*	*0.0146*	*0.0088*
Population density	-0.55	-0.57	-0.51	--
	0.0561	*0.0567*	*0.0535*	
Access to transit	-1.55	-1.52	-1.70	--
	0.4058	*0.4043*	*0.3915*	
Highway density	-0.13	-0.13	-0.13	--
	0.0231	*0.0239*	*0.0215*	
Number of children under 6	0.10	0.07	--	--
	0.1521	*0.1507*		
Number of children 6 to 18	0.03	0.03	--	--
	0.0595	*0.0595*		
Number of old	1.03	1.06	--	--
	0.0884	*0.0889*		
Reference person: Age	-0.01	-0.01	--	--
	0.0033	*0.0032*		
Reference person: Education	0.02	0.01	--	--
	0.0073	*0.0073*		
Constant	-5.75	-0.52	-4.75	-5.44
	0.3265	*0.4732*	*0.2514*	*0.1279*
Log likelihood	-16,917.8	-21,262.9	-16,784.9	-14,531.9
Number of observations	18,587	20,598	18,587	16,252
LR chi2(3)	11,079.7	7,031.88	11,345.5	6,935.9
Prob > chi2	0	0	0	0
Pseudo R2	0.2467	0.1419	0.2526	0.1927

Table 5-3 (cont.)

Part (c) – Parameter estimates for 3 car option				
3 car option	**Eq. #1 Coef.** *(std. err)*	**Eq. #2 Coef.** *(std. err)*	**Eq. #3 Coef.** *(std. err)*	**Eq. #4 Coef.** *(std. err)*
Log of income	2.62	-1.99	2.48	1.99
	0.0706	*0.3442*	*0.0630*	*0.0514*
Log of income squared	--	0.77	--	--
		0.0606		
Number of workers	2.00	1.94	1.57	1.27
	0.0837	*0.0834*	*0.0703*	*0.0512*
Yearly distance to work	0.17	0.16	0.14	0.11
	0.0162	*0.0161*	*0.0147*	*0.0089*
Population density	-0.72	-0.75	-0.67	--
	0.0647	*0.0653*	*0.0616*	
Access to transit	-1.50	-1.46	-1.69	--
	0.4444	*0.4427*	*0.4280*	
Highway density	-0.28	-0.28	-0.27	--
	0.0321	*0.0326*	*0.0300*	
Number of children under 6	-0.04	-0.06	--	--
	0.1661	*0.1647*		
Number of children 6 to 18	0.07	0.07	--	--
	0.0628	*0.0627*		
Number of old	1.08	1.12	--	--
	0.1000	*0.1004*		
Reference person: Age	0.03	0.02	--	--
	0.0037	*0.0037*		
Reference person: Education	-0.01	-0.02	--	--
	0.0080	*0.0080*		
Constant	-10.23	-3.15	-7.68	-8.49
	0.3825	*0.6273*	*0.3017*	*0.1774*
Log likelihood	-16,917.8	-21,262.9	-16,784.9	-14,531.9
Number of observations	18,587	20,598	18,587	16,252
LR chi2(3)	11,079.7	7,031.88	11,345.5	6,935.9
Prob > chi2	0	0	0	0
Pseudo R2	0.2467	0.1419	0.2526	0.1927

Table 5-4. Car ownership model – Marginal effects

Part (a) – Marginal effect of increasing income from \$40k to \$60k (based on Regression #1)				
Predicted probabilities	**Car ownership levels**			
	0 cars	**1 car**	**2 cars**	**3 cars**
Probability of owning $\underline{x}$ cars, $P_0(car_i)$ Income = \$40,000	0.80%	19.90%	61.84%	17.45%
Probability of owning $\underline{x}$ cars, $P_1(car_i)$ Income = \$60,000	0.20%	10.34%	67.33%	22.13%
Change in probability due to change in income (*)	-0.60pp	-9.56pp	5.49pp	4.67pp

Notes:
(*) in percentage points (pp)

Part (b) – Marginal effects for car ownership model, all variables (based Regression #1)				
Change in independent variable	**Increase in the probability of owning $\underline{x}$ cars, in percentage points**			
	0 cars	**1 car**	**2 cars**	**3 cars**
Increase income from \$40 to \$60k	-0.60	-9.56	5.49	4.67
Increase number of workers from 2 to 3	-0.77	-18.00	0.21	18.57
Increase population density by 10%	2.25	13.68	-7.14	-8.79
Increase access to transit by 10%	0.31	3.22	-2.93	-0.60
Increase Highway density by 10%	0.03	0.43	0.17	-0.63
Increase age of ref. person by 5 years	0.01	-0.45	-1.97	2.42
Increase education of ref. person by 5 years	-0.04	3.16	-1.11	-2.01
Increase total distance to work by 10%	-0.09	-0.34	0.24	0.20

Table 5-4. (continued)

Part (c) – Marginal effects for 6 cities
Increase in the probability of owning *x* cars (in percentage points)
by moving from city in column 1 to destination cities (columns 3-8)

City of origin	Cars owned	Destination city					
		Atlanta	Boston	Chicago	Houston	Phoenix	San Diego
Atlanta	0 car	0.00	0.64	2.11	-0.05	-0.18	0.04
	1 car	0.00	10.13	20.34	-1.41	-4.61	0.51
	2 car	0.00	-3.28	-12.81	-0.40	-0.60	-2.23
	3 car	0.00	-7.48	-9.64	1.87	5.38	1.68
Boston	0 car	-0.55	0.00	1.30	-0.59	-0.70	-0.52
	1 car	-9.13	0.00	9.59	-10.37	-13.13	-8.74
	2 car	1.38	0.00	-8.57	0.63	-0.27	-0.97
	3 car	8.30	0.00	-2.33	10.34	14.10	10.22
Chicago	0 car	-1.80	-1.26	0.00	-1.84	-1.94	-1.77
	1 car	-18.56	-9.54	0.00	-19.79	-22.51	-18.18
	2 car	9.70	8.47	0.00	8.93	7.98	7.35
	3 car	10.66	2.33	0.00	12.70	16.47	12.59
Houston	0 car	0.03	0.39	1.28	0.00	-0.07	0.05
	1 car	1.16	9.76	19.16	0.00	-2.56	1.55
	2 car	0.78	-0.27	-8.42	0.00	-0.97	-1.53
	3 car	-1.96	-9.88	-12.01	0.00	3.60	-0.07
Phoenix	0 car	0.48	2.10	5.55	0.35	0.00	0.59
	1 car	5.77	17.24	26.93	4.04	0.00	6.44
	2 car	-1.28	-7.78	-18.87	-1.09	0.00	-3.40
	3 car	-4.97	-11.56	-13.62	-3.30	0.00	-3.63
San Diego	0 car	-0.07	0.88	3.00	-0.14	-0.34	0.00
	1 car	-0.62	10.46	20.84	-2.22	-5.92	0.00
	2 car	2.15	-3.08	-13.54	2.11	2.68	0.00
	3 car	-1.46	-8.26	-10.30	0.26	3.58	0.00

Notes:
This table should be read as follows: Suppose a household moves from Atlanta (column 1) to Chicago (column 5). The household's probability of owning 0 or 1 car increases by 2.1 and 20.3 percentage points, respectively. At the same time, this household's probability of owning 2 or 3 cars decreases by 12.8 and 9.6 percentage points, respectively.

Table 5-5. Vehicle miles traveled, Regression results
(Dependent variable=VMT) [(a)]

Variable	#1 1-car option Coeff. *(Std. Err)*	#2 2-car option Coeff. *(Std. Err)*	#3 3-car option Coeff. *(Std. Err)*
Log of income	1.893	3.126	3.890
	(0.1917)	*(0.3211)*	*(0.6258)*
Log gas price	-1.353	-3.326	-1.078
	(1.8288)	*(2.4420)*	*(5.0176)*
Population density	-0.213	-0.320	0.809
	(0.2001)	*(0.2634)*	*(0.5709)*
Access to transit	-3.984	-2.274	-6.890
	(1.3304)	*(1.7116)*	*(3.4337)*
City area	-0.046	0.023	-0.048
	(0.0437)	*(0.0575)*	*(0.1210)*
City shape	0.870	1.428	0.104
	(0.6997)	*(0.8702)*	*(1.8132)*
Highway density	-0.160	0.362	0.033)
	(0.0749)	*(0.1668)*	*(0.3644)*
Number of drivers	0.895	0.861	2.836
	(0.2769)	*(0.3627)*	*(0.4436)*
Number of children	-0.241	1.256	-1.037
	(0.5522)	*(0.4614)*	*(1.1053)*
Number of kids 6-18	0.304	0.066	-0.371
	(0.1948)	*(0.1744)*	*(0.3349)*
Number people over 65	-1.459	-1.378	-1.249
	(0.2007)	*(0.2730)*	*(0.6288)*
DTW x number of workers	0.315	0.345	0.356
	(0.0238)	*(0.0159)*	*(0.0290)*
Mills ratio	0.390	0.051	-0.139
	(0.0884)	*(0.1439)*	*(0.1555)*
Constant	5.884	4.199	3.930
	(1.0970)	*(1.8844)*	*(3.9945)*

Notes:
[(a)] Results in this table are based on Regression #1 in Table 3.

Table 5-6. Vehicle miles traveled - Marginal effects

Variable	1 Car household	2 Car household	3 Car household
Income	0.19	0.15	0.14
Price of gasoline	-0.13	-0.16	-0.04
Population density	-0.02	-0.01	0.02
Access to transit	-0.21	-0.06	-0.13
City area	-0.01	0.00	0.00
City shape	0.06	0.05	0.00
Highway density	-0.04	0.04	0.00
Number of drivers	0.11	0.08	0.27
Number of children	0.00	0.01	0.00
Number of kids 6-18	0.01	0.00	-0.01
Number people over 65	-0.06	-0.01	-0.01
Yearly distance to work	0.08	0.15	0.16

Table 5-7. Vehicle miles traveled - Marginal effects from moving between cities

Effect on travel demand of moving a household from
city of origin (column 1) to destination city (columns 3-8)

City of origin	Cars owned	Destination					
		Atlanta	Boston	Chicago	Houston	Phoenix	San Diego
Atlanta	1 car	0	-548	-1259	391	-87	-120
	2 car	0	15	-1388	-263	-967	-933
	3 car	0	460	783	-680	-1545	-899
Boston	1 car	603	0	-794	1031	508	472
	2 car	-15	0	-1403	-277	-981	-948
	3 car	-443	0	310	-1097	-1928	-1308
Chicago	1 car	1609	920	0	2095	1501	1460
	2 car	1635	1653	0	1328	500	539
	3 car	-734	-303	0	-1371	-2180	-1577
Houston	1 car	-366	-874	-1521	0	-447	-478
	2 car	270	286	-1153	0	-722	-688
	3 car	723	1211	1552	0	-923	-234
Phoenix	1 car	88	-466	-1176	486	0	-33
	2 car	1082	1098	-473	788	0	37
	3 car	1802	2337	2712	1009	0	754
San Diego	1 car	381	-605	-1119	-75	-162	0
	2 car	801	-910	-1217	5	107	0
	3 car	716	87	1179	-48	-540	0

Notes:
This table is read as follows: Suppose a household moves from Atlanta (column 1) to Chicago (column 5). Household owning 1 car will travel, on average, 1,259 miles less per year The household's probability of owning 0 or 1 car increases by 3 and 13 percentage points, respectively. At the same time, this household's probability of owning 2 or 3 cars decreases by 9 and 7 percentage points, respectively.

Table 5-8. Overall marginal effects from moving young household in Atlanta

Description	1-car household	2-car household	3-car household
Original predicted probabilities [b]	20%	67%	12%
Original predicted VMT, in thousands [c]	11.6	19.8	22.2
Total (weighted) VMT for young household[d]	18,076	Miles per year	
Case 1: increase income by 50%			
Percentage Change in probabilities	-6.9pp	5.4pp	2.7pp
Predicted probabilities after income increase	13%	72%	14%
Predicted VMT after increase, in thousands	12.4	21.1	23.7
Total (weighted) VMT	20,179	Miles per year	
Change in VMT from increase in income	2,102.4	Miles per year	
Case 2: Increase DTW by 50%			
Percentage Change in probabilities	-1.1pp	1.3pp	0.5pp
Predicted probabilities after increase in DTW	13%	72%	14%
Predicted VMT after increase, in thousands	12.4	20.7	23.1
Total (weighted) VMT	19,197	Miles per year	
Change in VMT from increase in DTW	1120.7	Miles per year	
Case 3: Additional worker in household			
Percentage Change in probabilities	-13.1pp	-2.4pp	16.8%pp
Predicted probabilities after increase in workers	7%	64%	29%
Predicted VMT after increase, in thousands	11.6	19.8	22.2
Total (weighted) VMT	19.821	Miles per year	
Change in VMT from increase in workers	1744.6	Miles per year	
Case 4: Household moves from Atlanta to Chicago			
Percentage Change in probabilities	15.3pp	-17.0pp	-6.7pp
Predicted probabilities after move	35%	50%	5%
Change in VMT from move	10.3	18.8	22.8
Total (weighted) VMT	14.064	Miles per year	
Change in VMT from changing cities	-4012.1	Miles per year	

Notes:
[a] Young household in Atlanta is defined as having four individuals (reference person, spouse and two children). The household head is the only worker in the household, earns $40,000 per year, and lives 11 miles from work (one way)
[b] Predicted probabilities are estimated using equation #1 in Table 5-3
[c] Predicted VMT are estimated using equation in Table 5-5
[d] Total (weighted) VMT is estimated as follows: VMT = Σ Prob(CARS = i) x VMT_i

Table 5-9. Regression results from Train (1986)

Part (a) – Regression results, Car ownership model (K. Train, 1986		
Variable	Coefficient estimates One-vehicle household	 Two-vehicle household
Log income	1.05	1.57
Number of workers	1.08	1.5
log household size	0.181	0.197
Number of transit trips per capita	-0.0009	-0.0021
average utility	0.635	0.635
Constant	-1.79	-4.95

Part (b) – Marginal effects, car ownership model			
Variable	**0 cars**	**1 car**	**2 cars**
Income	-0.35%	-4.59%	4.94%
Number of workers	-0.27%	-10.21%	10.48%

6 Concluding comments

The objective of this dissertation is to analyze the impact of land use on transportation conditional on household location and car ownership. Much of the research on VMT has focused on the effect of car ownership on VMT, and very little has been said about the interaction between household location and VMT. In particular, I examine household's distance to work (DTW) within a city. The short-to-medium term constraints imposed by household location are likely to be more important economically and for policy than the automobile stock. I approached the above objective in two separate manners: first I treat distance to work as endogenous and study the interaction between VMT and DTW. Second, I study the overall effect of DTW on car ownership and VMT. This approach allows me to untangle the direct and indirect effects of DTW on VMT via car ownership.

6.1 Summary of results

Regarding the first issue of examining the simultaneous interaction of DTW and VMT, I reach two broad conclusions. First, those city characteristics that might be expected to affect commutes or the jobs-housing match (other than the city's physical size) have remarkably little effect on households' distance-to-work. Variables like city shape, or the joint jobs-housing distribution have little apparent effect on city-average distance-to-work. Only commute speed has a substantial effect on DTW. Our measure of congestion (MEDSPEED) suggests that people will locate around 8.8 percent further from their work when commute speed increases by 10 percent. In other words, by increasing commute speed from a national average of 29.5 mph to 32.5 mph, I expect people to live on average 1.24 miles further their jobs than they do now. That is, the mean one-way distance-to-work would increase from 9.7 to 10.9 miles

Second, I conclude that distance-to-work provides a quite modest constraint on overall household vehicle miles traveled. A one percent increase in distance-to-work leads to a 0.33 percent increase in VMT. In level terms, a one mile increase in one-way distance-to-work for one worker, which translates into roughly 480 additional commuting miles per year, leads to an annual increase of about 206 vehicle miles. These figures mean that a reduction in distance-to-work would have only small effects on overall vehicle miles traveled.

With respect to the second approach to analyze the effect of distance to work on travel demand conditional, I find that household location affects travel demand in two separate ways: first, it affects the number of vehicles owned by a household, and second, it affects the total number of miles driven by each car ownership level. As households move further away from their jobs, their travel demand is expected to increase. But the increase in travel demand will be less than proportional to the increase in distance to work. Our results show that increasing distance to work by 10 percent will only increase VMT by 1.4 percent. One explanation for this small effect could be to the fact that “commute miles” are only a small proportion of total travel demand. According to my data, approximately 23% of the average household’s annual vehicle miles traveled are due to its commute. Finally, of all my measures of city form, access to transit and population density have the highest effects on VMT, but the effects are still small.

In addition to the above, there are two other results that I find noteworthy and that I feel have received insufficient attention from the literature. First, I find that conditional on distance-to-work, people do not drive (much) more in physically larger cities. This result may not be surprising, since non-commuting "chores" can mostly be done locally, regardless of a city's size, but the size and nature of this conclusion has not been estimated to my knowledge. Previous research has either not examined the city area effect (despite, I feel, its seemingly obvious role) or, in the few cases where it has been included, has not emphasized it (BCMV). One implication of my finding is that household migration – mostly from physically small to large cities – will likely have substantial effects on nationwide VMT. This effect has not been much remarked on.

I also find that non-working households have a considerably smaller VMT-gas-price elasticity. Previous literature has not focused on the work decision, despite the fact that non-working households drive approximately 10,000 miles less per year than working households. I estimate separate VMT equations for workers and non-workers (i.e., no workers in the household.) Of course, this distinction also entails my recognizing that the work decision is endogenous.

6.2 Directions for future research

There are several key areas worth investigating in the future: First, an obvious extension of the model in Chapter 5 would be to tackle the endogeneity issue of distance to work by explicitly modeling this decision. Under these circumstances, it could be assumed that households simultaneously select where to live, how many cars to own, and how much to drive to maximize their utility. The structural equations resulting from this problem are listed below:

$$\text{Prob (CARS} = i) = \text{Prob } (Ui > Uj) \quad \text{for } i \neq j \text{ and } i=0,1,2, \text{ or } 3 \qquad [6\text{-}1]$$

$$\ln(DTW) = X1\,\beta + \varepsilon 2 \qquad [6\text{-}2]$$

$$VMT = X2 + \delta\, DTW + \varepsilon 3 \qquad [6\text{-}3]$$

Estimation of this model is not very straightforward, but several approaches can be analyzed. For instance, the model could be estimated via maximum likelihood, though the set up is not very easy.

A second natural extension of the work developed in this dissertation would be to model commute mode. This could be done following the work of Bhat (2005), Train (1986), and Mannering and Winston (1985) who have applied nested logit models to study car ownership and type choice. This framework could be used to model car ownership and commute mode. This approach could lead to very interesting policy questions related to commute and non-commute VMTs and policies to reduce VMTs.

A third direction is to tackle a different set of questions. Two such questions in particular are suggested by my research. First, I wonder how these conclusions would change if I characterized household location using Time-to-Work (TTW) rather than DTW. TTW may be a more accurate measure of the costs of living far from one's work than is DTW. It has two potential drawbacks, however: (i) its role in contributing to externalities is weaker, and (ii) it cannot be naturally measured on the same scale as VMT and DTW allow. Second, the high degree of household mobility in the U.S. suggests a future research agenda that explicitly studies the DTW or TTW decisions of movers and the implications of such mobility for VMT.

Appendices

7 Appendix 1 –Switching regression when one of the regimes is distributed bivariate normal

7.1 Switching regression model

In the following model, the behavior of the agents is described by three equations working under two regimes. There is one criterion function that determines which of the two regimes is applicable. The special thing about the particular model described next is that in regime 1, behavior is characterized by a system of two equations. Furthermore, these equations are assumed to be distributed bivariate normal. The following model is based on the switching model described in Maddala (1983), page 223.

The criterion function is defined by equation A.1 below:

$$I = \begin{cases} 1 & \text{if } \varepsilon_1 > -\gamma Z \\ 0 & \text{if } \varepsilon_1 < -\gamma Z \end{cases} \tag{A.1}$$

The model is completed by the following two regimes:

$$\begin{array}{lll} \text{Regime 1:} & \begin{cases} VMT = X_2\beta_2 + \delta DTW + \varepsilon_2 \\ DTW = X_3\beta_3 + \varepsilon_3 \end{cases} & \text{iff } I = 1 \\ \text{Regime 2:} & VMT = X_4\beta_4 + \varepsilon_4 & \text{iff } I = 0 \end{array} \tag{A.2}$$

Following the model described in 4, the variance-covariance matrix for the system described in Equation A.2 is defined as:

$$V = \begin{bmatrix} 1 & \sigma_{12} & 0 & \sigma_{14} \\ & \sigma_2^2 & \sigma_{23} & 0 \\ & & \sigma_3^2 & 0 \\ & & & \sigma_4^2 \end{bmatrix} \qquad \text{A.3}$$

where *V* is an symmetric matrix, and, furthermore, it is assumed that $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$

The likelihood function for the model defined by equations A.1 to A.3 is:

$$L(\beta,\sigma,\rho) = \prod \left[\int_{-\gamma Z}^{\infty} g(\varepsilon_1,\varepsilon_2,\varepsilon_3)d\varepsilon_1 \right]^{I_i} \left[\int_{-\infty}^{-\gamma Z} f(\varepsilon_1,\varepsilon_4)d\varepsilon_1 \right]^{1-I_i} \qquad \text{A.4}$$

where $g(\varepsilon_1,\varepsilon_2,\varepsilon_3)$ and $f(\varepsilon_1,\varepsilon_4)$ are trivariate and bivariate normal distributions, respectively. It can be shown that equation A.4 can be expressed as:

$$L(\beta,\sigma,\rho) = \prod_{I=1} g_{23}(\varepsilon_2,\varepsilon_3)\cdot\Phi\big(-w(\varepsilon_2,\varepsilon_3)\big)\cdot\prod_{I=0} f_4(\varepsilon_4)\cdot\Phi\big(w(\varepsilon_4)\big) \qquad \text{A.5}$$

where

$$w(\varepsilon_2,\varepsilon_3) \equiv \frac{-\gamma Z - \mu^*}{\eta_1}$$

$$w(\varepsilon_4) \equiv \frac{-\gamma Z - \rho_{14}\frac{\varepsilon_4}{\sigma_4}}{\sqrt{1-\rho_{14}^2}}$$

where $g_{23}(\varepsilon_2,\varepsilon_3)$ is a bivariate normal distribution function such that, first, $g(\varepsilon_1,\varepsilon_2,\varepsilon_3) = g(\varepsilon_1 \mid \varepsilon_2,\varepsilon_3)\cdot g_{23}(\varepsilon_2,\varepsilon_3)$, second, $f_4(\varepsilon_4)$ is a normal distribution function such that $f(\varepsilon_1,\varepsilon_4) = f(\varepsilon_1 \mid \varepsilon_4)\cdot f_4(\varepsilon_4)$; and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal. The log likelihood function is therefore:

$$\ln L(\beta,\sigma,\rho) = \sum_{I=1}\{\ln g_{23}(\varepsilon_2,\varepsilon_3) + \ln\Phi(-w(\varepsilon_2,\varepsilon_3))\} + \sum_{I=0}\{\ln f_4(\varepsilon_4) + \ln\Phi(w(\varepsilon_4))\} \qquad \text{A.6}$$

where each of the functions is defined above.

7.2 Derivation of the Likelihood function

In order to show that equation A.4 can be expressed as equation A.5, I will derive alternative expressions for each of the two terms in the square brackets on the right hand side of equation A.4. First, I begin by showing that the first integral on the right hand side of equation A.4 equals:

$$\int_{-\gamma Z}^{\infty} g(\varepsilon_1,\varepsilon_2,\varepsilon_3)d\varepsilon_1 = g_{23}(\varepsilon_{2i},\varepsilon_{3i})\cdot\Phi\big(-w(\varepsilon_2,\varepsilon_3)\big) \qquad \text{A.7}$$

To do this, note that the trivariate normal density function $g(\varepsilon_1,\varepsilon_2,\varepsilon_3)$ may be written in terms of its conditional and marginal distributions, as follows:

$$\begin{aligned}\int_{-\gamma Z}^{\infty} g(\varepsilon_1,\varepsilon_2,\varepsilon_3)d\varepsilon_1 &= \int_{-\gamma Z}^{\infty} g_1(\varepsilon_1 \,|\, \varepsilon_2,\varepsilon_3)\bullet g_{23}(\varepsilon_2,\varepsilon_3)d\varepsilon_1 \\ &= g_{23}(\varepsilon_2,\varepsilon_3)\bullet\int_{-\gamma Z}^{\infty} g_1(\varepsilon_1 \,|\, \varepsilon_2,\varepsilon_3)d\varepsilon_1\end{aligned} \qquad \text{A.8}$$

where the conditional distribution can be expressed as:

$$g_1(\varepsilon_1 | \varepsilon_2, \varepsilon_3) \sim N[\mu^*, \eta_1^2]$$ A.9

where

$$\mu^* = \mu_1 - \frac{\eta_1}{\eta_2}(\varepsilon_2 - \mu_2) - \frac{\eta_1}{\eta_3}(\varepsilon_3 - \mu_3)$$

$$\eta_1 = \sqrt{\frac{(1 - \rho_{12}^2 - \rho_{23}^2)}{(1 - \rho_{23}^2)}}$$

$$\eta_2 = -\frac{\sigma_2 \sqrt{(1 - \rho_{23}^2)(1 - \rho_{12}^2 - \rho_{23}^2)}}{\rho_{12}}$$

$$\eta_3 = \frac{\sigma_3 \sqrt{(1 - \rho_{23}^2)(1 - \rho_{12}^2 - \rho_{23}^2)}}{(\rho_{23}\rho_{12})}$$

The variables μ_1, μ_2 and μ_3 are the means of $\varepsilon_1, \varepsilon_2$ and ε_3 respectively. Therefore, assuming that $\mu_1 = \mu_2 = \mu_3 = 0$, the probability density function for equation A.8 is

$$\int_{-\gamma Z}^{\infty} g_1(\varepsilon_1 \mid \varepsilon_2, \varepsilon_3) d\varepsilon_1 = \int_{-\gamma Z}^{\infty} \frac{1}{\sqrt{2\pi} \bullet \eta_1} \exp\left\{-\frac{1}{2}\left[\frac{\varepsilon_1 - \mu^*}{\eta_1}\right]^2\right\} d\varepsilon_1$$ A.10

where $\mu^* = -\frac{\eta_1}{\eta_2}\varepsilon_2 - \frac{\eta_1}{\eta_3}\varepsilon_3$ and $\eta_1, \eta_2,$ and η_3 are defined in equation A.7 above.

Define the following transformation

$$z = \frac{\varepsilon_1 - \mu^*}{\eta_1}$$ A.11

where

$$dz = \frac{d\varepsilon_1}{\eta_1}$$ A.12

$$\varepsilon_1 \to -\gamma Z \Rightarrow z \to \frac{-\gamma Z - \mu^*}{\eta_1} \equiv w(\varepsilon_2, \varepsilon_3)$$

$$\varepsilon_1 \to \infty \Rightarrow z \to \infty$$

Therefore, by replacing the equations in A.12 into equation A.10, I obtain

$$\int_{-\gamma Z}^{\infty} g(\varepsilon_1 \mid \varepsilon 2, \varepsilon_3)\, d\varepsilon_1 = \int_{-\gamma Z}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{\varepsilon_1 - \mu^*}{\eta_1}\right]^2\right\} \frac{d\varepsilon_1}{\eta_1} \tag{A.13}$$

$$= \int_{w(\varepsilon_2,\varepsilon_3)}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z^2\right\} dz$$

$$= 1 - \Phi\left(w(\varepsilon_2, \varepsilon_3)\right)$$

$$= \Phi\left(-w(\varepsilon_2, \varepsilon_3)\right)$$

Finally, replacing equation A.13 into A.8 yields

$$\int_{-\gamma Z}^{\infty} g(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 = g_{23}(\varepsilon_{2i}, \varepsilon_{3i}) \bullet \Phi(-w(\varepsilon_2, \varepsilon_3)) \tag{A.14}$$

$$= g_{23}(\varepsilon_{2i}, \varepsilon_{3i}) \bullet \Phi\left(\frac{\gamma Z + \mu^*}{\sigma^*}\right)$$

where $g_{23}(\varepsilon_{2i}, \varepsilon_{3i})$ is the pdf of a bivariate normal distribution, $\Phi\left(w(\varepsilon_2, \varepsilon_3)\right)$ is the cdf of a standard normal distribution, and $w(\varepsilon_2, \varepsilon_3)$ is defined in equation A.12 above.

Second, I will show that the second integral on the right hand side of equation A.4 equals

$$\int_{-\infty}^{-\gamma Z} f(\varepsilon_1, \varepsilon_4) d\varepsilon_1 = f_4(\varepsilon_4) \cdot \Phi\left(w(\varepsilon_4)\right) \tag{A.15}$$

To do this, note that the bivariate normal $f(\varepsilon_1, \varepsilon_4)$ can be expressed in terms of its conditional and marginal distributions, as follows:

$$\int_{-\infty}^{-\gamma Z} f(\varepsilon_1, \varepsilon_4) d\varepsilon_1 = \int_{-\infty}^{-\gamma Z} f_4(\varepsilon_4) \bullet f_1(\varepsilon_1 \mid \varepsilon_4) d\varepsilon_1 \tag{A.16}$$

$$= f_4(\varepsilon_4) \bullet \int_{-\infty}^{-\gamma Z} f_1(\varepsilon_1 \mid \varepsilon_4) d\varepsilon_1$$

where $f_1(\varepsilon_1 \mid \varepsilon_4) \sim N\left[\mu_1 + \rho_{14} \frac{\sigma_1}{\sigma_4}(\varepsilon_4 - \mu_4), \sigma_1(1-\rho_{14}^2)\right]$, and μ_1 and μ_4 are the means of ε_1 and ε_4, respectively. By assumption, I have that $\mu_1 = \mu_4 = 0$ and $\sigma_1^2 = 1$. Therefore $f_1(\varepsilon_1 \mid \varepsilon_4) \sim N\left[\frac{\rho_{14}}{\sigma_4}\varepsilon_4, (1-\rho_{14}^2)\right]$. The function $f_4(\varepsilon_4)$ is distributed normal with mean 0 and variance σ_4^2. Under these conditions, the integral on the right hand side of equation A.16 can be expressed as:

$$\int_{-\infty}^{-\gamma Z} f_1(\varepsilon_1 \mid \varepsilon_4) d\varepsilon_1 = \int_{-\infty}^{-\gamma Z} \frac{1}{\sqrt{2\pi(1-\rho_{14}^2)}} \exp\left\{ -\frac{1}{2}\left[\frac{\varepsilon_1 - \frac{\rho_{14}}{\sigma_4}\varepsilon_4}{\sqrt{(1-\rho_{14}^2)}} \right]^2 \right\} d\varepsilon_1 \qquad \text{A.17}$$

Define the following transformation

$$z = \frac{\varepsilon_1 - \rho_{14}\frac{\varepsilon_4}{\sigma_4}}{\sqrt{1-\rho_{14}^2}} \qquad \text{A.18}$$

where

$$dz = \frac{d\varepsilon_1}{\sqrt{1-\rho_{14}^2}} \qquad \text{A.19}$$

$$\varepsilon_1 \to -\infty \Rightarrow z - \infty$$

$$\varepsilon_1 \to -\gamma Z \Rightarrow z \to \frac{-\gamma Z - \rho_{14}\frac{\varepsilon_4}{\sigma_4}}{\sqrt{1-\rho_{14}^2}} \equiv w(\varepsilon_4)$$

Therefore by replacing equations A.18 and A.19 into equation A.17, I obtain

$$\int_{-\infty}^{-\gamma Z} f(\varepsilon_1 | \varepsilon_4) d\varepsilon_1 = \int_{-\infty}^{-\gamma Z} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[\frac{\varepsilon_1 - \frac{\rho_{14}}{\sigma_4}\varepsilon_4}{\sqrt{(1-\rho_{14}^2)}} \right]^2 \right\} \frac{d\varepsilon_1}{\sqrt{1-\rho_{14}^2}} \quad \text{A.20}$$

$$= \int_{-\infty}^{w(\varepsilon_4)} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} z^2 \right\} dz$$

$$= \Phi\left(w(\varepsilon_4)\right)$$

where $w(\varepsilon_4)$ is defined in equation A.19 above. Replacing A.20 into A.16 yields the expression in A.15, which is what I were trying to show.

Finally, replacing equations A.7 and A.15 into equation A.4 yields the following expression for the likelihood function:

$$L(\beta, \sigma, \rho) = \prod_{I=1} g_{23}(\varepsilon_{2i}, \varepsilon_{3i}) \cdot \Phi(-w(\varepsilon_2, \varepsilon_3)) \cdot \prod_{I=0} f_4(\varepsilon_4) \cdot \Phi(w(\varepsilon_4)) \quad \text{A.20}$$

where

$$w(\varepsilon_2, \varepsilon_3) \equiv \frac{-\gamma Z - \mu^*}{\eta_1}$$

$$w(\varepsilon_4) \equiv \frac{-\gamma Z - \rho_{14}\frac{\varepsilon_4}{\sigma_4}}{\sqrt{1-\rho_{14}^2}}$$

where $\mu^* = -\frac{\eta_1}{\eta_2}\varepsilon_2 - \frac{\eta_1}{\eta_3}\varepsilon_3$. Taking the log of the likelihood function yields:

$$\ln L(\beta, \sigma, \rho) = \sum_{I=1} \{\ln g_{23}(\varepsilon_2, \varepsilon_3) + \ln \Phi(-w(\varepsilon_2, \varepsilon_3))\} + \sum_{I=0} \{\ln f_4(\varepsilon_4) + \ln \Phi(w(\varepsilon_4))\} \quad \text{A.21}$$

where $w(\varepsilon_2, \varepsilon_3)$ and $w(\varepsilon_4)$ are defined in A.20 above.

7.3 Maximum likelihood estimation

The model described in the previous section can in principle be estimated via maximum likelihood. A STATA program was developed to estimate the parameters. This program is listed in the annex to the appendix. Unfortunately, the program does not converge when the NPTS data are used. Simulations using Monte Carlo methods were carried out to determine whether the program was working correctly. The results of estimating the model in Chapter 4, using generated data are listed in the table below:

Table A-1. Simulations using Monte Carlo for the Model used in Chapter 4. Estimation is done using FIML and LIML

	Variable			
	X21	X22	Y3	Constant
True beta	0.35	-0.29	1.25	1.16
Beta (FIML)	0.349819	-0.28751	1.210815	1.185928
Std Dev (FIML)	0.330569	0.339508	0.326116	0.568624
Beta (LIML)	0.350138	-0.29158	1.186442	1.243583
Std Dev (LIML)	0.531846	0.570238	0.820107	0.416682

For the simulation, a dataset with 19 parameters and 1000 observations was created. The model was estimated 1000 times, and the average and standard deviations are presented for the equation that in Chapter 4 represents VMT. The model was estimated both using FIML and LIML. Note that the FIML estimation approximates the true betas very well and the standard error is smaller than when LIML is used. But LIML is still a good estimator, as the true beta is contained within a standard error of the estimated beta.

Annex to Appendix 1

Table A-1. Stata Code used to estimate model by maximum likelihood

```
***************************************************************************
*                               SIMULATIONS                               *
*                                                                         *
* FIML and LIML estimation of a swithching regression model where one of  *
* the regimes is a system of equations                                    *
***************************************************************************
clear
set more off

capture log close
log using $stata\pwj\FIMLandLIML_10000iter_09082007,replace


*-------------------------------------------------------------------------*
* STEP 1 - CREATE COEFFICIENTS                                            *
*-------------------------------------------------------------------------*
capture program drop gencoef2
program gencoef2  /* 1 parameter */
version 8.2
    local k `1'
    local kplus1=`k'+1

*.........................................................................*
* Create matrix where original ("true") coefficients will be stored.      *
*.........................................................................*
    matrix beta1=J(4,`kplus1',.)

*.........................................................................*
* Create coefficients (betas) for the "k" independent variables (Xij) and  *
* the coefficient for the constant term. Organize results in a vector.     *
*.........................................................................*
    forvalues j=1/4{
    forvalues i=1/`kplus1' {
        matrix beta1[`j',`i']=int(100*invnorm(uniform()))/100
    }
    }

end
*-------------------------------------------------------------------------*



*-------------------------------------------------------------------------*
* STEP 2 - CREATE FICTIONAL DATA (X, e, and Y)                            *
*-------------------------------------------------------------------------*
capture program drop gendata  /* 1 parameter */
program gendata
version 8.2
  local obs `1'
  local k   `2'                 /* No. of indep. vars excluding constant */
  local s12 `3'

  local kplus1=`k'+1


*.........................................................................*
* Create error term                                                       *
*.........................................................................*
  matrix mu  = (   0,      0,     0,     0)

  matrix var = (    1, `s12',     0,  0.57\     ///
               `s12',     4,   0.8,     0\     ///
                   0,   0.8,  1.44,     0\     ///
                0.57,     0,     0,     2)

  drawnorm e1 e2 e3 e4,n(`obs') m(mu) cov(var)
```

```
*.........................................................................*
* Create independent variables X_ij and dependent variable Y              *
*.........................................................................*

 forvalues j=1/4{
    forvalues i=1/`k'{
       gen x`j'`i'=uniform()
    }
  }

end
*-------------------------------------------------------------------------*



***************************************************************************
* STEP 3 - ESTIMATE REGRESSION AND RECOVER PARAMETERS                     *
***************************************************************************
display c(current_time)
clear
local maxiter= 1000
local obs    = 1000
local coef   = 2                   /* No. of indep. vars excluding constant */
local eqns   = 4


scalar alfa  = 1.25
local  s12   = 1.2

local kplus1 = `coef' + 1
local maxobs = max(`maxiter',`obs')

gencoef2 `coef'                    /* Create coefficients                   */

matrix list beta1

*tempfile bliml bfiml
quietly{

*.........................................................................*
* Begin iteration procedure                                               *
*.........................................................................*
  forvalues x=1/`maxiter'{
  gendata `obs' `coef' `s12'

*.........................................................................*
* Generate Y variables                                                    *
*.........................................................................*
  gen y1 = beta1[1,1]*x11 + beta1[1,2]*x12 + beta1[1,3] + e1
  gen y3 = beta1[3,1]*x31 + beta1[3,2]*x32 + beta1[3,3] + e3
  gen y4 = beta1[4,1]*x41 + beta1[4,2]*x42 + beta1[4,3] + e4

  gen y2 = beta1[2,1]*x21 + beta1[2,2]*x22 + beta1[2,3] + scalar(alfa) * y3 + e2


  gen I=y1>0

*.........................................................................*
* LIML estimation                                                         *
*.........................................................................*
  probit I x1*
  predict Ihat,xb

  reg y3 x3*
  predict y3hat

  gen mills1 = normden(Ihat)/norm(Ihat)  if I==1
  gen mills0 = normden(Ihat)/norm(-Ihat) if I==0

  reg y2 x2* mills1 y3hat if I==1
```

```
  noisily display in green "LIML round `x' completed"

  preserve
*.......................................................................*
* Create new file with estimated parameters for LIML                   *
*.......................................................................*
      matrix bhat_l=e(b)
      local kplus1 = colsof(bhat_l)

          if `x'==1{
              svmat bhat_l
              set obs `maxobs'
              keep bhat*
              save c:\bliml,replace
/*        end if */    }
          else{
              use c:\bliml,clear
              forvalues y=1/`kplus1'{
                  replace bhat_l`y'= bhat_l[1,`y']  in `x'
/*            end forvalues */  }
              save c:\bliml,replace
/*        end else */    }

  restore


*.......................................................................*
* FIML estimation                                                       *
*.......................................................................*
  ml model lf swregtri_lf_v4 (I = x1*) (y2 = x2* y3) (y3 = x3*) (y4 = x4*)  ///
                           /sigma2 /sigma3 /sigma4                          ///
                           /rho12  /rho23  /rho14
  ml search
  ml maximize

  noisily display in green "FIML round `x' completed"

*.......................................................................*
* Create new file with estimated parameters for FIML                   *
*.......................................................................*
      matrix bhat_f=e(b)
      local kplus1 = colsof(bhat_f)

        if `x'==1{
            svmat bhat_f
            set obs `maxobs'
            keep bhat*
            save c:\bfiml,replace
/*      end if */    }
        else{
            use c:\bfiml,clear
            forvalues y=1/`kplus1'{
                replace bhat_f`y'= bhat_f[1,`y']  in `x'
/*          end forvalues */  }
            save c:\bfiml,replace
/*      end else */    }



*.......................................................................*
* Show some output while the "loop" runs (for the entertainment value)  *
*.......................................................................*
    if mod(`x',50)==0{
      noisily display in yellow "Round `x' of `maxiter'"
/*  end if      */    }
/*end forvalues */  }
 /*end quietly   */  }

use c:\bliml,clear
  preserve
    collapse (mean)b*
```

```
    mkmat _all,matrix(bliml_ave)
  restore
  preserve
    collapse (sd)b*
    mkmat _all,matrix(bliml_sd)
  restore

use c:\bfiml,clear
  preserve
    collapse (mean)b*
    mkmat _all,matrix(bfiml_ave)
  restore
  preserve
    collapse (sd)b*
    mkmat _all,matrix(bfiml_sd)
  restore


*.....................................................................*
* Organize matrices into nice output                                  *
*.....................................................................*
display c(current_time)

matrix fe1=bfiml_ave[1,1..3]
matrix fe2=bfiml_ave[1,4..7]
matrix fe3=bfiml_ave[1,8..10]
matrix fe4=bfiml_ave[1,11..13]
matrix fe5=bfiml_ave[1,14..19]
matrix fe1sd=bfiml_sd[1,1..3]
matrix fe2sd=bfiml_sd[1,4..7]
matrix fe3sd=bfiml_sd[1,8..10]
matrix fe4sd=bfiml_sd[1,11..13]
matrix fe5sd=bfiml_sd[1,14..19]
matrix E1=beta1[1,1...] \ fe1 \ fe1sd
matrix E3=beta1[3,1...] \ fe3 \ fe3sd
matrix E4=beta1[4,1...] \ fe4 \ fe4sd

* true coeffs VMT eqn
matrix A1 = beta1[2,1..2]
matrix A2 = (scalar(alfa))
matrix A3 = beta1[2,3]
matrix A  = A1,A2,A3
matrix E2 = A \ fe2 \ fe2sd \ bliml_ave[1,1..4] \ bliml_sd[1,1..4]
matrix rownames E2 = "True beta" "Beta (FIML)" "Std Dev (FIML)" "Beta (LIML)" "Std Dev
(LIML)"
matrix colnames E2 = "X21" "X22" "Y3" "Constant"
matrix list E2
matrix dir
matrix list beta1
matrix list bfiml_ave
matrix list bfiml_sd
matrix list bliml_ave
matrix list bliml_sd
log close
exit
```

References

1. Alonso, William. 1964. Location and land use. Cambridge: Harvard University Press.
2. Anas, Alex. 1981. The Estimation of Multinomial Logit Models of Joint Location and Travel Mode Choice from Aggregated Data. Journal of Regional Science, v21 n2, pp.223-42
3. Anas, Alex. 1983. Discrete choice theory, information theory and the multinomial logit and gravity models. Transportation Research Part B, v.17, pp.13-23
4. Badoe, Daniel and Eric J. Miller. 2000. Transportation-Land Use Interaction: Empirical Findings in North America, and Their Implications for Modeling, Transportation Research Part D, v5, pp.235-263.
5. Bento, Antonio M., Maureen L. Cropper, Ahmed Mushfiq Mobaraq, and Katja Vinha. 2005. The impact of urban spatial structure on travel demand in the United States. The Review of Economics and Statistics, v.87.
6. Berkowitz, Michael K., Nancy T. Gallini, Eric J. Miller, and Robert A. Wolfe. 1990. Disaggregate analysis of the demand for gasoline. Canadian Journal of Economics, v23, n2, pp.253-275.
7. Boarnet, Marlon, and R. Crane. 2001. The influence of land use on travel behavior: specification and estimation strategies. Transportation Research Part A, v35, n9, pp.823-45.

8. Boarnet, Marlon G. and Sharon Sarmiento. 1998. Can Land Use Policy Really Affect Travel Behavior? A Study of the Link Between Non-Work Travel and Land Use Characteristics. Urban Studies, v35, n7, pp.1155-1169.

9. Cervero. 1989. Jobs–housing balancing and regional mobility, American Planning Association Journal, Spring, pp.136–150.

10. Chatman, D.. 2002. The Influence of Workplace Land Use and Commute Mode Choice on Mileage Traveled for Personal Commercial Purposes," Unpublished manuscript.

11. Clark, W. A. V. and W. F. J van Lierop. 1986. Residential Mobility and Household Location Modeling, in P. Nijkamp, "Handbook of Regional and Urban Economics. Amsterdam, North-Holland.

12. Dahl, C., 1995 Demand for Transportation Fuels: A Survey of Demand Elasticities and Their Components, *Journal of Energy Literature* 1(2).

13. De Jong, G.C., Fox, J., Daly, A., Pieters, M., Smit, R., 2004. A comparison of car ownership models. Transport Reviews, v24 n4, pp.379–408.

14. DeSalvo, Joseph. 1977. Theory of locally employed urban household. Journal of Regional Science, v17, pp.345-356

15. DeSalvo, Joseph. 1985. A model of urban household behavior with leisure choice. Journal of Regional Science, v25, pp.159-174

16. Domencich, Tom and Daniel McFadden. 1975. Urban Travel Demand: A Behavioral Analysis. North-Holland Publishing Co.

17. Dubin, Jeffrey A. and Daniel L.McFadden. 1984. An econometric analysis of residential electric appliance holdings and consumption. Econometrica, v52 n2, pp.345-362.

18. Dunphy, Robert T. and Kimberly Fisher. 1996. Transportation Congestion and Density: New Insights. Transportation Research Record1552, pp.89-96.

19. Eberts, Randall W. 1981. An Empirical Investigation of Intraurban Wage Gradients. Journal of Urban Economics, v10 n1, pp.50-60

20. Eberts, Randall W. and Timothy J. Gronberg. 1982. Wage Gradients, Rent Gradients, and the Price Elasticity of Demand for Housing: An Empirical Investigation. Journal of Urban Economics, v12 n2, pp.168-76

21. Eliasson, Jonas and Lars-Goran Mattsson. 2000. A model for integrated analysis of household location and travel choices. Transportation Research Part A, v34, pp.375-394.

22. Fujita, Masahisa . 1989. Urban economic theory: Land use and city size. Cambridge; New York and Melbourne: Cambridge University Press.

23. Gabriel and Rosenthal. 1989. Household Location and Race: Estimates of a Multinomial Logit Model. The Review of Economics and Statistics, n71 n2, pp. 240-249.

24. Giuliano, Geneviva. 2003. Travel, location and race/ethnicity. Transportation Research Part A, v37, pp.351–372

25. Giuliano, Genevieve and Kenneth Small. 1993. Is the Journey to Work Explained by Urban Structure?. Urban Studies, v30, n9, pp.1485-1500.

26. Goodwin, P., Joyce Dargay, and M. Hanly, “Elasticities of Road Traffic and Fuel Consumption with Respect to Price and Income: A Review,” *Transport Review* 24(3) (2004): 275-92

27. Greening, L., T. Hann, J. Formby, and D. Cheng, “Use of region, life-cycle and role variables in the short-run estimation of the demand for gasoline and miles traveled,” *Applied Economics* 27 (1995): 643-656.

28. Gronberg, Timothy J. and Jack Meyer. 1982. Spatial Pricing, Spatial Rents, and Spatial Welfare. Quarterly Journal of Economics, v97 n4, pp.633-44

29. Handy, Susan L. 1992. Regional versus local accessibility: neo-traditional development and its impact for non-work travel. The Built Environment, v18, pp.253–267.

30. Heckman, James J. 1979. Sample selection bias as a specification error. Econometrica, v47, n1, pp.153-162.

31. Heckman, James J. 1980. Sample selection bias as a specification error: An application to the estimation of female labor supply functions, in James Smith, ed, Female Labor Supply. Princeton, NJ: Princeton University Press, pp.206-248.

32. Henderson, J. Vernon. 1985. Economic Theory and the Cities. Orlando, FL: Academic Press.

33. Herrin, William E. and Clifford R. 1992. Testing the Standard Urban Model of Residential Choice: An Implicit Markets Approach. Journal of Urban Economics, v31 n2, pp.145-63

34. Holtzclaw, J. 1990. Explaining urban density and transit impacts on auto use. Paper presented by National Resources Defense Council and the Sierra Club to the State of California Energy Resources Conservation and Development Commission, docket no. 89-CR-90, Sacramento.

35. Holtzclaw, J. 1994. Using residential patterns and transit to decrease auto dependence and costs. Unpublished manuscript. New York: National Resources Defense Council.

36. Ihlandfeldt, K. R. 1992. Job Accessibility and the Employment and School Enrollment of Teenagers. Kalamazoo, Michigan, W. E. Upjohn Institute for Employment Research

37. Kayser, Hilke. 2000. Gasoline Demand and Car Choice: Estimating Gasoline Demand Using Household Information, *Energy Economics* 22, pp.331-348

38. Khattak, A., and V. Amerlynck, "Are Travel Times and Distances to Work Greater for Residents of Poor Urban Neighborhoods?" Unpublished manuscript, 1999.

39. Kockelman, Kara M.. 1997. Travel Behavior as a Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from the San Francisco Bay Area. Transportation Research Record 1607, pp.117-125..

40. Lerman, Steven R. 1976. Location, housing, automobile ownership and mode to work: A joint choice model. Transportation Research Record.

41. Levinson, D., and Ajay Kumar.1997. Density and the Journey to Work. Growth and Change, v28, pp.147-72

42. Linneman, P. and P. Graves. 1983. Migration and Job Change: A Multinomial Logit Approach. Journal of Urban Economics v14, pp.263-279.

43. Madden, Janice. 1981. Why Women Work Closer to Home. Urban Studies, v18, pp.181-94

44. Mannering, Fred, and Clifford J. Winston, "A dynamic empirical analysis of household vehicle ownership and utilization," *RAND Journal of Economics* 16(2) (1985): 215-235

45. Massey, Douglas and Nancy Denton. 1988. The Dimensions of Residential Segregation. Social Forces, v67, n2, pp.281- 315.

46. Mattsson, Lars-Goran. 1984. Equivalence between welfare and entropy approaches to residential location. Regional Science and Urban Economics, v.14, pp.147-173.

47. McFadden, Daniel. 1974. The Measurement of Urban Travel Demand, Journal of Public Economics, v3, pp.303-328.

48. McFadden, Daniel. 1978. Modeling the Choice of Residential Location, in A. Karquist, et al (eds.), *Spatial Interaction Theory and Planning Models*. North Holland Publishing Company.

49. Muth, Richard E., 1969. Cities and Housing. The Spatial Pattern of Urban Residential Land Use. Chicago: The University of Chicago Press..

50. Nationwide Personal Transportation Survey (NPTS). 1995.

51. Newberry, D., "Road Damage Externalities and Road User Charges," *Econometrica* 56(2) (1988): 295-316.

52. Newman, Peter and Jeffrey Kenworthy. 1998. Sustainability and Cities: Overcoming Automobile Dependence. Island Press, Washington, DC.

53. Newman, Peter and Jeffrey Kenworthy. 1989. Cities and Automobile Dependence: An International Sourcebook. England, Aldershot.

54. O'Sullivan, Arthur. 2007. Urban Economics, 6th ed. McGraw-Hill

55. Pickrell, D., and P. Schimek, "Growth in Motor Vehicle Ownership and Use: Evidence for the Nationwide Personal Transportation Survey, Journal of Transportation and Statistics (1999) 1-7.

56. Pushkarev, Boris and Jeffrey Zupan. 1977. Public Transportation and Land Use Policy. Bloomington, Indiana: Bloomington University Press.

57. Rouwendal, Jan and E Meijer. 2001. Preferences for housing, jobs, and commuting: A mixed logit analysis. Journal of Regional Science, v41, pp.475–505.

58. Schimek, Paul. "Household Motor Vehicle Ownership and Use: How much does Residential Density Matter?" Transportation Research Record 1552 (1996) 120-125.

59. Sermons, M William and Frank S. Koppelman. 1998. A Factor Analytic Approach to Incorporating Systematic Taste Variation into Models of Residential Location Choice. Transportation Research Record, v1617, pp.194-202.

60. Sermons, M.William and N. Seredich. 2001. Assessing traveler responsiveness to land and location based accessibility and mobility solutions. Transportation Research, Part D, v6, pp.417-28.

61. Smith, W., 1984. Mass transit for high-rise, high-density living. Journal of Transportation Engineering, v110 n6, pp.521–535.

62. Steiner, Ruth. 1995. Residential Density and Travel Patterns: A Review of the Literature. Transportation Research Record 1466, pp.37-43.

63. Train, Kenneth E. 1986. Qualitative Choice Analysis: Theory, Econometrics and an Application to Automobile Demand, MIT Press.

64. Turnbull, Geoffrey. 1992. Location, Housing, and Leisure Demand under Local Employment. Land Economics, v68 n1, pp.62-71.

65. White, Michelle J. 1988. Location Choice and Commuting Behavior in Cities with Decentralized Employment. Journal of Urban Economics, v24, n2, pp.129-152.

66. Wilson, A. G. 1998. Land-Use/Transport Interaction Models: Past and Future .Journal of Transport Economics and Policy, v32 n1, pp.3-26