Transportation Research Part A 43 (2009) 180–196

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Neighborhoods, cars, and commuting in New York City: A discrete choice approach

Deborah Salon

Institute of Transportation Studies and the Department of Agricultural and Resource Economics, University of California, Davis 95616, United States

ARTICLE INFO

Article history:
Received 22 March 2007
Received in revised form 22 August 2008
Accepted 28 October 2008

Keywords:
Mode choice
Car ownership
Location choice

ABSTRACT

Cities around the world are trying out a multitude of transportation policy and investment alternatives with the aim of reducing car-induced externalities. However, without a solid understanding of how people make their transportation and residential location choices, it is hard to tell which of these policies and investments are really doing the job and which are wasting precious city resources. The focus of this paper is the determinants of car ownership and car use for commuting. Using survey data from 1997 to 1998 collected in New York City, this paper uses discrete choice econometrics to estimate a model of the choices of car ownership and commute mode while also modeling the related choice of residential location.

The main story told by this analysis is that New Yorkers are more sensitive to changes in travel time than they are to changes in travel cost. The model predicts that the most effective ways to reduce both auto ownership and car commuting involve changing the relative travel times for cars and transit, making transit trips faster by increasing both the frequency and the speed of service and making auto trips slower – perhaps simply by allowing traffic congestion. Population density also appears to have a substantial effect on car ownership in New York.
© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Heavy reliance on the private automobile for urban transportation causes substantial externalities, the most prominent being traffic congestion, air pollution, and, many would argue, loss of a sense of community. Recognizing this, urban planners and economists have repeatedly suggested investments and policies that encourage the use of alternatives to the private automobile for urban transportation. Cities both in the United States and around the world are trying out a multitude of transportation policy and investment alternatives with the aim of reducing car-induced externalities. However, without a solid understanding of how urban residents make their transportation and residential location choices, it is hard to tell which of these policies and investments are most likely to do the job and which will simply waste precious city resources.

This paper addresses the following question: What are the most effective policy levers to control car ownership and use in dense urban areas? To get at this question, I use the statistical framework of discrete choice econometrics to model the joint choice of residential location, car ownership, and commute mode. This model purposely incorporates as many variables that have clear policy relevance as possible, as well as individual characteristics of commuters as control variables. Although related work has been done, the present analysis is rare in that it focuses on both car ownership and car use while also endogenizing residential neighborhood choice. This is important because the choice of where to live is fundamentally linked to the choices of whether to own and use a car; analyses that do not explicitly model the joint nature of these decisions may produce biased results. The only previous research known to the author that jointly models the three decisions modeled here was published in 1976 (Lerman).

E-mail address: ddsalon@gmail.com

0965-8564/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tra.2008.10.002

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

The remainder of this paper is organized as follows. Section 2 presents a brief review of the existing literature on car ownership and use. Section 3 presents the data and research methodology. Section 4 introduces the estimated models and highlights specific policy-relevant findings, and Section 5 concludes with suggestions for future research directions in this area.

2. Existing literature on car ownership and use

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

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

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

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

Bhat and Pulugurta (1998) estimate discrete choice models of auto ownership for four metropolitan areas using data collected between 1987 and 1991. de Jong (1990) postulates and empirically estimates a model of the demand for cars and vehicle kilometers traveled by zero- and single-car households in the Netherlands using household survey data from 1985. Schimek (1996) uses a two-stage procedure to estimate jointly the demand for vehicles and the demand for vehicle kilometers traveled using 1990 household survey data from across the US.

While these models are based on disaggregate data, and therefore could potentially estimate the effects on car ownership and use of policy-sensitive variables, the authors do not include many such variables in their models. Schimek includes the policy-sensitive variables of transit availability and population density, while de Jong includes only the fixed and variable cost components of driving. Bhat and Pulugurta include dummy variables for suburban and urban residential neighborhoods.

On the car use side, there have been numerous studies on the topic of transport mode choice from all over the world. Findings from a few such studies are directly compared to those from the present study in the results section of this paper. Most of these previous studies of mode choice, however, take both car ownership and residential location to be exogenous to their modeling framework. In fact, there is only one study that I am aware of that did jointly model these decisions (Lerman, 1976), and it has been suggested by others that this should be done (Oum et al., 1992).

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

Studies that model mode choice jointly with either car ownership or residential location include Train (1980), Thobani (1984), and Ben-Akiva and Bowman (1998). Train estimated a nested logit model of the choices of car ownership and commute mode using 1975 data from the San Francisco Bay Area. Train's study includes many of the same policy-sensitive variables that are the focus of the present analysis. These include travel time and cost by transport mode, accessibility of the home to non-work destinations by transport mode, and the distance between the home location and the Central Business District. Train finds that many of these variables are statistically significant predictors of the choices of auto ownership and commute mode. Thobani follows Train to conduct a similar analysis for Karachi, Pakistan. Ben-Akiva and Bowman estimate a joint model of residential location choice and daily activity schedule choice using travel survey data from the Boston area in 1991. Their focus is on the impact of accessibility measures calculated from the activity model on the residential location model.

In the creation of the model presented in this paper, direction was taken from all of the studies cited here. There are, however, a number of differences between existing work and that presented here. These include contextual differences such as the year, the different physical context of New York compared with other cities, and the difference between the car ownership and use levels in the datasets. In Train's data, for instance, 93% of surveyed households owned a car, and 81% used a private car for their commute trip. The corresponding values for the 1997–1998 data from New York City are 61% and 30%, respectively. The other large difference between the literature since 1977 and the present paper is the inclusion in the model of the choice of residential location.

3. Methodology and data

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

In the real world, both changing one's residential neighborhood and changing one's car ownership status are highly costly activities in terms of both time and money. Because of these high transaction costs, many researchers have modeled commute mode choice as if residential neighborhood and car ownership status were exogenous variables. This approach often yields reasonable results, particularly in places where there is little variation in car ownership levels between households and in accessibility by non-auto modes throughout a region. New York City does not have these characteristics.

This paper presents the full results of a joint multinomial logit estimation of the choices of commute mode, car ownership status, and residential neighborhood. This model treats all choices as endogenous, modeling them as a single joint choice. The basis for the multinomial logit model is random utility theory (see Ben-Akiva and Lerman (1985) and Train (2003) for further details on logit model theory). Under this theory, it is assumed that each individual chooses the alternative that yields the highest payoff in terms of utility. Under this model, a utility function based on the attributes $\mathbf{x}_{nj}$ of J alternatives as well as the characteristics $\mathbf{s}_n$ of N individuals is postulated to have both deterministic and stochastic parts,

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

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

$$\begin{aligned}
P_{ni} &= Prob(U_{ni} > U_{nj} \ \forall j \neq i) \\
P_{ni} &= Prob(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \ \forall j \neq i) \\
P_{ni} &= Prob(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \ \forall j \neq i) \\
P_{ni} &= \int_{\epsilon} I(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \ \forall j \neq i) f(\epsilon_n) d\epsilon_n
\end{aligned}$$

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

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

where C_n represents the choice set for individual n.

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

Joint choice models have compound choice sets, meaning that each alternative in the choice set is composed of more than one sub-choice alternative. In the present model, each element of the compound choice set contains one commute mode alternative, one car ownership status alternative, and one residential location alternative as defined by a census tract. For example, one alternative is walk to work, own zero cars, live in census tract 23, and a separate alternative would be walk to work, own one car, live in census tract 23.

The choice set for the model estimated here has 6 commute mode alternatives, 3 car ownership status alternatives, and over 2000 residential census tract alternatives. Therefore, even though each sub-choice has a manageable number of alternatives, the full compound choice set is unmanageably large with nearly 40,000 alternatives. To reduce the choice set to one that is computationally manageable, this paper follows McFadden (1978). For each commuter, a random sample is taken of 9 of the residential location sub-alternatives that she did not choose. The residential location choice set for each commuter is then 10 possible locations: the location actually chosen plus the 9 randomly sampled not-chosen locations. For each commuter, then, the estimated compound choice set includes all 18 mode-car ownership combinations and 10 possible census tract locations, making a modeled choice set of 180 compound alternatives. The multinomial logit models in this paper were estimated using Stata 8.0 software.

3.1. Data

Incorporated into the present analysis are data extracted from seven separate data sources. The main data source is the Regional Travel – Household Interview Survey (RT-HIS), conducted in the fall of 1997 and the spring of 1998 by NuStats International and jointly commissioned by the New York Metropolitan Transportation Council and the North Jersey Transportation Planning Authority (NYMTC, 2000). The analysis in this paper is based on the 2621 commuters who completed the survey and reside within the five boroughs of New York City. Households completed both a 24-hour travel diary on an assigned day and a lengthy telephone interview that collected information about their socioeconomic situation, their residential location, and their travel habits. While this survey information is admittedly a bit dated, it remains the most recent set of in-depth travel survey data available for New York City, and serves as the basis for the city's own transportation modeling and planning efforts.

The RT-HIS dataset provides the individuals doing the choosing in the model – the dependent variables – as well as most of the independent variables used to explain the travel mode sub-choice and some of the independent variables used to explain the car ownership and residential location sub-choices. The rest of the independent variables in the model come from a variety of other data sources including the US Decennial Censuses of 1990 and 2000, the 1997 Business Patterns Census, the New York City Department of Finance (Community Cartography, 2004), the New York City Department of City Planning (2004), and New York City Transit (2005). All of these data sources are geographically referenced, and Geographic Information System (GIS) technology was used to merge the data from these disparate sources into a single dataset.

Some of the variables included in the model estimation are extracted directly from the data sources listed here. Others, however, were calculated from the raw data from these sources using statistical techniques and GIS software. Approximately 25% of NYC households in the RT-HIS refused to provide their incomes; an auxiliary regression was estimated to impute these missing values. Also, because surveyed individuals provided travel time and cost information only for the alternatives they chose (the trips they actually took), it was necessary to estimate the travel time and cost information for the alternatives not chosen. For this purpose, the network analyst function of GIS was used to calculate distances along New York's street network for commute trips from the neighborhoods not chosen to the work location. These distances were translated into travel times and costs using average speed and cost estimates that varied by time of day, mode, and origin and destination boroughs. These were estimated using speed and cost information from both commute and non-commute trips that were actually taken by the individuals surveyed.

There are two aspects of this dataset that make it particularly appropriate for the present analysis. First, it is large, making it possible to estimate a model with many parameters. Second, there is substantial variation in both car ownership and car use for commuting even within New York City, making parameter estimates robust. Table 1 summarizes the distribution of the choice of car ownership status and commute mode in the sample used for the estimation in this paper. Manhattan exhibits its extremely low car use for commuting, while more than half of Staten Island commuters use cars. Car ownership levels exhibit a similar spatial pattern.

3.2. Weighting the observations

In the case where a random sample has been collected, there is no need to weight the observations. On the other hand, in the case where a certain sub-population has been oversampled, using this dataset as if it were a random sample of the population can bias the results. The RT-HIS sample was obtained using a complex stratification scheme based on a combination

Table 1
Weighted shares of car ownership and commute mode in sample.

Residence location	NYC	Manhattan	Other Boros	Staten Island
# of HH (unweighted)	1865	803	580	512
# of HH (weighted)		372	1415	78
Car ownership				
0	38%	66%	32%	10%
1	42%	27%	46%	35%
2	21%	7%	22%	55%
# of trips (unweighted)	2621	1027	825	769
# of trips (weighted)		508	1990	123
Mode choice				
Walk/bike	8%	19%	5%	5%
Auto passenger	3%	5%	3%	6%
Auto driver	29%	9%	32%	56%
Bus	16%	13%	16%	21%
Subway with walk access	35%	43%	35%	4%
Subway with other access	9%	11%	9%	9%

of location and mode choices, and was not completely random. To obtain unbiased results from a non-random sample, it is necessary to weight the observations so that the weighting scheme in a sense "undoes" the weights of the stratification scheme, making the results representative of the underlying population.

The weighting scheme used here is based on residential location, and the methodology follows Manski and Lerman (1977). The weight for each neighborhood is the percent of the population in that neighborhood (according to the 2000 Census) divided by the percent of the sample in the neighborhood, as in the following equation:

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

If a neighborhood is represented in the sample exactly how it is represented in the population, the weight will be one. If the neighborhood is underrepresented (overrepresented) in the sample, the weight will be greater than (less than) one. These weights are used in the estimation by multiplying each of the probabilities P_{ni} by the neighborhood weight for that individual, and using these weighted probabilities to create the joint probability function to be maximized. The neighborhoods used are the 195 groups of census tracts identified as "neighborhoods" by the New York City Department of City Planning in its 2007 PlaNYC report. These neighborhoods are used for weighting purposes only; the model is estimated using census tracts as the residential location alternatives.

3.3. Model selection part 1: joint choice vs. individual choices

This section asks whether it is important to model of the choices of residential location, car ownership status, and commute mode in a single joint choice model, or whether separate models of these three choices would suffice. This is a particularly valuable question to answer because separate choice models are far simpler to estimate than joint choice models. To answer this question, seven multinomial logit choice models were estimated that include all of the possible sub-models of the full joint choice models:

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

The estimation results for all of these models are not available in this paper due to space constraints, but are available from the author upon request. The three that are included here are the full model and the indiviudal choices of commute mode and car ownership, and they are reported in Tables 3, A.1, and A.2. To assess the importance of modeling these choices jointly, I compare each model's predicted probabilities for the alternatives that were actually chosen. This is done by comparing the average predicted probability for the chosen alternative in the joint choice model with the product of the average predicted probabilities for the chosen sub-alternatives from the single choice models (see Table 2).[1]

This comparison method is relatively simple. First, the joint choice model is estimated. Then, I take the average of the predicted probabilities for the chosen alternatives. Since the model is estimated using neighborhood weights, the averages here are weighted as well, using this same weighting scheme. For the comparison, it is necessary to also estimate the single choice models for each sub-choice, and calculate the weighted average of these predicted probabilities for the chosen sub-alternatives. The goodness-of-fit comparison is between the weighted average probability for the compound alternative and the product of the weighted average probabilities for the sub-alternatives. The following mathematical expression represents the comparison:

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

where $P_n(\cdot)$ is the weighted probability the individual n chooses alternative $(\cdot)$; l is the location choice, c is the car ownership choice; m is the mode choice; $y_{nj} = 1$ if individual n chooses compound alternative j; $y_{nj} = 0$ otherwise; and $y_{nl}, y_{nc}, and\, y_{nm}$ are defined in an analogous manner.

As shown in Table 2, the joint choice models perform better for both the full compound choice case and for the location-mode choice case. For the car-mode and the location-car choice cases, however, the separate models perform better than the joint choice model. This indicates that the car ownership choice portion of the model is not performing well and that it is reducing the performance of the other modeled choices. However, the current research is focused on car ownership status

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

Table 2
Goodness-of-fit comparison.

Model	Joint	Product of separate
Weighted average predicted probabilities for the chosen alternative		
Full compound choice	0.061	0.053
Location-mode compound choice	0.126	0.093
Car-mode compound choice	0.195	0.215
Location-car compound choice	0.117	0.124

as well as car use for commuting, and it is therefore important to be able to test hypotheses regarding car ownership choice. For this reason, I chose to continue to include as endogenous the choice of car ownership status in the present model.

3.4. Model selection part 2: nested vs. multinomial logit

One limitation of the multinomial logit model is that it assumes that the model satisfies the Independence of Irrelevant Alternatives (IIA) assumption. This assumption is described best by example. Suppose that a commuter chooses between walk, car, bus, and subway for her mode of transport. If any of these alternatives became unavailable (for instance, if the commuter sprained her ankle and could no longer walk), then the probabilities of the other alternatives would necessarily increase. The IIA assumption is that the probabilities of the remaining alternatives would increase by the same proportion. If, however, there is some difference in the proportional increase in probabilities, the IIA assumption is violated. If the subway alternative were removed, for instance, one might expect that a disproportionate percent of the probability of choosing subway for a given individual might be allocated to bus because these two alternatives are both transit. This would be a violation of the IIA assumption, and it occurs because the unobserved utility (the model's error term) is correlated between the alternatives of subway and bus.

As in this example, violation of the IIA assumption can happen even in a single choice model, but it is especially likely in a joint choice situation. For instance, in the present application, it makes sense that there would be some correlation among the mode sub-alternatives that all have the same residential location and car ownership status. If walk to work, own one car, live in census tract 23 were removed because of a sprained ankle, it would be disproportionately likely that the commuter would choose drive to work, own one car, live in census tract 23, rather than any alternative that would require her to change car ownership status or residential location. It is important to note here that the IIA assumption is violated only if the correlation between alternatives is not accounted for by the explanatory variables in the model.

One way to relax the IIA assumption is to estimate a nested logit model, rather than a multinomial logit model. The nested logit allows for structured correlation across the unobserved utility of a subset of the alternatives in the choice set. These subsets of the alternatives are the "nests". Within each nest, the alternatives are assumed to be closer substitutes for each other than they are for the alternatives outside of the nest, and inclusive value parameters are estimated that indicate the extent to which this assumption is true. Inclusive value parameter estimates between zero and one indicate greater substitution between alternatives within a nest than between alternatives in different nests. Estimates that are not significantly different from one indicate no difference between the nested and the multinomial logit specification. Estimates that are negative or greater than one are usually interpreted to be inconsistent with random utility theory. For details on nested logit model specification, see Ben-Akiva and Lerman (1985) or Train (2003). For a clear exposition of the interpretation of inclusive value parameter estimates, see Train et al. (1987), Bosang (1999), or Gangrade et al. (2002).

For each of the model specifications presented here that endogenizes two of the sub-choices, two nested versions of the model (one with each of the sub-choices as the upper level of the model) were estimated in addition to the non-nested version, and results were compared. In most of the nested versions of these models, the estimates of the inclusive value coefficients are largely either not significantly different from one or were substantially larger than one. In all nested logit estimations with inclusive value estimates between zero and one, the elasticity results were not substantively different from those obtained in the joint multinomial logit estimations.

A three-level nested logit specification was not estimated. Due to the size of the dataset used here and the number of variables in the model, it is no trivial matter to estimate a three-level nested model.[2] Furthermore, from the order of the levels in the two-level versions that yielded inclusive value estimates between zero and one, it was not clear how to specify such a model. Since the elasticity results were not substantively different between nested and non-nested versions of the two-level models, I take this as evidence that the overall results of the multinomial logit model presented in this paper are robust.

3.5. Limitations of the model

The present analysis is limited by a couple of simplifications of the choice framework. Multiple-worker households are not modeled differently from single-worker households, even though the relationship in a multiple-worker household between residential neighborhood choice and travel choices is likely to involve a compromise between the workers. The simplifying

[2] Two major statistical software packages were unable to do so (NLogit 3.0 and Stata 8.0), and the matrix programming language that was used (GAUSS) to estimate the two-level nested versions of the model was unable to load the full dataset.

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

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

4. Results

Table 3 presents the estimated coefficients for the multinomial logit model of the joint choice of residential location, car ownership status, and commute mode. Tables 4 and 5 present the elasticities and marginal effects calculated using the model results.

This section begins by describing how the explanatory variables were chosen for the model and how to properly interpret the estimated coefficients. Then, I offer an interpretation of the model's estimated coefficients, elasticities, and marginal effects. The elasticity results are compared to those found in similar studies in the literature.

4.1. Explanatory variables

Explanatory variables included in the model were chosen based on a combination of economic theory and data availability. Variables that influence commute mode choice and car ownership status are meant to represent the relative "prices" of the alternatives in money, time, and convenience. Variables that influence residential location are meant to capture the relative attractiveness of neighborhoods in terms of attributes such as cost, transit access, and local availability of services. Additional variables that influence residential location choice include characteristics of the inhabitants of each location. As will be immediately apparent from examination of Table 3, there are some included variables that are not statistically significant. These variables remain in the model because they were statistically significant in alternative model specifications and/or there is theoretical basis for their inclusion.

Many of the coefficients in the model are estimated separately for low- and high-income commuters, and some are estimated separately for commuters with children and for other subpopulations. Segmenting the model in this way explicitly allows for some structured heterogeneity of preferences.

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

Many of the independent variables are interaction variables, and they should be interpreted according to the following examples. The generic variables have the most intuitive interpretations. A variable is generic if it can be measured for all alternatives. A generic variable in the current model is "Travel Cost", and its negative sign for both low- and high-income commuters indicates that as commute cost for any alternative rises, the utility of that alternative falls.

Coefficients on alternative-specific variables are interpreted to have meaning only for the alternative specified, and only relative to the utility of the omitted alternative. By assumption, the coefficient for the omitted alternative is zero. For instance, the negative coefficient on "Subway Lines At Work for Bus" means that as the number of subway lines near the workplace rises, the utility of the bus alternatives goes down *relative to the omitted auto mode alternatives*. In another example, the positive coefficient on "Household Size for Two or More Cars" means that as the household size rises, the utility of having two or more cars in the household rises *relative to the omitted zero car alternatives*.

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

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

In addition to the coefficients that are listed in Table 3, the model also includes 53 alternative-specific constants. These are dummy variables that serve the purpose of normalizing the model so that it will be sure to reproduce the sample shares of the actual choices of the sample.[3]

[3] Usually, there are $J-1$ alternative-specific constants (ASCs), where J is the total number of alternatives in the model. However, the current model includes only 10 of the approximately 2000 residential location alternatives for each commuter. To avoid the impossibility of estimating approximately 40,000 ASCs, the residential location alternatives are aggregated into three groups for the purpose of including ASCs: Manhattan, Staten Island, and the Rest of the City. The estimated number of ASCs is therefore 53: 6 mode alternatives times 3 car ownership alternatives times 3 aggregate residential location alternatives minus one. This simplification is only made for the calculation of the ASCs; the residential location alternatives in the model are census tracts.

Table 3
Multinomial logit model of the full joint choice of residential location, car ownership status, and commute mode.

Commute mode choice variables	Coefficient	SE	Coefficient	SE
	Income < $25,000 per HH member		Income > $25,000 per HH member	
Travel cost	−0.128***	0.012	−0.118***	0.014
Walking time	−1.280***	0.090	−1.392***	0.121
Waiting time	−12.442***	2.365	−16.449***	2.621
Riding time	−1.827***	0.086	−2.186***	0.144
Not segregated by income				
Long walk (>10 min)	−0.406***	0.087		
Subway lines at home for walk	0.221***	0.079		
Subway lines at home for bus	0.028	0.083		
Subway lines at home for subway	0.106	0.067		
Subway lines at work for walk	−0.017	0.040		
Subway lines at work for bus	−0.143***	0.033		
Subway lines at work for subway	0.038	0.024		
Bus lines at home for walk	0.004	0.012		
Bus lines at home for bus	−0.005	0.011		
Bus lines at home for subway	0.000	0.010		
Bus lines at work for walk	−0.021*	0.011		
Bus lines at work for bus	−0.001	0.007		
Bus lines at work for subway	0.019***	0.006		
Car ownership status choice variables				
Income for one car	0.215**	0.106	0. 276***	0.062
Income squared for one car	0.003	0.013	−0.009**	0.004
Income for two or more cars	0.735***	0.135	0. 722***	0.083
Income squared for two or more cars	−0.027*	0.014	−0. 029***	0.005
Not segregated by income				
Household size for one car	0.025	0.043		
Household size for two or more cars	0.330***	0.052		
Subway lines at home for one car	−0.093*	0.048		
Subway lines at home for two or more cars	−0.127	0.078		
Subway lines at work for one car	0.034*	0.018		
Subway lines at work for two or more cars	−0.008	0.025		
Bus lines at home for one car	−0.015	0.009		
Car ownership status choice variables, cont				
Bus lines at home for two of more cars	0.003	0.011		
Miles to Midtown Manhattan for one car	0.061***	0.021		
Miles to Midtown Manhattan for two or more cars	0.070***	0.025		
Retail density for one car	0.661**	0.273		
Retail density for two or more Cars	−0.677	0.484		
Employment density for one car	−0.006**	0.003		
Employment density for two or more cars	−0.001	0.004		
Population density (L) for one car	0.063	0.021		
Population density (L) for two or more cars	0.063	0.021		
Population density (H) for one car	−0.004***	0.002		
Population density (H) for two or more cars	−0.014***	0.002		
Residential location choice variables				
Housing price per income	−0.830***	0.187	−3. 741***	0.577
Median income	−0.118***	0.043	0.060	0.046
Percent vacant and industrial land	−0.687**	0.314	−0. 864**	0.423
Percent college-educated	−0.079	0.417	2.559***	0.430
Not segregated by income				
Percent college-educated if kids in HH	1.439***	0.477		
Average number of building stories	−0.079***	0.028		
Subway lines at home	−0.164**	0.069		
Bus lines at home	0.023***	0.007		
Miles to Midtown	0.052***	0.019		
Miles to subway station	0.086**	0.041		
Percent owner-occupied	−1.463***	0.267		
Percent owner-occupied if homeowner	3.240***	0.238		
Percent non-white if white HH	−1.930***	0.156		
Percent white if non-white HH	−2.472***	0.173		
Residential location choice variables, cont.				
Percent under 18	−4.206***	0.614		
Percent under 18 if kids in HH	6.004***	0.932		
Percent married households	1.865***	0.274		
Employment density	−0.001	0.002		
Retail density	−0.190	0.199		

(*continued on next page*)

Table 3 (*continued*)

Commute mode choice variables	Coefficient	SE	Coefficient	SE
	Income < $25,000 per HH member		Income > $25,000 per HH member	
Retail density if kids in HH	−0.150	0.242		
Population density (L)	−0.062***	0.020		
Population density (H)	0.012***	0.001		
Population density (L) if kids in HH	−0.040***	0.013		
Population density (H) if kids in HH	0.000	0.002		
Plus alternative-specific constants[a]				
Estimation summary information				
Observations	2621			
Alternatives[b]	180			
Pseudo R^2	0.297			

There are 53 alternative specific constants in this model, representing all combinations of commute mode and car ownership, and three residential location groups (Manhattan, Staten Island, and the Rest of New York City).
The 180 compound alternatives consist of 6 mode alternatives, 3 car ownership status alternatives, and 10 census tract alternatives sampled from the full set of over 2000 possible census tracts.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

Table 4
Elasticities of car ownership and car use for commuting in Full Joint Model.

	Car use	Zero car	One car	Two+ car
Five boroughs of New York City				
Population density (home)	−0.15	0.31	−0.03	−0.39
Car commute cost (incl. parking)	−0.33	0.11	−0.03	−0.11
Non-car commute cost	0.10	−0.03	0.01	0.03
Car commute time	−0.50	0.18	−0.04	−0.19
Non-car commute time	1.22	−0.39	0.10	0.40
Walking time	0.20	−0.08	0.02	0.08
Waiting time	0.53	−0.18	0.05	0.17
Riding time	0.48	−0.13	0.03	0.15
Income	n/a	−0.58	0.15	0.60
Manhattan only				
Population density (home)	−0.23	0.34	−0.15	−0.75
Car commute cost (incl. parking)	−0.79	0.11	−0.06	−0.20
Non-car commute cost	0.13	−0.01	0.01	0.02
Car commute time	−0.70	0.10	−0.05	−0.22
Non-car commute time	1.58	−0.15	0.08	0.30
Walking time	0.26	−0.04	0.01	0.09
Waiting time	0.73	−0.10	0.07	0.13
Riding time	0.58	−0.02	0.00	0.07
Income	n/a	−0.46	0.37	0.56
Staten Island only				
Population density (home)	−0.05	0.28	0.06	−0.16
Car commute cost (incl. parking)	−0.23	0.16	0.01	−0.07
Non-car commute cost	0.07	−0.04	0.00	0.02
Car commute time	−0.38	0.27	0.00	−0.10
Non-car commute time	0.94	−0.68	−0.04	0.29
Walking time	0.17	−0.11	−0.02	0.06
Waiting time	0.38	−0.27	−0.01	0.11
Riding time	0.38	−0.29	−0.01	0.12
Income	n/a	−0.95	−0.03	0.39
Rest of New York city				
Population density (home)	−0.14	0.30	−0.02	−0.37
Car commute cost (incl. parking)	−0.28	0.10	−0.02	−0.10
Non-car commute cost	0.10	−0.04	0.01	0.03
Car commute time	−0.49	0.20	−0.05	−0.19
Non-car commute time	1.19	−0.47	0.11	0.43
Walking time	0.20	−0.10	0.03	0.08
Waiting time	0.52	−0.20	0.05	0.18
Riding time	0.47	−0.17	0.03	0.16
Income	n/a	−0.62	0.11	0.62

Table 5
Non-marginal effects of car ownership and car use for commuting in Full Joint Model (units are percentage points).

	Car use	Zero car	One car	Two+ car
Five boroughs of New York city				
Home population density (plus 5470 people/sq. mile)	−0.35	0.61	0.22	−0.83
Home subway lines (plus one line)	−2.36	2.66	−1.24	−1.42
Car commute cost (plus 15 cents)	−0.31	0.12	−0.04	−0.08
Non-car commute cost (plus 15 cents)	0.26	−0.10	0.03	0.06
Car commute time (plus 2.6 min)	−1.35	0.53	−0.19	−0.34
Non-car commute time (plus 3.8 min)	2.97	−1.08	0.36	0.72
Walk time for walkers (plus 1.5 min)	0.07	−0.04	0.02	0.02
Walk time for transit riders (plus 45 s)	0.22	−0.08	0.03	0.05
Wait time for transit riders (plus 30 s)	1.56	−0.56	0.19	0.38
Ride time for transit riders (plus 2.6 min)	1.18	−0.43	0.14	0.29
Income (plus $4250)	n/a	−2.25	0.58	1.67
Manhattan only				
Home population density (plus 5470 people/sq. mile)	−0.22	0.72	−0.07	−0.65
Home subway lines (plus one line)	−1.72	2.74	−1.61	−1.13
Car commute cost (plus 15 cents)	−0.21	0.08	−0.04	−0.04
Non-car commute cost (plus 15 cents)	0.18	−0.07	0.04	0.03
Car commute time (plus 2.6 min)	−1.01	0.39	−0.19	−0.20
Non-car commute time (plus 3.8 min)	2.25	−0.82	0.50	0.33
Walk time for walkers (plus 1.5 min)	0.06	−0.03	0.00	0.03
Walk time for transit riders (plus 45 s)	0.16	−0.06	0.03	0.02
Wait time for transit riders (plus 30 s)	1.18	−0.43	0.26	0.17
Ride time for transit riders (plus 2.6 min)	0.87	−0.32	0.19	0.13
Income (plus $4250)	n/a	−2.04	1.09	0.95
Staten Island only				
Home population density (plus 5470 people/sq. mile)	−0.14	−0.03	0.74	−0.72
Home subway lines (plus one line)	−2.40	1.77	−0.06	−1.72
Car commute cost (plus 15 cents)	−0.35	0.08	0.01	−0.10
Non-car commute cost (plus 15 cents)	0.29	−0.08	−0.00	0.08
Car commute time (plus 2.6 min)	−1.60	0.38	0.05	−0.44
Non-car commute time (plus 3.8 min)	3.26	−0.83	−0.04	0.87
Walk time for walkers (plus 1.5 min)	0.10	−0.02	−0.01	0.03
Walk time for transit riders (plus 45 s)	0.25	−0.06	0.00	0.07
Wait time for transit riders (plus 30 s)	1.73	−0.44	−0.02	0.46
Ride time for transit riders (plus 2.6 min)	1.31	−0.33	−0.02	0.35
Income (plus $4250)	n/a	−1.50	−0.25	1.75
Rest of New York city				
Home population density (plus 5470 people/sq. mile)	−0.39	0.62	0.26	−0.88
Home subway lines (plus one line)	−2.52	2.70	−1.22	−1.48
Car commute cost (plus 15 cents)	−0.32	0.13	−0.05	−0.08
Non-car commute cost (plus 15 cents)	0.28	−0.11	0.03	0.07
Car commute time (plus 2.6 min)	−1.42	0.58	−0.20	−0.37
Non-car commute time (plus 3.8 min)	3.14	−1.16	0.35	0.81
Walk time for walkers (plus 1.5 min)	0.07	−0.04	0.03	0.02
Walk time for transit riders (plus 45 s)	0.24	−0.09	0.03	0.06
Wait time for transit riders (plus 30 s)	1.64	−0.60	0.18	0.42
Ride time for transit riders (plus 2.6 min)	1.25	−0.46	0.14	0.32
Income (plus $4250)	n/a	−2.35	0.50	1.85

4.2. Interpretation of the estimated coefficients

Most of the statistically significant coefficients in the commute mode choice category of Table 3 have the expected signs. Higher travel costs and travel times lower the utility of the alternative for both high- and low-income commuters. Any alternative that includes a walk longer than 10 minutes is additionally undesirable. For the alternative-specific variables that represent transit access, all of the coefficients should be interpreted as being relative to the utility of the omitted auto mode alternative. Where there are more subway lines near work, the bus mode alternative becomes less attractive than all of the other mode options. Where there are more subway lines near home, there is a positive effect on the utility of walking relative to the other modes. This makes sense because areas with the highest number of subway lines in New York City are also areas with the highest walk accessibility.

In the car ownership choice category, again most of the signs on the statistically significant coefficient estimates are as expected. Higher income increases the utility of car-owning alternatives, and the effect shrinks as income rises (as evidenced by the negative coefficient on the squared terms). For both income categories, higher incomes have a stronger effect on owning two or more cars than on owning one car. Living farther from midtown Manhattan raises the utility of owning a car, and commuters from larger households have a higher utility of car ownership. Within the high population density range (more than 20,000 people per square mile), higher density lowers the utility of owning a car.

In the residential location choice category of variables as well, most of the statistically significant signs on the estimated coefficients make intuitive sense. Higher housing cost reduces the utility of a location, and physically undesirable characteristics such as tall buildings and vacant or industrial land also reduce the utility of a location. Within the lower population density range, higher densities reduce the attractiveness of a location. The opposite is true within the higher population density range, suggesting that there is a mid-range population density that is particularly undesirable. All else equal, it appears that New Yorkers would rather live farther from midtown Manhattan.

The estimated effects of neighbor characteristics on the attractiveness of a location describe a world in which people gravitate toward neighborhoods where their neighbors are similar to them. More educated neighbors increase the utility of a location for both higher income households and for households with children. A higher neighborhood percentage of people who are racially different from the commuter's household reduces the utility of the alternative. A higher percent of married neighbors raises the utility of a location. More children in the neighborhood reduce its attractiveness, but this effect is reversed for households with children. Likewise, a higher percentage of homeowners reduces a neighborhood's appeal, but this effect is reversed for homeowning households.

There are a few counterintuitive signs on the estimated coefficients related to residential location choice, however. For instance, all else equal, this model indicates that New Yorkers would rather live in neighborhoods with fewer subway lines and be farther from the subway stations. Lower income households appear to prefer living in neighborhoods with lower median incomes. One possible explanation for these findings is that the housing price data may be an imperfect representation of what RT-HIS survey respondants actually paid for housing, these other variables are correlated with the real housing price, and therefore they appear to have a negative effect on the attractiveness of a neighborhood.

4.3. Elasticities

Table 4 presents the elasticities of car ownership and use for commuting with respect to a number of variables in the estimated full joint choice model. The elasticities are the percent change in the probability of choosing a particular alternative when an independent variable is increased by 1%. Although they are not identical to demand elasticities, these elasticities can be interpreted in much the same way – these elasticities are the percent change in the market share (similar to demand) of the particular alternative when an independent variable is increased by 1%.

Calculation of appropriate elasticities from discrete choice models is not, however, a trivial matter. All of the elasticities presented in this paper are calculated in the following way. First, the coefficients that parameterize the model are estimated based on the actual data, and the weighted predicted probabilities for each alternative are calculated for each individual. These predicted probabilities are represented by $wtp0_{nj}$ in the equations that follow. Second, the independent variable for which the elasticity is being calculated is increased by 1%. Third, the predicted probabilities are recalculated. These predicted probabilities are represented by $wtp1_{nj}$ in the equations that follow. Note that the model is not re-estimated, rather the existing model estimates are used to predict new probabilities based on the altered underlying data. Fourth, both the original and the new predicted probabilities are summed over the alternatives that contain the relevant sub-choice. Fifth, the percent change in the probability of choosing the relevant alternative is calculated for each individual, represented as the individual elasticity estimates ϵ_n in the equations below. Finally, the individual elasticity estimates are averaged across the sample, weighted by the original probability for each individual of choosing the alternative.[4] The final elasticities are given by ϵ.

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

$$\epsilon = \sum_n \frac{\sum_{j\in J} wtp0_{nj}}{\sum_n \sum_{j\in J} wtp0_{nj}} \epsilon_n$$

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

The elasticities are shown for the entire sample and then separately for Manhattan residents, Staten Island residents, and the residents of the other boroughs. These subsample elasticities were calculated by extracting the subset of the sample that actually chose to live in each location, and calculating the probability-weighted elasticities for each of these subsamples. Note that the borough-level elasticities are calculated from the model estimated using the entire sample, and that, as is evident in Table 3, separate coefficients were not estimated for each of the city's subregions. By not estimating different coefficients for each area of the city, I assume that preferences are similar across the city, after controlling for commuter

[4] This weighting is necessary because, for example, a change from a 1% probability to a 2% probability will appear as a 100% increase in the probability, but actually represents almost zero change in the likelihood of choosing that alternative.

socioeconomic characteristics.[5] The differences in choice behavior (see Table 1) are assumed to come from differences in the transportation-land use contexts across city boroughs.

Overall, most of the elasticity estimates make intuitive sense and are broadly consistent with the range of estimates found in the literature, with the Staten Island elasticity estimates being closest to estimates of elasticities in less dense cities. This in itself is interesting because one might expect these numbers to be more different due to the often-cited uniqueness of the land use and transportation system in New York City. This consistency with existing literature is encouraging for the generalizability of the results from the current model.

The income elasticity of car ownership and the travel cost elasticity of car use are the numbers most commonly found in the literature, and they are used here for comparison purposes. Using aggregate data, Ingram and Liu (1997) found the income elasticity in global cities to be 0.5 for all levels of car ownership. Using household-level data, Schimek (1996) estimated this same number to be 0.221 for the US, and de Jong (1990) found a value of 0.42 for the Netherlands. Using a discrete choice framework, Bhat and Pulugurta (1998) find for Boston in 1991 that the income elasticities of car ownership were −0.938 for zero-car households, −0.189 for one-car households, and 0.281 for two-car households. These car ownership elasticities from discrete choice analysis are estimated separately as the effect on the probability of being in a zero-car household, being in a one-car household, and being in a household with two or more cars. Recall that these elasticities are weighted by the original probabilities. This means that the elasticity estimate – for example – for a zero-car household is dominated by the changes in probability that zero-car households experience when faced with a 1% increase in the variable of interest.

The elasticities of car ownership presented here are broadly consistent with these previous estimates. In direct comparison to Bhat and Pulugurta's work, they indicate that Staten Islanders are similar to Bostonites. The relevant comparison with Bhat and Pulugurta's work is found in Table A.4, which presents the elasticities of car ownership calculated from a model of the choice of car ownership only. These elasticities of car ownership with respect to income on Staten Island are −1.06, −0.33, and 0.30 for zero-, one-, and two-or-more-car households, respectively.

Zhang (2004) finds that the travel cost (including parking) elasticity of car use is −0.144 in Boston, and −0.242 in Hong Kong. These findings of Zhang are comparable to the results presented in Table A.3 of −0.24 for Staten Island, and −0.29 for the whole of New York City for travel cost elasticity of car use. Additional estimates of travel cost elasticity of car use from the literature include Asensio's (2002) estimate of −0.092 for Barcelona and Hensher and Ton's (2000) estimates of less than −1 for Sydney and Melbourne.

The elasticity estimates presented here that are distinct from estimates in the literature are those with respect to population density. Using his mode choice only model, Zhang (2004) finds that the population density elasticity of car use is −0.044 in Boston and −0.039 in Hong Kong, and Schimek (1996) finds this elasticity to be −0.069. In the models estimated for this paper, population density is not an explanatory variable in the mode choice segment of the model. However, for comparison purposes to Zhang's study, a mode choice only model run was conducted with population density for the auto modes as an explanatory variable. The calculated car use elasticity with respect to population density from this model is −0.14. Schimek finds that the population density elasticity of car ownership is −0.057. In contrast to this relatively modest estimate, the estimates presented in both Table 4 and the elasticity tables in Appendix A are surprisingly large.

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

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

4.4. Parking cost – the missing link?

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

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

car owners with high incomes will be more likely to choose off-street parking. There is a large middle income bracket, however, where car owners will sometimes choose on- and sometimes choose off-street parking.

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

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

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

4.5. Non-marginal effects

Coefficient estimates and elasticities are useful in providing validation for the model, but provide less-than-complete information for someone who seeks to understand the policy implications of this work. Elasticities tell us the predicted percent change in the dependent variable for a given change in a policy variable. As such, elasticities are useful indicators if both the expected change in the policy variable and the expected effect on choice behavior are proportional to the current levels. When this criterion is not met, indicators such as marginal and non-marginal effects can be more useful.

Table 5 presents the predicted non-marginal effects of changes in selected independent variables. These values indicate the percentage point change in the likelihood of choosing an alternative, given a non-marginal absolute change in the independent variable. For the subway line variables, the non-marginal change evaluated is the addition of one line. For the remainder of the variables, the non-marginal changes are 10% of that variable's median value, and the values of these changes are noted in Table 5. The effects were calculated according to the following equation:

$$Non-marginal\ effect = \frac{\sum_n (wtp1_{nj} - wtp0_{nj})}{N}$$

As is clear from a comparison of Table 5 with Table 4, the estimated elasticities and non-marginal effects tell a similar overall story about the factors that affect car ownership and use for commuting.

5. Policy implications and concluding remarks

This paper has identified effective policy levers to influence car ownership and use in New York City using a discrete choice econometric model of car ownership and use for commuting. The model endogenizes the choice of residential neighborhood, and purposefully includes more policy-sensitive variables than have those in previous studies. Broad consistency of the results with both economic theory and with the findings of previous studies is heartening for the possibility of generalizing these results to other cities using a simulation approach. This is an area for future research. This section elaborates on the implications of the model results for transport and land use policy, and describes how these results could be combined with policy cost and co-benefit information to provide a full evaluation of policy alternatives.

The main story told by this analysis is that New Yorkers are more sensitive to changes in travel time than they are to changes in travel cost. The model predicts that the most effective ways to reduce both auto ownership and car commuting involve changing the relative travel times for cars and transit, making transit trips faster by increasing both the frequency and the speed of service and making auto trips slower – perhaps simply by allowing traffic congestion. The largest elasticities and marginal effects of car use for commuting are with respect to Non-Car Commute Time, and the largest portion of this effect is from Waiting Time and Riding Time (see Tables 4 and 5).[6] These indicate that the elasticity of car use for commuting with respect to Non-Car Commute Time is approximately one, meaning that a 1% decrease in Non-Car Commute Time will yield a 1% decrease in the car mode share for commuting. The estimated effect of Car Commute Time is similar in magnitude (but opposite in sign) to the effect of Riding Time for transit riders. Table 5 also shows that improving transit by adding subway lines near the commuter's home location has a substantial effect on both commute mode choice and car ownership. This model esti-

[6] One might expect that the effect of car commute time changes would be similar and opposite of the effect of non-car commute time changes. A quick examination of Tables 4 and 5 will show that although the effect of increasing car commute time does decrease the probability of car use for commuting, the effect is consistently smaller than the effect of decreasing non-car commute time. This is because individuals consider not only the total time a trip takes, but also the amount of that time that is spent walking and waiting. For car commute time, the total trip time is simply the riding time. For non-car commute time, however, the total trip time is the sum of the three categories of time: walking, waiting, and riding time. Tables 4 and 5 contain the decomposition of these effects of travel time on car ownership and use for commuting.

mates that an additional subway line will reduce the commute mode share of cars by an average of 2.4 percentage points, and increase the share of zero-car households by 2.7 percentage points. As discussed above, population density also appears to have a substantial effect on car ownership, but this may be due to an omitted variable such as home parking cost.

For policymakers, the costs as well as the co-benefits of any policy must be considered along with its effectiveness in achieving a social goal. Some policy-relevant variables are much more costly to change than others, and some policies have non-monetary benefits and/or costs beyond reducing car ownership and use. To use the results of analyses such as that presented here to inform policy, they must be used together with monetary cost and non-monetary cost and benefit information to compare competing policy alternatives.

Acknowledgements

I would like to thank my dissertation committee members James Wilen, Daniel Sperling, and Susan Handy as well as participants in the PASI-TS workshop Francisco Martinez, Marcela Munizaga, and Jose Holguin-Veras for their encouragement and helpful comments on early drafts of this paper. In addition, I am grateful for the excellent suggestions from two anonymous reviewers, and for the support of both the Earth Institute and the Spatial Information Design Lab at Columbia University. This work was supported by a dissertation grant from the University of California Transportation Center as well as an Eisenhower Graduate Fellowship from the US Department of Transportation. Any and all errors are, of course, my own.

Appendix A

See Tables A.1–A.4.

Table A.1
Multinomial logit model of the Choice of Commute Mode.

Commute mode choice variables	Coefficient	SE	Coefficient	SE
	Income < $25,000 per HH member		Income > $25,000 per HH member	
Travel cost	−0.123***	0.013	−0.121***	0.016
Walking time	−0.750***	0.088	−0.985***	0.115
Waiting time	−13.448***	2.500	−16.814***	2.698
Riding time	−0.046	0.140	−0.869***	0.208
Not segregated by income				
Long walk (>10 min)	−0.691***	0.114		
Subway lines at home for walk	0.213**	0.085		
Subway lines at home for bus	0.035	0.088		
Subway lines at home for subway	0.106	0.074		
Subway lines at work for walk	0.012	0.041		
Subway lines at work for bus	−0.130***	0.033		
Subway lines at work for subway	0.050**	0.024		
Bus lines at home for walk	0.003	0.015		
Bus lines at home for bus	−0.009	0.013		
Bus lines at home for subway	0.015	0.011		
Bus lines at work for walk	−0.012	0.012		
Bus lines at work for bus	0.010	0.007		
Bus lines at work for subway	0.028***	0.006		
One car HH for auto	2.708***	0.252		
Two or more car HH for auto	3.388***	0.323		
One car HH for transit	−0.132	0.187		
Two or more car HH for transit	−0.282	0.284		
Staten Island HH for auto	−0.348	0.501		
Staten Island HH for transit	−0.991*	0.521		
Manhattan HH for auto	−0.568**	0.250		
Manhattan HH for transit	−1.032***	0.204		
Plus alternative-specific constants[a]				
Estimation summary information				
Observations	2621			
Alternatives[b]	6			
Pseudo R^2	0.318			

[a] *There are 5 alternative specific constants in this model, representing the commute mode alternatives.*
[b] There are 6 mode alternatives in this model.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

Table A.2
Multinomial logit model of the choice of car ownership status.

Car ownership status choice variables	Income < $25,000 per HH member		Income > $25,000 per HH member	
Income for one car	0.083	0.119	0.178**	0.070
Income squared for one car	0.021	0.014	−0.000	0.004
Income for two or more cars	0.594***	0.153	0.566***	0.095
Income squared for two or more cars	−0.007	0.017	−0.015***	0.006
Not segregated by income				
Household size for one car	−0.004	0.046		
Household size for two or more cars	0.285***	0.059		
Subway lines at home for one car	−0.080	0.053		
Subway lines at home for two or more cars	−0.146*	0.085		
Subway lines at work for one car	0.046**	0.019		
Subway lines at work for two or more cars	−0.017	0.026		
Bus lines at home for one car	−0.018*	0.011		
Bus lines at home for two of more cars	0.010	0.013		
Miles to Midtown Manhattan for one car	0.073***	0.023		
Miles to Midtown Manhattan for two or more cars	0.108***	0.029		
Retail density for one car	0.419	0.284		
Retail density for two or more cars	−0.860*	0.516		
Employment density for one car	−0.004*	0.002		
Employment density for two or more cars	−0.000	0.004		
Population density (L) for one car	0.052**	0.022		
Population density (L) for two or more cars	0.038	0.024		
Population density (H) for one car	−0.007***	0.002		
Population density (H) for two or more cars	−0.021***	0.003		
Auto commuter if one car in HH	2.619***	0.254		
Auto commuter if two or more cars in HH	3.220***	0.347		
Transit commuter if one car in HH	−0.086	0.185		
Transit commuter if two or more cars in HH	−0.225	0.306		
Staten Island HH if one car in HH	−0.725	0.456		
Staten Island HH if two or more cars in HH	−0.705	0.478		
Manhattan HH if one car in HH	−0.610***	0.209		
Manhattan HH if two or more cars in HH	0.106	0.318		
Plus alternative-specific constants[a]				
Estimation summary information				
Observations	2621			
Alternatives[b]	3			
Pseudo R^2	0.277			

[a] *There are 2 alternative specific constants in this model, representing car ownership level alternatives..*
[b] There are 3 car ownership alternatives in this model.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

Table A.3
Elasticities of car use for commuting in Mode Only Choice Model.

	Car use for commuting
Five boroughs of New York city	
Car commute cost (incl. parking)	−0.29
Non-car commute cost	0.08
Car commute time	−0.07
Non-car commute time	0.64
Walking time	0.11
Waiting time	0.47
Riding time	0.07
Manhattan only	
Car commute cost (incl. parking)	−0.68
Non-car commute cost	0.10
Car commute time	−0.14
Non-car commute time	0.92
Walking time	0.17
Waiting time	0.63
Riding time	0.11
Staten Island only	
Car commute cost (incl. parking)	−0.24
Non-car commute cost	0.05

(*continued on next page*)

Table A.3 (*continued*)

	Car use for commuting
Car commute time	−0.05
Non-car commute time	0.45
Walking time	0.11
Waiting time	0.28
Riding time	0.06
Rest of New York City	
Car commute cost (incl. parking)	−0.25
Non-car commute cost	0.08
Car commute time	−0.06
Non-car commute time	0.64
Walking time	0.10
Waiting time	0.47
Riding time	0.06

Table A.4
Elasticities of car ownership in Car Ownership only Choice Model.

	Zero-car ownership	One-car ownership	Two+ car ownership
Five boroughs of New York City			
Population density (home)	0.25	0.04	−0.41
Subway lines (home)	0.02	−0.01	−0.01
Income	−0.50	0.10	0.58
Manhattan only			
Population density (home)	0.26	−0.20	−1.30
Subway lines (home)	0.03	−0.04	−0.07
Income	−0.42	0.63	0.94
Staten Island only			
Population density (home)	−0.17	0.14	−0.04
Subway lines (home)	0.01	0.00	0.00
Income	−1.06	−0.33	0.30
Rest of New York city			
Population density (home)	0.25	0.06	−0.40
Subway lines (home)	0.02	0.00	−0.01
Income	−0.54	0.02	0.59

References

Asensio, J., 2002. Transport mode choice by commuters to Barcelona's CBD. Urban Studies 39, 1881–1895.
Ben-Akiva, M., Bowman, J.L., 1998. Integration of an activity-based model system and a residential location model. Urban Studies 3 (7), 1131–1153.
Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA.
Bhat, C.R., Pulugurta, V., 1998. A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. Transportation Research Part B 32, 61–75.
Bosang, L., 1999. Calling patterns and usage of residential toll service under self selecting tariffs. Journal of Regulatory Economics 16, 45–81.
Community Cartography, 2004. Real Property Assessment Database. Prepared using data from the NYC Department of Finance.
Dargay, J., Gately, D., 1999. Income's effect on car and vehicle ownership, worldwide: 1960–2015. Transportation Research Part A 33, 101–138.
de Jong, G., 1990. An indirect utility model of car ownership and private car use. European Economic Review 34, 971–985.
Gangrade, S., Pendyala, R.M., McCullough, R.G., 2002. A nested logit model of commuters' activity schedules. Journal of Transportation and Statistics 5 (2/3).
Goldberg, P.K., 1995. Product differentiation and oligopoly in international markets: the case of the U.S. automobile industry. Econometrica 63, 891–951.
Hensher, D.A., Ton, T.T., 2000. A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. Transportation Research Part E 36, 155–172.
Ingram, G.K., Liu, Z., 1997. Motorization and the Provision of Roads in Countries and Cities. Technical Report, World Bank. Policy Research Working Paper No. 1842.
Ingram, G.K., Liu, Z., 1999. Determinants of motorization and road provision. In: Essays in Transportation Economics and Policy. Brookings Institution Press, Washington, DC. Chapter 10.
Lerman, S.R., 1976. Location, housing, automobile ownership, and mode to work: a joint choice model. Transportation Research Record: Journal of the Transportation Research Board 610, 6–10.
Mannering, F., Winston, C., 1985. A dynamic empirical analysis of household vehicle ownership and utilization. Rand Journal of Economics 16, 215–236.
Manski, C.F., Lerman, S.R., 1977. The estimation of choice probabilities from choice based samples. Econometrica 45 (8), 1289–1316.
Manski, C.F., Sherman, L., 1980. An empirical analysis of household choice among motor vehicles. Transportation Research Part A 14, 349–366.
McFadden, D., 1978. Modeling the choice of residential location. In: Spatial Interaction Theory and Planning Models. North-Holland, Amsterdam, pp. 75–96.
New York City Department of City Planning, 2004. Street map GIS files. Updated version available as of 8/08 at <www.nyc.gov/html/dcp/html/bytes/applbyte.shtmlk>.
New York City Transit, 2005. Subway and bus line maps and schedules. Updated version available as of 8/08 at <www.mta.nyc.ny.us/nyct/maps>.
NYMTC, 2000. Regional Travel - Household Interview Survey. New York Metropolitan Transportation Council (NYMTC) and North Jersey Transportation Planning Authority (NJTPA). Prepared by NuStats International in association with Parsons Brinckerhoff Quade & Douglas, Inc.

Oum, T.H., Waters, W., Yong, J.-S., 1992. Concepts of price elasticities of transport demand and recent empirical estimates: an interpretive survey. Journal of Transport Economics and Policy 26, 139–154.
Schimek, P., 1996. Household motor vehicle ownership and use: how much does residential density matter? Transportation Research Record 1552, 120–125.
Thobani, M., 1984. A nested logit model of travel mode to work and auto ownership. Journal of Urban Economics 15 (3), 287–301.
Train, K., 1980. A structured logit model of auto ownership and mode choice. The Review of Economic Studies 47 (2), 357–370.
Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press.
Train, K.E., McFadden, D.L., Ben-Akiva, M., 1987. The demand for local telephone service: a fully discrete model of residential calling patterns and service choices. The RAND Journal of Economics 18 (1), 109–123.
US Census Bureau, 1990. 1990 Decennial United States Census of Population and Housing. Washington, DC.
US Census Bureau, 1997. 1997 US Business Patterns Census. Washington, DC.
US Census Bureau, 2000. 2000 Decennial United States Census of Population and Housing. Washington, DC.
Waddell, P., 1993. Exogenous workplace choice in residential location models – is the assumption valid? Geographical Analysis 25 (1), 65–82.
Zhang, M., 2004. The role of land use in travel mode choice: evidence from Boston and Hong Kong. Journal of the American Planning Association 70, 344–360.