The influences of past and present residential locations on vehicle ownership decisions

Gregory S. Macfarlane, gregmacfarlane@gatech.edu
Laurie A. Garrow*, laurie.garrow@ce.gatech.edu
Patricia L. Mokhtarian, patmokh@gatech.edu

Georgia Institute of Technology, School of Civil and Environmental Engineering, 790 Atlantic Drive, Atlanta, GA 30332-0355, United States

*Corresponding author. Tel.: +1 (404) 385 6634.

ABSTRACT

This study explores the relationship between historical exposure to the built environment and current vehicle ownership patterns. The influence of past exposure to the built environment on current vehicle ownership decisions may be causal, but there are alternative explanations. Households may primarily select to live in neighborhoods that facilitate their vehicle ownership preferences, or they may retain preferences that they have developed in the past, irrespective of their current situations. This study seeks to control for these alternative explanations by including the built environment attributes of households' past residences as an influence on vehicle ownership choices. We use a dataset from a credit reporting firm that contains up to nine previous residential ZIP codes for households currently living in the 13-county Atlanta, Georgia, metropolitan area. Results show that past location is significant, but of marginal influence relative to the attributes of the current location. From a practical perspective, our results suggest that models that include current but not past neighborhood attributes (also controlling for standard socioeconomic variables) can forecast vehicle ownership decisions reasonably well. However, models that include both current and past neighborhood attributes can provide a more nuanced understanding of the built environment's potentially causal influences on vehicle ownership decisions. This better understanding may provide more realistic forecasts of responses to densification or other travel demand management strategies.

Highlights

- Prior addresses can capture past exposure to density and non-auto modes.
- Past exposure to density slightly decreases the probability of owning more than two autos.
- Past exposure to non-auto modes decreases the probability of owning more than one auto.
- The influence of current neighborhood attributes is stronger than past exposure.

1 Introduction

Society's dependence on private vehicles creates several negative externalities. From an economic perspective, traffic congestion cost the U.S. economy $121 billion in lost wage productivity in 2011 (Schrank et al., 2012). Economic externalities from vehicle dependence may be even more pronounced among certain demographic groups; for instance, the reduced ability of low-income households to obtain vehicles is often viewed as a factor contributing to low economic mobility (Leonhardt, 2013; Matas et al., 2009). In addition, a large portion of urban air pollution is due to transportation-related emissions and can contribute to global climate change (Chapman, 2007), respiratory ailments in the general public (Buckeridge et al., 2002), and other negative externalities.

Previous studies have explored relationships between vehicle ownership decisions and the built environment (e.g., Dieleman et al., 2002; Ewing and Cervero, 2001; Giuliano and Dargay, 2006; Van Acker and Witlox, 2010). These studies have hypothesized that households who live in denser or more mixed-use areas can access a larger number of activity locations by walking, biking, or taking public transit, reducing the need for one (or more) vehicles. Policies designed to increase densities or land use mix are therefore often viewed as mechanisms for reducing vehicle ownership and/or vehicle usage, which in turn would help reduce emissions (Kenworthy and Laube, 1996; Norman et al., 2006; Stone, 2009), potentially alleviate congestion, and improve transportation equity (Sanchez et al., 2003).

To isolate the autonomous influence of the built environment on vehicle ownership decisions, it is important to control for other possible causal influences. On one hand, self-selection could explain part of the observed correlation between the built environment and travel behavior; that is, individuals who prefer to own fewer vehicles may choose to live in denser or more mixed neighborhoods *so that* they can own fewer vehicles. Density in this situation facilitates, rather than causes, a particular behavior. If this is true, then incentivizing or requiring density through zoning or tax policies may not change vehicle ownership in a meaningful way, *unless preferences also change*. On the other hand, however, individuals' preferences for vehicle ownership may, in fact, evolve over time as they are exposed to more dense and mixed neighborhoods and learn about non-vehicle transportation options. If this is true, then building denser or more mixed-use developments could eventually lead to lower levels of vehicle ownership, although the short-term effectiveness of using density as a planning tool for positive environmental and economic changes could be diminished.

In this study, we use multinomial logit models to predict the (non-zero) number of vehicles owned by a household as a function of the head of the household's socioeconomic characteristics, current neighborhood density (defined as housing units per square kilometer in the ZIP code) and the current neighborhood's use of non-vehicle transportation modes (as indicated by the ZIP code's non-vehicle commuting mode share). To this base model we add variables that characterize the head of household's historical *exposure* to both density and non-vehicle transportation alternatives (using the same definitions for each metric). We use the results to assess the practical implications of omitting prior built environment information in vehicle ownership choice models.

Our study is distinct from the majority of prior studies reported in the literature in that we use data from existing third-party sources (namely prior addresses and socio-demographic information maintained by credit reporting firms) to explore these research questions. Our paper contributes to the literature by demonstrating how existing data sources can be mined to explore nuanced questions, such as which functional form representing the influence of prior residences best fits models of vehicle ownership. However, because we use existing revealed preference data, we are unable to explicitly model the role of attitudes and preferences on vehicle ownership decisions, as prior studies based on household surveys have been able to do.

The paper is organized into several sections. Section 2 describes how our study relates to and contributes to the literature, presenting a conceptual model of vehicle ownership response to a change in residential location. Section 3 provides an overview of the analysis database and data processing assumptions. Sections 4 and 5 follow, presenting the econometric methodology and results, respectively. Section 6 discusses the practical implications of the results, and Section 7 presents study limitations and directions for future research. The paper concludes with a summary of key findings and implications for practice. An appendix details the results of sensitivity tests we conducted to verify the robustness of results to different data processing and modeling assumptions.

2 Literature Review

Many forecasting models used in professional practice view vehicle ownership as a strictly utilitarian phenomenon. In this perspective, a given household has a need for vehicles established by the size of the household (or its number of workers) and the availability of vehicle alternatives (which, in turn, is a function of the built environment). The household acquires the necessary vehicles subject to income constraints (as an example, see Potoglou and Susilo, 2008). In contrast, a growing body of literature shows that attitudes and preferences play a significant role in vehicle choice and use. Some people have a propensity to own vehicles that are faster or more stylish than strictly necessary (Lois and López-Sáez, 2009). Others choose vehicles that signal environmental-political preferences (Sexton and Sexton, 2014). More generally, it has been shown that people tend to own vehicles that are similar to those driven by their neighbors (Adjemian et al., 2010).

Separating the effects of these attitudes from the effects of the built environment is difficult, as people select their built environment based at least partially on these attitudes. Studying this *self-selection* problem has become an important research objective in transportation behavior modeling. Cao et al. (2009) and Mokhtarian and Cao (2008) extensively review the literature on self-selection, and Cao and Cao (2013) give a recent presentation of self-selection in a vehicle ownership context. Handy et al. (2009) and Pinjari et al. (2009) likewise address environmental self-selection, though in transportation contexts other than vehicle ownership. Self-selection may be addressed in a vehicle ownership model by considering the vehicle choice and residential location choice as occurring simultaneously (e.g., Eluru et al., 2010; Roorda et al., 2009), through structural equations modeling (e.g., Bagley and Mokhtarian, 2002; Cao et al., 2007b), or by including attitudes as exogenous variables when they can be observed.

For example, Cao et al. (2007a) surveyed 547 households that had recently (in the previous year) moved into a group of Northern California neighborhoods representing either traditional or suburban land use characteristics. These characteristics included indicators such as the age and style of homes, street connectivity, and distance to various commercial establishments. The survey asked the households about their current and previous vehicle ownership levels as well as their attitudes toward travel behaviors (e.g., "I need a car to do many of the things I like to do") and neighborhood design (e.g., "I prefer shopping areas within walking distance"). The authors showed that these attitudes were more predictive of household vehicle ownership than were objective measurements of the neighborhoods. Additionally, the households in the survey tended to relocate to either traditional or suburban areas in a pattern consistent with both their previous land use and their expressed attitudes. Nevertheless, the authors allow that preferences might change with experience, stating "it is possible that the built environment also plays an additional indirect role by influencing these attitudes over time" (Cao et al., 2007a p.846).

Figure 1 presents a conceptual model of how the built environment could play this additional indirect, "learned preferences" role. The figure shows vehicle ownership (VO) decisions at two periods in time. A residential relocation occurs between time z=1 and time z=2. At any given point in time, vehicle ownership (VO) is an outcome of the socioeconomic state of the household (SE), the built environment (BE) where the household is located, and the attitudes and preferences of the household members (AT). Learned preferences are the impacts that behavior (choices of BE and VO) has on AT; that is, they are a conceptual mechanism for the dynamic (time-delayed) feedback from behavior to attitudes. On one hand, these observed transportation behaviors (choices of BE and VO) may be a proxy for access to alternative modes. On the other hand, they could be considered a measure of the social environment. Self-selection is more complex, but could be characterized as the way in which the (contemporaneous) impact that a certain variable (AT or even SE) has on one behavior (BE) subsequently influences another behavior (VO). The learned preference mechanism is confounded with self-selection in empirical observation when, as is the case here, we: (1) envision a dynamic process, and (2) do have past behavior (BE) measures, but (3) do not have contemporaneous AT measures.

[Insert Figure 1 here]

Weinberger and Goetzke (2010) emphasized a learned preference perspective, examining the relationship between implied exposure to vehicle independence at a previous residential location and current vehicle ownership using national data from the 2000 Census "long" form survey (U.S. Census Bureau, 2000a). The 2000 Census asked a sample of respondents "Did this person live in this house or apartment 5 years ago (on April 1, 1995)?" If the respondent answered "no," a follow-up question obtained the address of the location where the individual lived on April 1, 1995. Weinberger and Goetzke (2010) showed that people who reported prior addresses in San Francisco, Chicago, Philadelphia, Boston, New York, and Washington, D.C. owned fewer vehicles than other households, all else equal. The authors attributed this effect to these individuals having learned a preference for, or positive attitude toward, transit use when living in those transit-rich locations, a preference that persisted after a change to a new built environment.

Note that in both Cao et al. (2007a) and Weinberger and Goetzke (2010), exposure measures that describe characteristics of the built environment at past residential locations are used to isolate the impact of the current built environment on vehicle ownership decisions. A household's previous exposure to density, for example, may either indicate what its historical preferences have been (under a self-selection theory) or may signify experiences that have shaped those preferences (under a learned preferences theory); it may not always be possible to empirically discriminate between the theories. Intuitively, we expect that individuals who have frequently changed residences would be more likely to be exposed to new environments and thus more likely to form new preferences for vehicle ownership. However, we are unable to confirm this intuition with the available data.

In this paper, we build on two previous papers (Cao et al., 2007b; Weinberger and Goetzke, 2010) that have used previous addresses to investigate the influence of past and present residential locations on current vehicle ownership decisions.[1] In contrast to these previous studies, our dataset contains multiple prior residential locations as well as the length of time spent at each residential location. This enables us to derive a wide set of exposure metrics and test the robustness of results in each metric.

3 Data

Data that would provide researchers with a better ability to infer the autonomous influence of the built environment on vehicle ownership decisions would ideally contain at least the following two elements: (1) a history of several previous addresses, including the dates of relocation; and, (2) a nationwide scope of previous addresses given at a small geographic resolution. Our analysis database meets these two conditions. The database is compiled from four primary sources: vehicle information is obtained from the state's registration database, socioeconomic information is obtained from a targeted marketing (TM) firm, residential move histories are compiled from a credit reporting firm, and residential densities and commuting mode shares are calculated from Census data. A schematic figure of how these databases relate to each other is given in Figure 2. Our final analysis dataset contains 227,830 households, constituting a 12% sample of the 13-county metropolitan Atlanta region. Definitions for each of the variables used in the study, as well as descriptive statistics for these variables, are provided in Tables 1 and 2. The majority of the variables shown in the tables have a straightforward interpretation. This section provides definitions for variables that merit additional discussion and provides an overview of key assumptions used to

[1]Analyzing movers in the Puget Sound Transportation Panel dataset, Krizek (2003) considers the change in various travel behaviors (such as vehicle miles traveled) resulting from changes in neighborhood accessibility, but he does not specifically examine vehicle ownership decisions.

compile the analysis database that may influence the representativeness of our analysis database and/or the interpretation of results.

[Insert Figure 2 here]

[Insert Table 1 here]

[Insert Table 2 here]

3.1 Motor Vehicle Database

The number of vehicles owned by a household was obtained from the Georgia Department of Revenue's Motor Vehicle Division (called the DMV), which maintains records for all vehicles registered with the state. For the purposes of this study, we consider as vehicles only passenger cars and light-duty trucks (not, for instance, cargo vans or motorcycles). We drew a simple random sample of vehicles registered to addresses in the 13-county metropolitan Atlanta area from this database, removed duplicate registration addresses, and simply sampled replacements. By this method, addresses with multiple vehicles have a proportionally higher probability of being sampled; Section 3.6 discusses the representativeness of our sample and potential biases resulting from sampling and processing assumptions.[2] We then appended to each sampled address the number of vehicles registered at that address as of December, 2010. Home addresses were a prior to merging the databases; however, approximately 1.5% of the addresses had to be excluded as it appeared that apartment numbers were omitted. During the merge process, this resulted in a large number of vehicles being associated with multi-unit buildings. Although it was not possible to determine exactly which addresses represented unique households (versus a multi-unit building), we assumed those addresses with five or fewer vehicles represented unique households. Multi-unit buildings where the units had distinct street addresses remain in the analysis database.

3.2 Targeted Marketing Data

Target marketing (TM) firms compile information about individuals from a variety of sources (e.g., public records, credit records, credit card transactions). These data are often sold to advertising companies to customize marketing campaigns to potential customers. These TM databases contain the majority of household and individual demographic fields that are used in travel demand forecasting applications. TM data has been used in several prior travel demand studies (e.g., see Binder et al., 2014; Kressner and Garrow, 2012).

The household's residential address in the TM database was joined to the registration address in the DMV database; vehicles registered to non-residential addresses were left

[2]Specifically, since our final sample is representative with respect to median and mean income, the bias inherent to this approach is apparently either negligible or essentially counteracted by the bases on which some of the sampled addresses were discarded.

unmatched and discarded from the analysis.[3] The TM database provides socioeconomic information about the number of adults in the household, the number of children in the household, the household annual income and ethnicity,[4] the age of the head of household, and housing tenure. As an explanatory variable in the vehicle ownership model, we include the number of adults in the household in relation to the number of vehicles that could be potentially owned by the household. Specifically, for a potentially-chosen number of vehicles j in household i, we define $Insufficiency_{ij}$ as

$$\max(0, Adults_i - j).$$

For example, if there are two adults in the household, the insufficiency variable would take a value of one for the "one vehicle" alternative ($j = 1$), zero for the "two vehicles" alternative, and zero for the "three or more vehicles" alternative. In this sense, the insufficiency variable represents a measure of competition for limited vehicle resources in the household.

For the TM database used in this study the variable indicating the number of children is ambiguous in that a "0" value may indicate either a known zero count or an unknown value.[5] An analysis against Census data (U.S. Census Bureau, 2012a) revealed that 38% of households in the Atlanta region have children, whereas in our data only 28% do. We examined the robustness of model results to different imputation methods for this variable as part of the analysis.

3.3 Move Histories

The TM firm who provided data for this analysis has a close relationship to a credit reporting firm. The credit reporting firm maintains a database of current and previous addresses at the ZIP code level for (in theory) up to four adults living in the household. A maximum of nine previous ZIP codes for each adult is available. The date (month and year) that each address changed is also available. For households with more than one adult listed in the TM records, only 40% contained an address history for a second adult. Using this second history would potentially allow us to study mismatched households: for instance, a household where one partner has lived in dense urban areas and the other has not. However, using the address history for the second adult in the household would require us to make strong assumptions about household formation and dynamics, so we restricted our analysis to the move histories for the heads of households.

We made several assumptions to construct coherent address paths from the source data. The credit reporting firm noted that the ZIP codes and associated move-in dates may not be in sequential order for all households. Of the records in our dataset, 34% required reordering. Also, by comparing the current ZIP codes in the TM and credit reporting databases, we

[3]Vehicles may be registered to non-residential addresses; however, given that TM data are compiled from personal credit reports and loan applications, it is highly unlikely that a work address would be reported in the TM database.

[4]Ethnicity is given at the household level and may be either observed or modeled.

[5]The TM marketing firm recently updated the algorithm it uses to populate the "number of children" field, so this is not expected to be a limitation in subsequent studies.

noted inconsistencies in how the credit company recorded the “most recent” address. Although the majority of “most recent” ZIP codes matched, 27% did not. We assumed this implied that the credit reporting database was missing the most recent move, i.e., that the ZIP code obtained from the TM database was accurate. We calculated the move-in date for these households using two assumptions: the first assumption represents the “earliest” move-in date and the second assumption represents the “latest” move-in date. The earliest move-in date assumes the move occurred one month after the head of household’s last known address change. The latest move-in date assumes the move occurred one month before the data were collected, or on December 1, 2010. We tested the sensitivity of exposure metric calculations to these two assumptions. Finally, some households in the database (1.4%) had consecutive moves in the same month; we forced the length of residence in this case to be one day.

The prior ZIP codes provide comprehensive coverage of the U.S. On average, the heads of households in the database have lived at 2.5 previous addresses. There are 14,785 unique prior ZIP codes in the database, which represents 44% of all ZIP codes in the U.S.; ZIP codes from all 50 states, the District of Columbia, and Puerto Rico are represented. The large majority of ZIP codes (89%) are from the southeastern U.S.,[6] specifically Georgia (73%) and the Atlanta metropolitan statistical area (MSA) itself (71%). Excluding prior addresses in the Atlanta MSA, 32% are from other MSAs that are ranked in the top ten by population and 62% are from MSAs that are ranked in the top 50. The six cities identified by Weinberger and Goetzke (2010) as having high transit accessibility account for 21% of the non-Atlanta prior ZIP codes, suggesting that a significant subset of our sample has some experience with very large metropolitan areas and – potentially – rail mass transit systems.

3.4 Census Data

As the built environment and observed transportation behavior may independently affect preference development, it is important to have a measurement of each. Our measurement of land use is density, which we define as the count of housing units in each ZIP code divided by the land area (in square kilometers) of the ZIP code. We select housing units – rather than population or households – as a more direct measure of the built environment. Though density is an imperfect measure of the built environment, it is both readily calculated from available data and (doubtless for that reason) commonly used (Transportation Research Board, 2009). As a measure of the availability or feasibility of non-vehicle transportation modes, we define the non-vehicle mode share as the proportion of workers who *do not* drive or carpool to work in each ZIP code. These people may use public transit, walk, bike, etc. We calculate these metrics from Census data. For the current ZIP code and for residences occupied since 2005, we use tables from the American Community Survey (ACS) (U.S.

[6]Defined by Census as being the area bounded by and including Texas, Oklahoma, Arkansas, Kentucky, West Virginia, and Maryland.

Census Bureau, 2011, 2012b).[7] For residences occupied earlier than 2005, we draw instead from the 2000 Census and its "long" form sample (U.S. Census Bureau, 2000b, 2000c).

3.5 Past Exposure

This study tests the hypothesis that a head of household's previous experience with density or exposure to non-vehicle modes can be used to predict the household's current vehicle ownership. We do this by creating a household *past exposure metric*, E_i, for household i. This metric applies some rule to characterize the density or non-vehicle mode share at all of the previous addresses in which the head of household is known to have resided. As the existing literature provides little insight into *how* prior experience shapes preferences for vehicle ownership, we develop an array of plausible exposure metrics. Because households may be observed for different lengths of time and/or have a different number of prior addresses, we use metrics that do not require comparisons across households.

To describe our past exposure metrics, we first need to formalize notation. Given household i that lived at previous ZIP code number z ($z = 1,2,3,\dots,Z_i$), the density or non-vehicle mode share associated with the previous ZIP code, d_{iz}, is defined for up to nine previous addresses. However, because not all households have that many addresses in the database, we define Z_i as the maximum number of previous addresses available for household i. The corresponding length of residence at ZIP code number z is denoted as t_{iz}. Households with no previous addresses, and therefore no observable past exposure, are excluded from the study.

3.5.1.1 Duration

The duration past exposure score is the arithmetic mean of the attribute, weighted by the time spent in the corresponding ZIP code. Longer residences carry more weight, but each additional day has an unknown marginal importance α:

$$E_i = \frac{\sum_{z=1}^{Z_i} d_{iz} \times t_{iz}^{\alpha}}{\sum_{z=1}^{Z_i} t_{iz}^{\alpha}}.$$

In this formulation, $\alpha = 1$ represents a constant marginal effect with each day carrying equal weight. For $\alpha > 1$ there is an increasing marginal effect for each day, thus accentuating the relative importance of long residences; this may represent individuals developing an "addiction" to the attribute. For $0 < \alpha < 1$ there is a diminishing marginal effect, increasing the relative importance of shorter residences. Finally, the case where $\alpha = 0$ reduces the exposure score to the simple arithmetic mean, which indicates that the exposure is duration-insensitive and that each prior residence has equal weight.

[7]The 5-year aggregation for 2007-2011 is used as it is the earliest ACS product to provide tables by ZIP Code Tabulation Area (ZCTA).

3.5.1.2 *Decay*

The decay past exposure score assumes that older addresses affect current preferences or behavior differently than recent ones, determined by the estimable parameter γ:

$$E_i = \frac{\sum_{z=1}^{Z_i} d_{iz} \times t_{iz} \times z^{\gamma}}{\sum_{z=1}^{Z_i} t_{iz} \times z^{\gamma}}.$$

In this formulation $\gamma = -1$ represents a geometric decay function, with each day spent at the $z = 3$ (third prior) address contributing $1/3$ as much to the exposure as each day at the $z = 1$ (first prior) address. For $-1 < \gamma < 0$ memory decays less quickly than geometric, and $\gamma > 0$ implies that older addresses are *more* important to current behavior. For completeness, the case where $\gamma = 0$ implies a simple time-weighted mean of the attribute, corresponding to $\alpha = 1$ for the duration metric.

3.5.1.3 *Extreme*

The extreme past exposure score assumes that individuals' behavior is most influenced by the highest density or non-vehicle share they have ever experienced, or

$$E_i = \max(d_{iz}), z = 1, \dots, Z_i.$$

3.5.1.4 *Longest*

The longest past exposure score assumes that individuals' behaviors are simply a function of the place z^* at which they have resided the *longest*, or that

$$E_i = d_{iz^*},$$

where z^* is the prior residence for $t_{iz^*} = \max_z(t_{iz}), z = 1, \dots, Z_i$.

3.6 Representativeness

Our assembled database contained records for 417,538 households (representing 22% of all Atlanta households). Although we were forced to delete records with one or more missing fields, the resulting dataset of 227,830 complete records was still representative of the Atlanta region as a whole in important respects. To illustrate this point, in our estimation sample the median income is $62,500 per year and the mean is $80,196; for vehicle-owning respondents to the ACS in the 13-county Atlanta region,[8] the comparable figures are $62,000 and $81,693, respectively. The distributions of the number of household adults and the ages of adults in the household are also similar between the estimation sample and Census data. However, our sample is biased toward homeowners (95% versus 78%) and Whites (75% versus 67%). This bias was also seen in a study by Kressner and Garrow (2015), which compared a "complete" sample of TM data at the block group level with Census data for the 13-county Atlanta region; that is, the over-representation of homeowners and Whites is seen in the "full TM database" as well as in the "reduced analysis dataset" ultimately used for this study.

[8]Specifically, the collection of Public Use Microsample Areas (PUMAs) that contain the 13 counties as defined by the Environmental Protection Agency for air quality mitigation purposes.

To assess whether the length of time at the current residence calculated from the credit reporting database was representative, we compared the distribution of this variable to housing tenure fields available in the TM and Census data. The TM data contain a variable that is the length of time the household has lived at its current address. This variable is correlated with (correlation 0.62), but not identical to, the length of time calculated from credit records. Figure 3 shows a comparison of the length of residence as recorded in the TM database, the credit records, and as measured in the 2009 American Housing Survey (AHS). The AHS data are for the entire Southeast region (the smallest relevant geography available to the public), but Atlanta residents are likely more mobile than the Southeast region in general. The credit records show, on average, shorter residences than either the AHS or TM records. A plausible explanation is that some people move away only temporarily (e.g., college students living on campus during the academic year and at their parents' home during break); in this case, their credit records may show multiple relocations that may not be self-reported.

[Insert Figure 3 here]

Although the sample is not representative of the population in every way, this is less of a concern when the purpose of the sample is to uncover relationships among variables (as it is here) than when it is purely to describe a population (Babbie, 2009; Groves, 1989, Chapter 1). For example, if we were using the sample to estimate the true share of various races in the population it would be problematic, but a model based on the sample can properly predict vehicle ownership *given race*. In particular, when the model is multinomial logit, or MNL (as it is here), Manski and Lerman (1977) showed that under certain conditions, the MNL parameter estimates obtained from a stratified sample will be consistent and unbiased relative to the MNL estimates obtained from a simple random sample. Thus, we do not expect that the estimated effect of exposure measures on the number of vehicles owned by the household will be impacted by the non-representativeness of our estimation database.

4 Empirical Model

The number of vehicles y_i that household i owns is assumed to be a function of the socioeconomics of the household SE_i, the built environment around the current residence BE_i, and the attitudes or preferences of the household AT_i:

$$y_i = f(SE_i, BE_i, AT_i) \ .$$

In this study, AT_i is unobserved, but we use the exposure to a built environment BE_i as a proxy for it (following the conceptual model in Figure 1).

For cases where y represents discrete ordinal outcomes, as in the current situation, it is natural to model the *probability* that y takes on a given value, using an ordinal response model or a discrete choice model such as the MNL (McFadden, 1974). Several prior studies that have compared ordinal response and MNL models for vehicle ownership have found the MNL models to be superior (Bhat and Pulugurta, 1998; Potoglou and Susilo, 2008); we thus follow this convention and use MNL models. In this case, the f function represents the utility of owning y_i vehicles, and the parameters of that function (as well as, potentially, some of the explanatory variables) can differ with y. More formally, in the MNL, the utility U for household i in choosing alternative j from choice set J is a linear function of $\boldsymbol{x}_{ij}$, $U_{ij} = \boldsymbol{\beta}'_j\boldsymbol{x}_{ij} + \epsilon_{ij}$, where $\boldsymbol{x}_{ij}$ comprises the SE, BE, and AT variables described above. If ϵ_{ij} is distributed independently and identically with a Gumbel (or extreme value type I) distribution, the probability of individual i choosing alternative j is given as:

$$P(y = j \,|\boldsymbol{x}_{ij}) = \frac{e^{\boldsymbol{\beta}'_j\boldsymbol{x}_{ij}}}{\sum_{k\in J} e^{\boldsymbol{\beta}'_k\boldsymbol{x}_{ik}}}.$$

We estimate the MNL model using the "mlogit" package for R (Croissant, 2013). As a measure of model fit, we use the McFadden likelihood ratio index with respect to constants,

$$\rho_C^2 = 1 - \frac{\log(\mathcal{L}_\beta)}{\log(\mathcal{L}_C)}$$

where $\mathcal{L}_\beta$ is the final model likelihood, and $\mathcal{L}_C$ is the likelihood of a constants-only (market share) model.

5 Results

Our presentation of results proceeds as follows. First, we present results for a base model that includes only current address attributes, representing the most common situation in which only those attributes are available to the analyst. Next, we compare these results to a set of models that each add one of the past exposure metrics developed in Section 3.5 to the base model. These specifications allow us to assess whether the interpretation of results is robust to different exposure metrics. We use the results from the best fitting model that includes current neighborhood attributes and past exposure metrics to isolate the autonomous effect of the built environment on vehicle ownership. An appendix to the paper presents a sensitivity analysis of the results.

5.1 Base Model

Table 3 presents a base model that captures the effects of demographic characteristics, current density, and non-vehicle mode share on vehicle ownership. Having insufficient vehicles in a household brings a fairly substantial disutility, as shown by the strongly negative coefficient value (the greater the number of adults who would be without vehicles — which will be larger for smaller numbers of vehicles — the greater the disutility). Each additional child increases the probability of owning multiple vehicles. An increase[9] in household income is correlated with an increase in the probability of owning more than one vehicle. All else equal, Asians and Hispanics are more likely to own multiple vehicles than Whites, African Americans, and Other ethnic groups. Households that rent are more likely to

[9]The natural logarithmic transformation models a constant effect of a given *percentage* change in income, representing a diminishing marginal effect of each additional dollar of income on utility. The same logic applies to density and alternative share.

own fewer vehicles. There is a parabolic effect of householder age: the utility of owning multiple vehicles increases with age to a point (32 years old for two vehicles, 54 for three or more), and then declines again as the head of household ages. Finally, the density and the non-vehicle mode share associated with the household's current ZIP code show their expected negative coefficients. As density or the number of people commuting without a vehicle increases, the probability of owning two or three or more vehicles decreases. This finding supports many previous studies (Bhat and Guo, 2007; Cao et al., 2007a; Giuliano and Dargay, 2006) that have shown individuals living in dense areas with non-vehicle transportation options tend to own fewer vehicles, all else being equal.

[Insert Table 3 here]

5.2 Models that Include Past Exposure Metrics

The models presented in Table 4 control for learned preference and/or self-selection effects by introducing past exposure metrics. The models include the same variables shown in Table 3; however, no coefficients changed by an order of magnitude or by a level of significance[10] and are therefore not shown with the exception of the log of the current density and the log of current non-vehicle mode share (which do change). Of particular note, in no case did the signs of these coefficients change, showing that the directional interpretation of the influence of these variables on vehicle ownership is consistent across all of the exposure formulations.

[Insert Table 4 here]

Overall, the four models (representing different exposure score metrics) have ρ_C^2 fit statistics that are similar, indicating that the experiences the exposure metrics seek to capture are robust to different specifications. However, the model based on the "Extreme" exposure formulation, which simply considers the prior residence with the highest density and the prior residence with the highest non-vehicle mode share that the head of the household has experienced, fits the data the best (by a slight but statistically significant[11] margin). Multiple duration exposure formulations were estimated by varying the value α from 0 to 1.3 in 0.1 unit increments. Interestingly, among these duration exposure formulations, the one that fit the data the best is the case where $\alpha = 0.1$, or a very lightly weighted average of all of the previous ZIP codes. Similarly, the multiple decay formulations were estimated by varying γ from -1 to 1 in 0.1 unit increments; the one that fit the data the best is the case where $\gamma = 0$, or an unweighted average.

Across all four exposure measure models, the current housing unit density and the current non-vehicle mode share retain their expected signs and are highly significant. For example, all of the models show that between two households with identical past exposure and all else equal, the household currently living in a neighborhood with higher densities and/or more non-vehicle transportation alternatives is significantly less likely than the other household to own multiple vehicles.

[10]We define four significance levels: $p < 0.10, p < 0.05, p < 0.01$, and $p < 0.001$.

[11]This is determined by a non-nested likelihood ratio test of the type proposed by Vuong (1989).

The influence of past exposure metrics on vehicle ownership is less clear. Prior exposure to higher densities decreases the probability of owning three or more vehicles; however, the same variable has no discernible impact on the probabilities of owning one versus two vehicles, all else being equal. With one exception, the coefficients associated with the past non-vehicle mode share are all negative; thus, individuals who have been exposed to higher non-vehicle mode shares are less likely to own multiple vehicles. However, only three out of the eight parameter estimates associated with the past non-vehicle mode share are significant at the 0.05 level, and thus the relative influence of prior exposure to non-vehicle mode shares is modest.

6 Analysis of Current and Past Effects

The previous section focused on establishing the statistical significance of the association between prior exposure to the built environment and vehicle ownership. Both current and prior exposure metrics were found to be significant in describing vehicle ownership to some degree. These findings can have multiple interpretations in the theoretical context of either learned preferences or self-selection. In this section, we examine the consequences of these findings for forecasts and practical applications.

In the previous section, the base model contained the current neighborhood characteristics, representing the typical built environment variables available in studies of this type. We then added the past exposure metrics, and analyzed how the estimated impact of the current neighborhood traits changed (very little). However, it is also of interest to proceed in the opposite direction: take as a base model one containing the past exposure metrics, and then add the current neighborhood traits and analyze how the impact of the former variables changes. This represents the conceptualization that past experience may be a more fundamental influence on behavior than current circumstances, and should thus be accounted for first. Table 5 presents three models: a model with only current attributes, a model with only past exposure metrics, and a model with both. These models generate two important observations. First, the model using the current coefficients only has a slightly (though significantly, in a statistical sense[12]) better fit than the model with the past coefficients only; the model with both sets has the best fit. Second, both sets of coefficients change when they are placed in a model together, but the past coefficients change more. Both observations suggest that current neighborhood attributes are better independent predictors than past attributes, but that past experiences may have a moderating effect.

[Insert Table 5 here]

As can be seen from Figure 1, planners who consider only current built environment attributes in their models thus estimate parameters that conflate the effects of that current built environment with the past experiences of the residents. As discussed in Section 2, past experiences may indicate either the residents' self-selection preferences or learned preferences. In either case, it is important to determine the consequences of this conflation for

[12]Determined by the same non-nested test previously used.

transportation forecasts. A model with only current attributes will suffice *if the relationship between past and present (i.e., the nature of that conflation) remains unchanged*. Table 6 presents the aggregate error that results from considering several scenarios where this relationship breaks down, revealing potential shortcomings to the current-only model. These are based on comparing the reference predictions from the "extreme" model in Table 5 to those obtained by applying the same model to scenarios in which we have deliberately altered the past exposure metrics. Taking the "three or more vehicles" category to mean "three," the households in our analysis database collectively own 449,145 vehicles;[13] this number is used as a reference scenario. In the "Past equals present" scenario, we assume that all households have always lived in a neighborhood precisely like their current one, imitating a condition where the past experiences of the households are ignored or unavailable to the researcher. In the "Mean past" scenario, we assume that all households have past exposure scores equal to the sample mean, and in the "Random past" scenario, the past exposure scores of all households are randomly shuffled. The final four scenarios effectively bound the expected minimum and maximum changes in vehicle ownership predicted by the model by replacing the past exposure scores with the 1^{st}, 5^{th}, 95^{th}, or 99^{th} percentile in the estimation dataset. Assuming all individuals have previously lived in low density areas with small non-vehicle mode shares (corresponding to the 1^{st} and 5^{th} percentiles) results in a 3.8% to 5.8% increase in vehicles. Similarly, assuming all individuals have previously lived in high density areas with large non-vehicle mode shares (corresponding to the 95^{th} and 99^{th} percentiles) results in a 3.2% to 6.0% decrease in the vehicles. These results may be unique to our data or region, but we suspect that other large U.S. cities will show similar patterns.

[Insert Table 6 here]

Each of these scenarios is absurd in its own way, and yet on aggregate has a small to modest impact on forecasted vehicles. One explanation is that built environment attributes — past or present — have an absolutely small effect on vehicle ownership relative to socioeconomic characteristics. As an example, consider a representative household with two white adults of median age and income, who own their home and have no children. This household is expected (via the extreme model) to own 2.174 vehicles if the head of household has average past exposure scores and the current residence is in a neighborhood with a density and non-vehicle mode share combination that corresponds to the 15^{th} percentile in each metric. However, if this same household lives in a neighborhood with a density and non-vehicle mode share combination that respectively correspond to the 85^{th} percentiles, it would be expected to own 2.034 vehicles.[14] This corresponds to a decrease in the number of vehicles owned of about 6.4% (an average elasticity of -0.014[15]). Conversely, if the current neighborhood's density and non-vehicle mode share are held constant at their means and the head of household's historical past exposure scores move from the 15^{th} to the 85^{th}

[13]In reality, the sampled households own 482,331 vehicles, since some households own four or five.

[14] Using the "base" model that includes only current information, these numbers are 2.204 and 2.026, respectively.

[15]This value is -0.018 with the base model; this larger elasticity indicates that the base model is more sensitive to changes in current density.

percentiles, the household would be expected to own 2.144 versus 2.066 vehicles. This corresponds to a decrease in the number of vehicles of only 3.6% (an average elasticity of -0.0074). Though the current attribute elasticity is 1.9 times greater than the previous exposure elasticity, it remains fairly small in absolute terms; for comparison, the average elasticity of vehicle ownership with respect to income is 0.025 and increasing the number of adults from two to three increases the expected vehicles from 2.097 to 2.435 (16%).

On one hand, this result is somewhat discouraging, as it suggests that exposure to higher densities and non-vehicle transportation options (either currently or in the past) has a relatively modest influence on vehicle ownership decisions. On the other hand, this result is encouraging from a practical perspective, as it suggests that models that include only current neighborhood attributes (in addition to standard socioeconomic information) should be able to reasonably predict vehicle ownership decisions, meaning planners do not likely need to compile extensive histories of prior residences to estimate these models. However, an extensive database of prior addresses does permit a richer analysis and the ability to isolate the autonomous effects of the current built environment on vehicle ownership decisions, as we have done in this study. This can be useful when designing strategies for reducing vehicle ownership in particular groups of individuals, or evaluating different policy mechanisms for reducing vehicle ownership. For instance, if a city has knowledge that its forecasted population growth will be mostly migrants from a particular city or region, a model including characteristics of this previous region might allow for more predictive certainty. Alternatively, cities pursuing densification strategies may consider adjusting their vehicle ownership forecasts based on our observation that the direct effect of the current built environment is somewhat more modest than a naïve current-only model would indicate.

7 Limitations and Future Directions

There are several limitations of our study. Due to difficulties in compiling a list of households that do not own vehicles, we needed to exclude zero-vehicle households from our analysis. However, a comparison of vehicle ownership and income calculated from the Census data provides additional insights into which types of households are missing from our database. Based on data collected in the 2011 ACS (U.S. Census Bureau, 2012a), 4.9% of households in the 13-county area do not own a vehicle. Of these zero-vehicle households, 69% are in the lowest income quintile and 85% are in the lowest two quintiles. Thus these households are primarily those who cannot afford to own a vehicle, rather than those who may be expressing a preference for a vehicle-free lifestyle. Studying the preferences of the wealthier 15% of zero-vehicle households (representing 0.72% of all households) would be an important direction for future research.

Another limitation is that this is a retrospective study of households that currently live in the Atlanta metropolitan area, but formerly lived elsewhere. Individuals who once lived in Atlanta, but now reside in areas outside Atlanta are not represented in our analysis database, and neither are those with no known previous residences. The ideal analysis of learned preferences and self-selection on vehicle ownership would be based on a national longitudinal panel in which changes in vehicle ownership and land use can be directly observed over time, together with changes in socioeconomic traits and attitudes. However, we are unaware of any such database that exists in the U.S.

Finally, it is important to note that this study relies entirely on pre-existing revealed preference data, specifically administrative records. On one hand, our study contributes to the literature by exploring how research questions can be investigated using passively collected data and other forms of "big data." On the other hand, our reliance on revealed preference data does not allow us to design a study "from scratch" and include measures of potentially important intervening variables. This may be one reason that the models show a small effect of prior built environment measures on current vehicle ownership – that is, the causal effect gets diluted along the way. As future research, it would be interesting to combine the revealed preference data with attitudinal/stated preference surveys (individuals could be contacted by mail using the addresses provided on the administrative records). The latter could provide a more complete picture of how current (and potentially past) attitudes and preferences influence vehicle ownership decisions.

These shortcomings notwithstanding, the empirical models we have estimated indicate that one's prior residential history influences one's preference (either learned or innate) for owning more or fewer vehicles, independent from the influence of the existing built environment, but this influence has a negligible aggregate impact. The effect of prior exposure to density and non-vehicle transportation modes is only about half of the effect of the current density and non-vehicle commute mode share. This finding does not preclude the possibility that people learn or develop preferences by other methods (such as sociodemographic changes, e.g. in household size, or social influence, or even self-introspection), which in turn may influence transportation behaviors. Studying such preferences directly is likely to be a fruitful research endeavor.

There are several ways in which this study can be extended or modified to investigate particular theories related to the influence of the built environment on an individual's preference development. It may be that some cities are more influential on preferences, or that individuals learn their preferences at different stages in life. For example, an exposure score that weights exceptionally dense regions more heavily, or that gives a bonus to the land use individuals experience in college might prove to be better predictors of vehicle ownership than the metrics used in this study. There is also the potential to examine the behavior of households where the members have substantially different past exposure scores; in this case it would be of great interest to see which preference "wins out," or whether the household expresses the behaviors of a low-density or a high-density household. In terms of self-selection, our analysis suggests that in the Atlanta region, individuals are — for the most part — living in neighborhoods that are similar to those they have lived in previously. This may not be the case for regions in which particular land uses are under-supplied in the market. If the supply for dense neighborhoods with non-vehicle transportation options cannot meet the demand, then many individuals who would like to live in these neighborhoods would be unable to do so. In this situation, the "extreme" or “one opportunity” the individual had to live in a dense area may be a stronger reflection of the individual’s preference, and thus more influential on predicting current vehicle ownership.

8 Conclusions

Transportation planners have identified urban densification as a policy mechanism – among other strategies – to reduce vehicle ownership and/or use. These professionals cite studies that have shown a correlation between high densities and low vehicle ownership. However, this correlation is by itself an insufficient basis for policy. If households moving into new densely-populated neighborhoods have previously developed a preference for *high* vehicle ownership, they may simply continue to express those preferences in their new neighborhood. Or it may be that the individuals who do move to new densely-populated neighborhoods would already have owned fewer vehicles anyway, because that is the lifestyle they have selected to live. Our study indirectly considers these questions, by relating a household’s past experiences to its current behavior. We have shown that preferences for vehicle ownership are, at least to some extent, a reflection of prior exposure. Households with prior exposure to higher densities and non-vehicle transportation modes are less likely to own multiple vehicles, all else equal. However, the measurable effect of the prior exposure to higher densities and non-vehicle shares, at an individual level or aggregated across the region, is modest.

The central policy implication of our study is that proposed increases in density may have a marginally smaller short-run effect on forecasted vehicle ownership than would be predicted by models that did not consider the new residents’ prior experiences. Another implication is that achieving low vehicle ownership rates in a new development is more likely with residents who have previously lived in high-density areas, or who have had exposure to non-vehicle transportation modes.

Our models showed that very large changes in current density and transportation mode share would have a small effect on vehicle ownership when compared with other attributes, such as the number of adults in the household. Thus, planners seeking to limit vehicle ownership should consider other policies in addition to (or instead of) increasing density and providing non-vehicle alternatives. Other strategies such as increasing registration fees (Chin and Smith, 1997) or restricting the flow of vehicle traffic (Salon, 2009) into certain areas are likely to have a greater effect on reducing vehicle ownership. The best policy agenda is likely to be a multi-faceted approach, of which densification may be only one part.

Acknowledgements

We would like to thank Brian Stone and the anonymous referees for helpful comments on earlier versions of this paper. This project was partially funded by a U.S. Department of Transportation Eisenhower Fellowship.

References

Adjemian, M. K., Lin, C.-Y. C., Williams, J., 2010. Estimating spatial interdependence in automobile type choice with survey data. *Transportation Research Part A: Policy and Practice 44* (9), 661–675.

Babbie, E. R., 2009. *The Practice of Social Research*, 12th Edition. Wadsworth Publishing Company, Belmont, CA.

Bagley, M. N., Mokhtarian, P. L., 2002. The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *The Annals of Regional Science* 36 (2), 279–297.

Bhat, C. R., Guo, J., 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological* 41 (5), 506–526.

Bhat, C. R., Pulugurta, V., 1998. A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. *Transportation Research Part B: Methodological* 32 (1), 61–75.

Binder, S., Macfarlane, G. S., Garrow, L. A., Bierlaire, M., 2014. Associations among household characteristics, vehicle characteristics and emissions failures: An application of targeted marketing data. *Transportation Research Part A: Policy and Practice* 59, 122–133.

Buckeridge, D. L., Glazier, R., Harvey, B. J., Escobar, M., Amrhein, C., Frank, J., 2002. Effect of motor vehicle emissions on respiratory health in an urban area. *Environmental Health Perspectives* 110 (3), 293–300.

Cao, J., and Cao, X., 2014. The Impacts of LRT, Neighbourhood Characteristics, and Self-selection on Auto Ownership: Evidence from Minneapolis-St. Paul. *Urban Studies 51*(10), 2068-2087.

Cao, X., Mokhtarian, P. L., Handy, S. L., 2007a. Cross-sectional and quasi-panel explorations of the connection between the built environment and auto ownership. *Environment and Planning A* 39 (4), 830–847.

Cao, X., Mokhtarian, P. L., Handy, S. L., 2007b. Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach. *Transportation* 34 (5), 535–556.

Cao, X., Mokhtarian, P. L., Handy, S. L., 2009. Examining the impacts of residential self-selection on travel behaviour: A focus on empirical findings. *Transport Reviews* 29 (3), 359–395.

Chapman, L., 2007. Transport and climate change: a review. *Journal of Transport Geography* 15 (5), 354–367.

Chin, A. T., Smith, P., 1997. Automobile ownership and government policy: The economics of Singapore's vehicle quota scheme. *Transportation Research Part A: Policy and Practice* 31 (2), 129–140.

Croissant, Y., 2013. mlogit: multinomial logit model. R package version 0.2-4. URL http://CRAN.R-project.org/package=mlogit

Dieleman, F. M., Dijst, M., Burghouwt, G., 2002. Urban form and travel behaviour: micro-level household attributes and residential context. *Urban Studies* 39 (3), 507 – 527. 25

Eluru, N., Bhat, C. R., Pendyala, R. M., Konduri, K. C., 2010. A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. *Transportation* 37 (4), 603–626.

Ewing, R., Cervero, R., 2001. Travel and the built environment. *Transportation Research Record* 1780, 87–114.

Giuliano, G., Dargay, J., 2006. Car ownership, travel and land use: a comparison of the US and Great Britain. *Transportation Research Part A: Policy and Practice* 40 (2), 106–124.

Groves, R. M., 1989. *Survey Errors and Survey Costs*. John Wiley & Sons, New York.

Handy, S., Cao, X., Mokhtarian, P. L., 2006. Self-selection in the relationship between the built environment and walking: Empirical evidence from northern California. *Journal of the American Planning Association* 72 (1), 55–74.

Hausman, J., McFadden, D., 1984. Specification tests for the multinomial logit model. *Econometrica* 52 (5), 1219–1240.

Kenworthy, J. R., Laube, F. B., 1996. Automobile dependence in cities: an international comparison of urban transport and land use patterns with implications for sustainability. *Environmental Impact Assessment Review* 16 (4-6), 279–308.

Kressner, J. D., Garrow, L. A., 2012. Lifestyle segmentation variables as predictors of home-based trips for Atlanta, Georgia, airport. *Transportation Research Record* 2266, 20–30.

Kressner, J. D., Garrow, L. A., 2015. Using third-party data for travel demand modeling: a comparison of targeted marketing, Census, and household travel survey data. *Transportation Research Record* 2442, 8-19.

Krizek, K. J., 2003. Relocation and Changes in Urban Travel. *Journal of the American Planning Association* 69 (3), 265–281.

Leonhardt, D., Jul. 22 2013. In climbing income ladder, location matters. *New York Times*.

Lois, D., López-Sáez, M., 2009. The relationship between instrumental, symbolic and affective factors as predictors of car use: A structural equation modeling approach. *Transportation Research Part A: Policy and Practice* 43 (9-10), 790–799.

Manski, C. F., Lerman, S. R., 1977. The Estimation of Choice Probabilities from Choice Based Samples. *Econometrica* 45 (8), 1977–1988.

Matas, A., Raymond, J.-L., Roig, J.-L., 2009. Car ownership and access to jobs in Spain. *Transportation Research Part A: Policy and Practice* 43 (6), 607–617.

McFadden, D. L., 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P. (Ed.), *Frontiers in Econometrics*. Academic Press, New York, pp. 105–142.

Mokhtarian, P. L., Cao, X., 2008. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological* 42, 204–228.

Norman, J., MacLean, H. L., Kennedy, C. A., 2006. Comparing high and low residential density: Life-cycle analysis of energy use and greenhouse gas emissions. *Journal of Urban Planning and Development* 132, 10–21.

Pinjari, A. R., Bhat, C. R., Hensher, D. A., 2009. Residential self-selection effects in an activity time-use behavior model. *Transportation Research Part B: Methodological* 43 (7), 729–748.

Potoglou, D., Susilo, Y. O., 2008. Comparison of vehicle-ownership models. *Transportation Research Record* 2076, 97–105.

Roorda, M., Carrasco, J., Miller, E. J., 2009. An integrated model of vehicle transactions, activity scheduling and mode choice. *Transportation Research Part B: Methodologica*l 43 (2), 217–229.

Salon, D., 2009. Neighborhoods, cars, and commuting in New York City: a discrete choice approach. *Transportation Research Part A: Policy and Practice* 43 (2), 180–196.

Sanchez, T. W., Stolz, R., Ma, J. S., 2003. Moving to equity: Addressing inequitable effects of transportation policies on minorities. The Civil Rights Project at Harvard University, Cambridge, MA.

Schrank, D., Eisele, B., Lomax, T., 2012. *Urban Mobility Report*. Texas A&M University Transportation Institute, College Station, TX.

Sexton, S. E., Sexton, A. L., 2014. Conspicuous conservation: The Prius halo and willingness to pay for environmental bona fides. *Journal of Environmental Economics and Management* 67(3), 303-317.

Small, K. A., Hsiao, C., 1985. Multinomial logit specification tests. *International Economic Review* 26 (3), 619–627.

Stone, B., 2009. Land use as climate change mitigation. *Environmental Science & Technology* 43 (24), 9052–9056.

Transportation Research Board (TRB), 2009. *Driving and the built environment: The effects of compact development on motorized travel, energy use, and CO2 emissions.* Transportation Research Board Special Report 298. Washington, D.C.: National Research Council.

U.S. Census Bureau, 2000a. Form D-61B: Census 2000 "long" form questionnaire.

U.S. Census Bureau, 2000b. H001. Housing Units: Census 2000 Summary File 1 (SF1) 100-percent data.

U.S. Census Bureau, 2000c. P030. Means of Transportation to Work for Workers 16 Years and Older: 2000 Census Summary File 3 (SF3) - Sample Data. All 5-Digit ZIP Code Tabulation Areas fully-or-partially within United States.

U.S. Census Bureau, 2011. B08101. Means of Transportation to Work by Age: 2011 ACS 5-year estimates. All ZIP Code Tabulation Areas within the United States.

U.S. Census Bureau, 2012a. American Community Survey 2006-2010 ACS 5-year PUMS files.

U.S. Census Bureau, 2012b. B25001. Housing Units: 2007-2011 American Community Survey 5-year estimates.

Van Acker, V., Witlox, F., 2010. Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship. *Journal of Transport Geography* 18 (1), 65–74.

Vuong, Q. H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57 (2), 307–333.

Weinberger, R., Goetzke, F., 2010. Unpacking preference: How previous experience affects auto ownership in the United States. *Urban Studies* 47 (10), 2111–2128.

Appendix: Sensitivity Analysis

We tested the sensitivity of model results to several data processing and modeling assumptions. Since we know that the number of children variable is incomplete for some households that do in fact have children, we applied various imputation methods to correct for this issue; however none of the imputation methods resulted in different model interpretations.

We used the duration exposure metric to test the sensitivity of results to our "previous move" assumptions, described in Section 3.3. Specifically, we calculated the duration exposure metric for the earliest possible move-in date and latest possible move-in dates using $\alpha = 1$, which applies an equal weight to each prior exposure day. The impact on the exposure metric was minimal: only 497 of the 227,830 records changed by more than 1%. Since results were robust to both of our previous move assumptions, we only report those models that use the latest move assumption in Table 4.

The MNL model imposes the "independence of irrelevant alternatives" (IIA) assumption, requiring that the unobserved characteristics of each alternative j (ε_{ij}) are independent of both the unobserved characteristics ($\varepsilon_{ij'}$) and the observed utility ($V_{ij'}$) for all other alternatives j'. This assumption may be violated for ordinally-related alternatives, since it is plausible that similar unobserved characteristics would influence the choice between adjacent alternatives in particular. To test whether the MNL model was appropriate, we applied a Hausman-McFadden IIA test (Hausman and McFadden, 1984) comparing the estimates of each exposure model to models with an identical specification but an alternative removed. The tests failed by producing a negative statistic; Small and Hsiao (1985, p. 619) point out that such computational failures are not uncommon with the Hausman-McFadden test, given that it "requires inversion of the difference between two closely related matrices [the variance-covariance matrices of the coefficient estimators for the reduced-choice-set and full-choice-set models], which may be non-positive-definite or nearly singular." We also tested a nested specification of the extreme model that allowed for correlated substitution between the two- and three-or-more-vehicle alternatives. The estimated nest substitution parameters were all greater than 1, implying a violation of random utility theory, and the other coefficients were not materially different from the extreme exposure MNL model. We therefore retain the MNL specification.

List of Figures

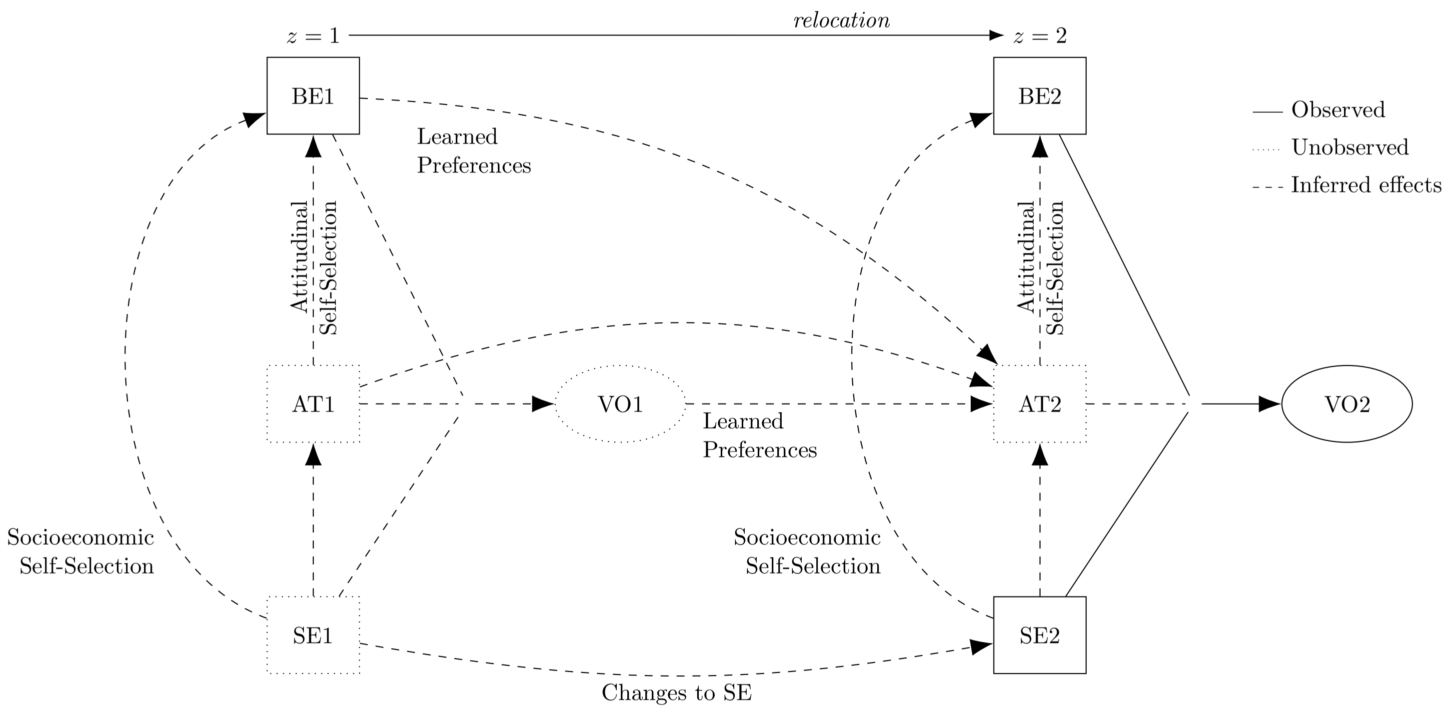

Figure 1: A conceptual model of self-selection and learned preferences for vehicle ownership decisions.

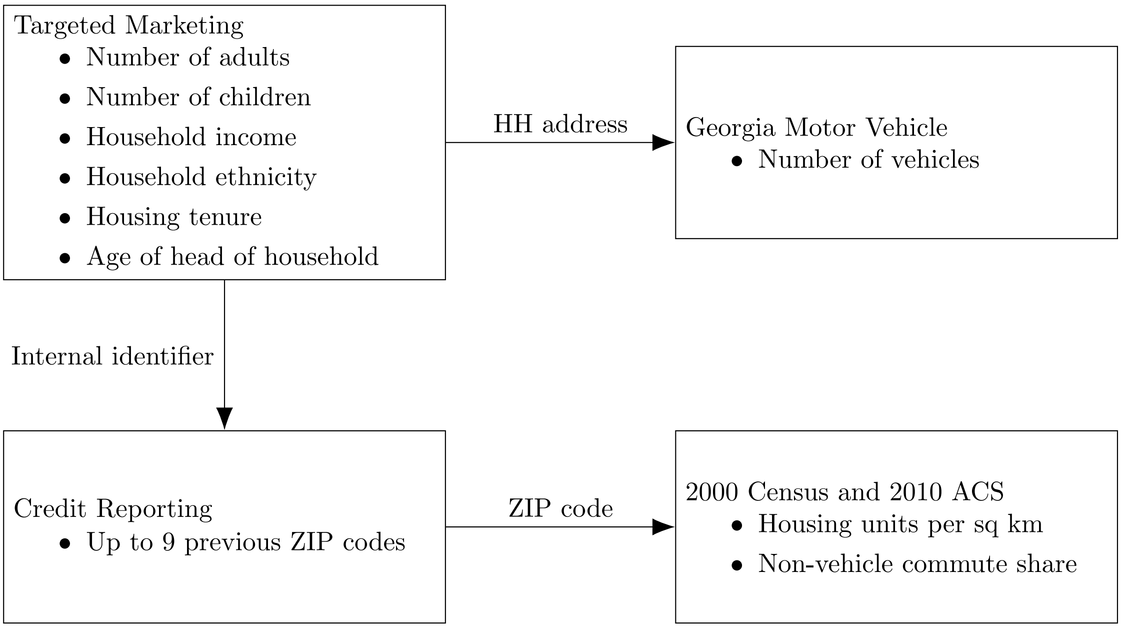

Figure 2: Compilation of analysis dataset.

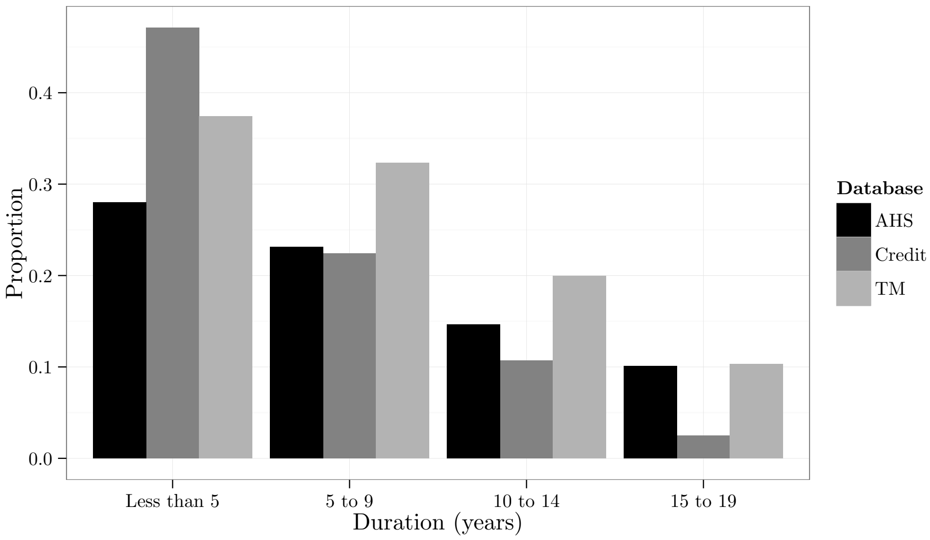

Figure 3: Length of residence in three different databases.

List of Tables

Table 1: Descriptive statistics for quasi-continuous variables.

Quasi-Continuous Variables	Source	Mean	Std. Dev.	1%	99%	Notes
Household adults	Targeted marketing	1.8	0.77	1	4	
Household children	Targeted marketing	0.49	0.92	0	4	
Household income (USD)	Targeted marketing	80,196	45,488	10,000	225,000	Given in ranges, bottom-coded at $10k and top-coded at $250k. We use the median of the range.
Householder age	Targeted marketing	50	13	26	82	The age in years of the primary adult in the household.
Current density	Census	299	258	20	1,183	The number of housing units in a ZIP code divided by area measured in square kilometers.
Exposure: density						
Duration $\alpha = 1$	Calculated using prior addresses from credit reports and density from Census data.	426	813	16	3,242	Each prior residence receives weight proportional to length of residence.
Duration $\alpha = 0$ *(Mean)*		424	716	19	3,100	Each prior residence receives equal weight.
Decay $\gamma = -0.5$		426	816	16	3,254	Older residences receive *less* weight than newer ones.
Decay $\gamma = 0.5$		426	816	16	3,241	Older residences receive *more* weight than newer ones.
Extreme		704	1,537	24	7,064	The highest density to which the householder has been exposed.
Longest		430	989	8.9	3,393	The density of the ZIP code with the longest time of residence.
Current non-vehicle share	Census	0.12	0.054	0.051	0.28	The percentage of work commuters who did not drive alone or carpool.
Exposure: non-vehicle share						
Duration $\alpha = 1$	Calculated using prior addresses from credit reports and non-vehicle shares from Census data.	0.11	0.083	0.032	0.46	Each prior residence receives weight proportional to length of residence.
Duration $\alpha = 0$ *(Mean)*		0.11	0.075	0.031	0.41	Each prior residence receives equal weight.
Decay $\gamma = -0.5$		0.11	0.083	0.032	0.46	Older residences receive *less* weight than newer ones.
Decay $\gamma = 0.5$		0.11	0.083	0.032	0.46	Older residences receive *more* weight than newer ones.
Extreme		0.15	0.13	0.035	0.76	The highest non-vehicle share to which the householder has been exposed.
Longest		0.11	0.096	0.031	0.55	The non-vehicle share in the ZIP code with the longest time of residence.

Table 2: Descriptive statistics for categorical variables.

Discrete Variables	Source	Number	%
Vehicles	Georgia Motor Vehicle		
1 Vehicle		69,899	31
2 Vehicle		91,522	40
3+ Vehicles		66,409	29
Ethnicity	Targeted marketing		
White		169,758	75
African-American		33,893	15
Asian		5,376	2.4
Hispanic		7,960	3.5
Other		10,843	4.8
Housing tenure	Targeted marketing		
Owner		214,194	94
Renter		2,645	1.2
Probable owner		2,585	1.1
Probable renter		8,406	3.7

Table 3: Base vehicle ownership model.

Generic Variables[1]	β		t-stat	
Insufficiency (# adults without vehicles)	-1.088^{***}		-146.0	
	2 Vehicles		3+ Vehicles	
Alternative-Specific Variables	β	t-stat	β	t-stat
(Intercept) *ref. 1 vehicle*	-4.236^{***}	-31.1	-10.496^{***}	-64.7
Number of Children	0.127^{***}	19.5	0.071^{***}	9.9
log(Income)	0.371^{***}	39.4	0.452^{***}	42.4
Household Ethnicity: *ref. White*				
African-American	-0.088^{***}	-5.5	0.050^{**}	2.8
Asian	0.426^{***}	11.5	0.725^{***}	18.5
Hispanic	0.318^{***}	10.4	0.658^{***}	20.3
Other	0.205^{***}	8.1	0.171^{***}	6.0
Housing Tenure: *ref. Known owner*				
Known renter	-0.623^{***}	-13.3	-0.852^{***}	-13.4
Probable renter	-0.568^{***}	-11.5	-1.077^{***}	-12.9
Probable owner	0.080^{**}	2.9	0.281^{***}	9.2
Householder Age	0.012^{***}	5.0	0.186^{***}	54.3
Householder Age^2	$-1.92e{-}04^{***}$	-8.3	$-1.73e{-}03^{***}$	-53.8
log(Current Density)	-0.119^{***}	-14.2	-0.212^{***}	-23.0
log(Current Non-Vehicle Share)	-0.119^{***}	-7.2	-0.417^{***}	-22.1
Number of Observations	227,830			
Null Log-likelihood	$-250,297$			
Constant Log-likelihood	$-247,926$			
Model Log-likelihood	$-225,528$			
ρ^2_C	0.0903			

** significant at $p < .01$; $^{***}p < .001$

[1] Generic variables take a single coefficient value applicable to all alternatives.

Table 4: Models incorporating exposure metrics.

	Duration ($\alpha = 0.1$)		Decay ($\gamma = 0$)		Extreme		Longest	
Variables	β	t-stat	β	t-stat	β	t-stat	β	t-stat
2 Vehicles								
log(Current Density)	-0.115^{***}	-12.7	-0.114^{***}	-12.7	-0.111^{***}	-12.6	-0.119^{***}	-13.5
log(Past Density)	-0.004	-0.5	-0.008	-1.0	-0.011	-1.4	0.002	0.3
log(Current Non-Vehicle Share)	-0.111^{***}	-6.5	-0.114^{***}	-6.7	-0.110^{***}	-6.5	-0.115^{***}	-6.8
log(Past Non-Vehicle Share)	-0.260^{*}	-2.3	-0.138	-1.3	-0.219^{**}	-3.0	-0.119	-1.4
$3+$ **Vehicles**								
log(Current Density)	-0.138^{***}	-13.9	-0.147^{***}	-14.9	-0.140^{***}	-14.4	-0.178^{***}	-18.4
log(Past Density)	-0.147^{***}	-16.3	-0.125^{***}	-15.1	-0.152^{***}	-18.3	-0.056^{***}	-8.6
log(Current Non-Vehicle Share)	-0.404^{***}	-20.9	-0.404^{***}	-21.0	-0.397^{***}	-20.8	-0.402^{***}	-21.0
log(Past Non-Vehicle Share)	0.054	0.4	-0.031	-0.3	-0.106	-1.2	-0.314^{**}	-3.3
Model Log-likelihood	$-225,258$		$-225,302$		$-225,150$		$-225,426$	
ρ^2_C	0.0914		0.0913		0.0919		0.0908	

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 5: Comparison of models that include current attributes and/or past exposure metrics.

	Current Only[1]		Past Only[2]		Both[3]	
Variables	β	t-stat	β	t-stat	β	t-stat
2 Vehicles						
log(Current Density)	-0.119^{***}	-14.2			-0.111^{***}	-12.6
log(Past Density)			-0.053^{***}	-7.3	-0.011	-1.4
log(Current Non-Vehicle Share)	-0.119^{***}	-7.2			-0.110^{***}	-6.5
log(Past Non-Vehicle Share)			-0.231^{**}	-3.2	-0.219^{**}	-3.0
$3+$ **Vehicles**						
log(Current Density)	-0.212^{***}	-23.0			-0.140^{***}	-14.4
log(Past Density)			-0.221^{***}	-28.2	-0.152^{***}	-18.3
log(Current Non-Vehicle Share)	-0.417^{***}	-22.1			-0.397^{***}	-20.8
log(Past Non-Vehicle Share)			-0.273^{**}	-3.3	-0.106	-1.2
Model Log-likelihood	$-225,528$		$-225,835$		$-225,150$	
ρ_C^2	0.0903		0.0891		0.0919	

* significant at $p < .05$; $^{**}p < .01$; $^{***}p < .001$

[1] This is precisely the Base model from Table 3.

[2] Using the "extreme" exposure metric.

[3] This is precisely the Extreme model in Table 4.

Table 6: Error resulting from change in relationship between past and current exposure.

Scenario	Total Vehicles	%Δ from Reference
Reference	$452,170$	
Past equals present	$457,537$	1.2
Mean past	$446,748$	-1.2
Random past	$452,016$	-0.034
1^{st} percentile	$478,380$	5.8
5^{th} percentile	$469,393$	3.8
95^{th} percentile	$437,524$	-3.2
99^{th} percentile	$425,169$	-6.0