



Munich Personal RePEc Archive

The Impact of Residential Density on Vehicle Usage and Fuel Consumption: Evidence from National Samples

Kim, Jinwon and Brownstone, David

Technical University of Denmark, University of California, Irvine

17 June 2013

Online at <https://mpra.ub.uni-muenchen.de/47785/>

MPRA Paper No. 47785, posted 26 Jun 2013 04:01 UTC

The Impact of Residential Density on Vehicle Usage and Fuel Consumption: Evidence from National Samples

Jinwon Kim

Department of Transport, Technical University of Denmark

David Brownstone¹

Department of Economics, University of California, Irvine

First Version: March, 2010

Current Version: June, 2013

Abstract

This paper investigates the impact of residential density on household vehicle usage and fuel consumption. We estimate a simultaneous equations system to account for the potential residential self-selection problem. While most previous studies focus on a specific region, this paper uses national samples from the 2001 National Household Travel Survey. The estimation results indicate that residential density has a statistically significant but economically modest influence on vehicle usage, which is similar to that in previous studies. However, the joint effect of the contextual density measure (density in the context of its surrounding area) and residential density on vehicle usage is quantitatively larger than the sole effect of residential density. Moving a household from a suburban to an urban area reduces household annual mileage by 18%. We also find that a lower neighborhood residential density induces consumer choices toward less fuel-efficient vehicles, which confirms the finding in Brownstone and Golob (2009).

Keywords: Household vehicle choice, simultaneous equations systems, residential density

JEL Codes: D12, C31, R41

¹Brownstone is corresponding author. Address for correspondence: Department of Economics, 3151 SSPA, University of California, Irvine, California 92697-5100, USA. Email: dbrownst@uci.edu. Fax: +1-949-824-2182. Financial support from the University of California Transportation Center is gratefully acknowledged. We wish to thank Jan Brueckner, Kenneth Small, and two anonymous referees for many useful comments on earlier drafts. Of course, the authors bear full responsibility for any errors or opinions expressed in this paper. The previous version of this paper has been circulated with the title, "the impact of residential density on vehicle usage and fuel consumption" (UCTC research paper number UCTC-FR-2010-05).

1. Introduction

How does urban sprawl affect household travel behavior? This paper addresses this question by investigating the impact of land use density on household annual mileage traveled and fuel consumption. Following previous studies (Brownstone and Golob, 2009, Boarnet and Crane, 2001), we use land use density as the measure of urban spatial structure (or urban sprawl). Although urban sprawl is not simply low density, land use density is highly correlated with almost all measures of urban sprawl (see Badoe and Miller 2001). Most of the previous studies that attempt to measure the influence of urban spatial structure on vehicle usage focus on specific regions in order to guarantee geographic homogeneity (Boarnet and Crane, 2001, Bhat and Guo, 2007, Brownstone and Golob, 2009, Salon, 2009). This study analyzes national level data, so we control for geographic heterogeneity by including a set of urban/rural dimension dummies (the contextual density measure) and rail transit dummies.

The most important econometric issue is possible residential self-selection, and this occurs if residents of high-density areas differ in some unobservable characteristics that influence travel behavior. One possible behavior that leads to self-selection bias would be households who dislike automobile travel locating in dense urban areas with good transit. Unless residential self-selection is controlled, the estimated influence of land use density on travel behavior may be spurious. We follow the same methodology as in Brownstone and Golob (2009) to correct for the self-selection bias by specifying a simultaneous equation model where residential density, household mileage traveled, and fuel consumption are jointly endogenous. These three endogenous variables are assumed to be influenced not only by other endogenous variables but also by a rich set of socio-demographic variables. Among various socio-demographic variables, education dummies are key variables to identify the simultaneous equations system. In particular, our final model is consistent with the assumption that education only impacts fuel use and not vehicle miles travelled or density conditioned on number of workers, children, income, race, number of drivers, and urban structure. This is equivalent to assuming that education only impacts the choice of fuel efficiency. This assumption is partially tested using overidentification tests described in Section 3. Our final model also assumes that density is exogenous to miles travelled or fuel use, which is consistent with the behavioral assumption that households first choose residential location (and therefore density) and then choose their vehicle type and use

conditional on this choice. This assumption is consistent with other disaggregate studies conditioning on a broad set of socioeconomic variables (Bento et al. ,2005, and Bhat and Guo, 2007) but is tested in this paper using various tests described in Section 3.

The other econometric issue that may result in biased coefficients is the non-random missing data in the key endogenous vehicle fleet characteristics (needed to compute fuel use). To correct for the bias caused by this problem, weights are estimated to compensate for the higher probability of missing data for households owning many vehicles and weighted estimation is used. The wild bootstrap method is used to estimate standard errors that are robust to heteroscedasticity.

Although we mostly adopt the methodologies that were used in Brownstone and Golob (2009), our study is more extensive in two aspects. First, we use national level data from the 2001 National Household Travel Survey (NHTS) instead of just the California subsample. The big advantage of using national level data is that with the increased sample size, we can specify a larger, more accurate model. We can also check whether travel behavior in a specific region is unique or not by comparing those two results. We provide comparisons between the results of this paper and those in the California analysis of Brownstone and Golob (2009), which share the same specifications. We do not use the more recent 2009 NHTS survey because this survey did not collect vehicle miles travelled using dual odometer readings. The 2009 NHTS then imputed vehicle miles travelled based on a single odometer reading and self-reported miles travelled for each vehicle using an imputation model based on the 2001 NHTS data. Our preliminary work with the 2009 NHTS vehicle miles travelled data show that this imputed variable is less reliable.

This paper further extends Brownstone and Golob (2009) by including various geographic control variables such as census region, MSA category, urban/rural dimension and rail transit dummies. Since these variables are included, we can investigate the effects of certain geographic or institutional conditions, such as the supply of rail transit, on travel behavior. We find that the urban/rural dimension variables have a large impact on vehicle usage. Our results are consistent with the view expressed in Transportation Research Board (2009) that changing density without changing other aspects of urban structure has very little impact on vehicle and fuel usage. However, our simulation results in Tables 8 and 9 show substantial impacts from changing urban structure (as measured by our urban/rural dimension variables).

The estimation results show that residential density has a statistically significant but economically modest influence on vehicle usage and fuel consumption, which is similar to that in other previous studies. However, the joint effect of the urban/rural dimension variable (contextual density measure) and residential density is much greater than the sole effect of residential density. A simulation moving a household across the urban/rural dimension affects household annual mileage traveled and fuel consumption significantly. This result suggests that residential density in a wider geographic scope, which takes into account of density of surrounding areas, is important in influencing household travel behavior. Compared to the California subsample result of Brownstone and Golob (2009), the influence of residential density on mileage traveled is slightly higher. However, the impact of residential density on vehicle type choice, i.e. tendency toward more fuel efficient vehicle choices for households in denser area, is quantitatively smaller than that in the California subsample.

1.1. Literature review

Studies of the effects of land use density (or other measures of urban spatial structure) on vehicle usage can be divided into aggregate and disaggregate studies. Transportation Research Board (2009) reviews many of these studies, and Ewing and Cervero (2010) carry out a meta-analysis of some disaggregate studies. Aggregate studies use spatially defined averages for all variables. One of the most cited papers is Newman and Kenworthy (1999), where the authors implemented a global survey of 37 cities to assess automobile dependence cost. The results indicate that cities with more car use, road provision, and urban sprawl have higher automobile dependence, which causes direct and indirect costs in terms of higher road expenditures, more time spent on commuting, and higher external costs from road deaths and emissions.

Disaggregated studies use household observations of vehicle usage and either city-wide, zonal, or neighborhood averages for urban form variables. Bento et al. (2005) specify disaggregate models of commute mode choice, automobile ownership and annual vehicle miles traveled (VMT). They construct diversified measures of urban form and transit supply: measures of city shape, density of the road network, spatial distribution of population, jobs-housing balance, and bus route and rail miles supplied. Using the 1990 National Personal Transportation Survey, they find that the impacts of any of the urban form measures on travel behavior are

frequently insignificant and small in magnitude.

Although disaggregate studies (Bento et al., 2005) that include a rich set of socioeconomic control variables are less subject to residential self-selection bias, it is still possible that residents in high density areas differ in some unobservable characteristics that influence their travel behavior. The only way to deal with this possibility is to construct a joint model of residential density and travel behavior. One of the first to do this is Boarnet and Crane (2001). They specify a demand function for travel in which the number of trips of different travel modes are influenced by the relative time costs (price of travel) and various socio-demographic “taste” variables. By comparing models where land use density is endogenous and exogenous, they find that the measured influence of land use on travel behavior is very sensitive to how endogeneity is treated.

Bhat and Guo (2007) specify a joint mixed multinomial logit model of residential location and number of household vehicles. Their model allows for residential self-selection effects (correlation between the error terms in their equations), but after controlling for a rich set of covariates they do not find any significant effects of residential self-selection. This result implies the necessity of including a rich set of socio-demographic variables to control for residential self-selection. Using San Francisco Bay Area data, they find statistically significant but quantitatively small impacts of built environment measures (street block density, transit availability, and transit access time) on vehicle ownership. Salon (2009) also addresses the simultaneity concern by modeling the joint choice of residential location, car ownership, and commute mode. Salon (2009) estimates her model using samples from New York City residents.

Finally, Brownstone and Golob (2009) directly model the joint choice of residential density and vehicle usage to control for potential residential selectivity. Unlike other previous studies, they also explicitly model vehicle fuel consumption to account for the possibility that residents of high density areas choose more fuel efficient vehicles. Additionally, by adopting a weighting approach, they correct for the bias caused by systematic missing data problems. Using the California subsample of the 2001 NHTS, they find a statistically significant but quantitatively small impact of residential density on household vehicle usage and fuel consumption.

Unlike the previous studies using subsamples of a specific region, this paper uses national level data from the 2001 NHTS. While this paper follows the methodology used in Brownstone

and Golob's (2009), our empirical model additionally includes various geographic control variables that are necessary in analyzing national level data. We first confirm the finding in the previous studies that residential density has a statistically significant and economically modest influence on vehicle usage. A new finding in this paper is that the urban/rural dimension dummies have considerable influences on household annual mileage and fuel consumption. Our findings suggest that the effect of density in a small geographic scope is limited but the joint influence of density and densities of surrounding areas are economically significant.

This paper is organized as follows. Section 2 discusses the data used in the study. Section 3 describes the empirical model and the procedure to correct for the self-selection bias and the non-random missing data problem. Section 4 gives estimation results, section 5 discusses additional robustness checks, and section 6 concludes.

2. Data

The National Household Travel Survey (NHTS) conducted by the U.S. Department of Transportation contains information on household travel behavior and various socio-demographic variables. The latest survey was conducted in 2008-9 and released in late 2010.² We used the 2001 NHTS sample for this paper since the 2009 NHTS has a less reliable measure of vehicle miles traveled (especially for older vehicles). Hong (2011) shows that applying the model specification in this paper to the 2009 NHTS yields very similar conclusions.

The 2001 NHTS consists of 69817 household observations. Of these households, 26083 are in the national sample and 43734 are from nine add-on areas. The add-on sample is added to help in-depth research for selected states. Among 26083 households in the national sample, 21350 (82%) households have full information on the key endogenous vehicle fleet characteristics (needed to compute fuel use). Of these households with full information, 4992 observations are randomly chosen for analysis.³ We are reserving the remainder of the sample to test the possibility that we have "overfit" the data during the process of selecting the best model.

²The NHTS data are described in detail on the NHTS website at <http://nhts.ornl.gov/>.

³ Since the random sample is chosen from households with full information, the random sample may not match the population.

2.1. Vehicle ownership, mileage traveled, and fuel consumption

Household annual mileage and fuel consumption are the key vehicle fleet characteristics that are used as the joint endogenous variables in our model. Household annual mileage is defined as total mileage per year of all household vehicles. Annual vehicle mileage traveled is recorded based on two odometer readings for each vehicle. Fuel usage computation is based on information about the make, model and vintage of all household vehicles. The U.S. Energy Information Administration matched the NHTS vehicles to other data giving fuel efficiency for all vehicles sold in the U.S.

Because household annual mileage and fuel consumption are calculated by summing up all household vehicle information, there is no need to explicitly use vehicle ownership as a variable. However, vehicle ownership is closely related to missing data on the key endogenous variables. For example, for households owning one vehicle, 93.5% have full information on the key endogenous variables. However, for households with two, three, four, and five or more vehicles, 87.5%, 66.9%, 51.4%, 29.6% have full information on the key vehicle fleet characteristics, respectively. The probability of having full information is a decreasing function of the number of household vehicles. Since vehicle ownership is closely related to the key endogenous variables, this suggests that the sample of households with complete energy information is not a random sample, and this can lead to biased results. The methodology to correct for the bias caused by this non-random missing data problem is presented in section 3.

2.2. Land-use density

The 2001 NHTS data include population and housing density at the census and block level. The percentage of renter occupied-housing units is provided at both the block group and tract group level, and jobs per square mile is provided at the tract level. As expected, the density variables are all highly correlated.

Table 1 presents percentage of households residing in each density group (housing density at the census and block level) by different geographic regions. As expected, more people reside in denser areas in big cities such as New York and Chicago. We can also find that the urban/rural dimension variable, which categorizes geographic regions into urban, second city, suburban,

town, and rural, is highly correlated with residential density. Residential density for rail cities is slightly higher than for non-rail cities.

Table 1: Percentage of households residing in density groups by different geographic regions

Housing units per square mile in						
Census block group	0 to 50	50 to 250	250 to 1K	1K to 3K	3K to 5K	over 5K
All households	17	16	22	30	8	7
Households residing in New York	1	9	20	27	9	34
Households residing in Chicago	2	7	21	38	13	20
Households residing in Urban	0	0	4	30	25	41
Households residing in Second city	1	6	24	51	12	6
Households residing in Suburban	0	4	31	52	9	3
Households residing in Town	8	36	37	36	1	0
Households residing in Rural	65	24	8	3	0	0
Households residing in Rail city	2	10	20	34	13	21
Households residing in Non-rail city	21	18	23	30	6	2

Source: 2001 NHTS, national sample (26038 observations)

2.3. Urban/rural dimension and rail transit dummies

Since there exists heterogeneity in geographic conditions such as climate, vehicle, fuel prices, and access to public transit, travel behavior may be influenced by these variables. For example, travel behavior in New York may be much different than in California. However, we do not know whether this difference comes from residential density or other geographic conditions. To control for these conditions, we use various geographic control variables such as census region, MSA category, urban/rural dimension and rail transit dummies.

The urban/rural dimension variables were devised by Claritas, Inc. to establish objective classifications of geographic locations that were less-boundary dependent and more life-style relevant. These variables define five major classifications: urban, suburban, second city, town, and rural. In turn, each of the nation's 226399 block groups was assigned to one of these categories. The classification is based on density in the context of its surrounding area, yielding a “contextual density measure”. The contextual density measure was based on population

density, not of specific block group, but of the larger geographic area not constrained by boundary definitions. Thus, the measure ensures relevance to the density experienced by people. A heuristic approach was taken to determine meaningful breaks for defining each category. See Miller and Hodge (1994) for more detailed description. Table 2 shows average land use densities and selected demographics by the urban/rural classification. Table 3 presents the two-way descriptive tables of block group residential density and density of wider geographical area indicated by the urban/rural dimension dummies.

The rail transit dummy indicates that the household lives in an MSA where rail transit is supplied. Among all national households, 24% have access to rail transit. 61% of households in urban areas have access to rail transit while only 3% in rural areas are offered rail transit. In many empirical specifications we tested, this rail transit dummy seldom works in explaining household travel behavior. We suspect that this is because the variable is too crude since it includes areas with only small light rail systems. So, we constructed another rail transit dummy variable which includes only the New York, Chicago, Boston, Washington, and Philadelphia MSAs. The 'Extensive Rail MSAs' dummy indicates that the household lives in one of these large old cities. These five MSAs are chosen because they have the most extensive subway and commuter rail systems. Among all national households, 15% reside in these Extensive Rail MSAs, and 35% of households in urban area live in these MSAs.

Table 2: Average land use densities and selected demographics by urban/rural dimension classifications

Urban/rural dimension	Urban	Second city	Suburban	Town	Rural
Housing units per sq. mile - Block group	4087	2050	1828	742	170
Housing units per sq. mile - Tract level	3959	1766	1669	539	85
Population per sq. mile - Block group	14095	5125	4873	1622	354
Population per sq. mile - Tract level	13472	4325	4336	1082	182
Workers per sq. mile – Tract level	3611	1747	1747	486	68
Household annual mileage traveled	16042	19323	22687	25416	27923
Household annual fuel consumption in gallons	714	888	1057	1207	1336
Gallons of fuel used per 100 mile traveled	4.88	5.03	4.97	5.08	5.22
Household income (\$10000)	5.48	5.19	7.15	6.08	4.59
MSA has rail	0.61	0.20	0.36	0.16	0.03
Extensive Rail MSAs ¹⁾	0.35	0.13	0.23	0.12	0.02

Source: 2001 NHTS, national sample (26038 observations)

Notes: 1) Extensive Rail MSAs include New York, Chicago, Boston, Washington, and Philadelphia.

Table 3: Two-way descriptive table of block group residential density and urban/rural dimension dummies

Urban/rural dimension /Residential density	Urban	Second city	Suburban	Town	Rural	Total
0 to 50	40	3791	14	489	2	4336
50 to 250	277	1408	273	2142	12	4112
250 to 1K	1139	462	1900	2189	113	5803
1K to 3K	2447	161	3267	1100	964	7939
3K to 5K	567	8	587	84	783	2029
Over 5K	294	1	189	10	1316	1810
Total	4764	5831	6230	6014	3190	26029

Source: 2001 NHTS, national sample (26038 observations)

2.4. Vehicle usage and land-use density

Table 4 presents vehicle usage characteristics sorted by residential density groups. Average household annual mileage decreases as residential density increases. Average household annual fuel consumption is also lower in denser areas. The negative correlation is slightly greater for fuel consumption than mileage, which suggests that households in denser area use more fuel efficient vehicles.

Average vehicles per household reveal similar patterns to mileage and fuel consumption, but the pattern for average number of drivers is less clear than others. Average household incomes in medium density areas are higher than those of denser areas. Table 5 gives descriptive statistics of all variables used in the analysis.

Table 4: Vehicle usage by residential density

Housing units per square mile in Census block group	0 to 50	50 to 250	250 to 1K	1K to 3K	3K to 5K	Over 5K
Annual mileage of all household vehicles	25786	24553	22218	18897	14982	13431
Annual fuel consumption in gallons	1308	1217	1067	895	717	599
Vehicles per household	1.95	1.97	1.86	1.63	1.34	0.79
Average number of drivers	1.79	1.81	1.78	1.59	1.40	0.85
Household income (\$10000)	4.57	5.56	6.30	5.50	4.67	4.42

Source: 2001 NHTS, average across estimation sample (4992 observations)

Table 5: Descriptive statistics

Variables	Entire sample			Estimation sample (N=4992)	
	Observations	Mean	Std. Dev.	Mean	Std. Dev.
Household fuel usage per year in gallons	21370	1073	974	937	964
Housing units per square mile in units of 1000 - census block group	26029	1.52	1.66	1.88	1.94
Total mileage per year on all household vehicles	23837	22989	21172	19323	20425
Annual household income in units of \$10000	26038	5.77	4.70	5.32	4.64
Number of children	26038	0.65	1.05	0.59	1.03
Number of workers	26038	1.32	0.97	1.18	0.94
1-worker household	26038	0.33		0.34	
2-worker household	26038	0.36		0.33	
3-worker household	26038	0.08		0.06	
Number of drivers	26038	1.82	0.81	1.59	0.87
1-driver household	26038	0.28		0.31	
2-driver household	26038	0.55		0.49	
3-driver household	26038	0.13		0.09	
Respondent has college degree	26038	0.43		0.42	
Respondent has postgraduate degree	26038	0.11		0.11	
Respondent is retired	26038	0.28		0.30	
Single-person household	26038	0.12		0.15	
Race is Asian	26038	0.02		0.02	
Race is Hispanic	26038	0.03		0.04	
Race is Black	26038	0.07		0.09	
Race is mixed White or Hispanic	26038	0.03		0.03	
MSA has rail	26038	0.24		0.29	
Extensive Rail MSAs ¹⁾	26038	0.15		0.21	
Household resides in Rural ²⁾	26029	0.22		0.19	
Household resides in Second City	26029	0.18		0.19	
Household resides in Suburban	26029	0.24		0.22	
Household resides in Town	26029	0.23		0.22	
Household resides in Urban	26029	0.12		0.18	

Notes: 1) Extensive Rail MSAs include New York, Chicago, Boston, Washington, and Philadelphia. 2) The urban/rural dimension variable divides urban form into 5 categories: urban, second city, suburban, town, and rural. The classification is based on the contextual density measure (density in the context of its surrounding area). 3) Variables with missing Std. Dev. are dummy variables.

3. Empirical model

3.1. Simultaneous equation system

We follow the same methodology as in Brownstone and Golob (2009) to measure the influence of residential density on vehicle and fuel usage by specifying a simultaneous equation model with three endogenous variables and many exogenous variables. The three endogenous variables are housing units per square mile in the census block group (residential density, D), total annual mileage per year of all household vehicles (M), and total household annual fuel consumption per year (F). We also include various socio-demographic and geographic control variables (X). The system can be written as

$$\begin{aligned}M_i &= A_{1,3}D_i + B_1X_i + \varepsilon_{1,i} \\F_i &= A_{2,1}M_i + A_{2,3}D_i + B_2X_i + \varepsilon_{2,i} \\D_i &= B_3X_i + \varepsilon_{3,i},\end{aligned}\tag{1}$$

where i indicates a household observation. The model above can be written as a matrix form,

$$\begin{aligned}y_i &= Ay_i + BX_i + \varepsilon_i \\Cov(\varepsilon_i) &= \Omega,\end{aligned}\tag{2}$$

where y_i is the vector of the three endogenous variables, A and B are coefficient matrices, and ε_i is a vector of residuals with an unrestricted correlation structure.

To identify the simultaneous equation, the model is assumed to be a recursive system by restricting the A matrix in (2). The recursive system in effect assumes that the household first chooses residential location (D) and then decides annual mileage and fuel consumption. There are three paths of influences of residential density on mileage and fuel consumption. First, residential density affects annual mileage because miles per vehicle will be greater due to the separation of households and activity sites, and households in lower density areas will choose to have more vehicles, controlling for socioeconomic and demographic differences. Second, residential density directly affects fuel usage in that households that choose to live in denser areas also choose to own more fuel efficient vehicles, partly due to higher costs of maintaining larger vehicles in dense area or the relative difficulty of maneuvering and parking large vehicles. Finally, there is a direct link from mileage to fuel usage. These endogenous effects define a recursive system, so there are no identification problems in the absence of error term correlations.

We also impose additional restrictions on the B matrix to over-identify the system, but these restrictions are based on removing several insignificant variables (see Table 6). We implement tests to make sure that the over-identifying constraints are not violated.

The key feature of this model is that it accounts for residential self-selections of households. Residential self-selection occurs if residents of high-density areas differ in some unobservable characteristics that influence travel behavior. If this residential self-selection is prevalent, the estimated influences of density are just sorting effects, not a causal influence. The simultaneous equation system assumes that household characteristics that influence household residential decision also influence household travel behavior. In the context of the model, residential self-selections imply that the errors in the residential density equation are correlated with the errors in other equations. However, if we include all relevant variables that influence both the residential density and household travel behavior, the error correlations may be zero even in the presence of residential self-selection.

The model is first estimated using 3SLS with the restriction that error correlations are zero once a rich set of socioeconomic and geographic exogenous variables are controlled for.⁴ Then, all of the over-identifying restrictions (including the restriction that error correlations are zero) pass the specification test described later in this section. The restriction of a diagonal matrix also passes a Hausman test that compares the OLS estimates and the 3SLS estimates without the restriction. These results indicate that the results cannot reject the null hypothesis that all of the error correlations are zero, consistent with the findings of other studies (Brownstone and Golob (2009), Bhat and Guo (2007)). However, it does not mean that there is no residential self-selection. Rather, it suggests the necessity of including a large set of socio-demographic variables and geographic control variables because these exogenous variables capture the residential self-selection effects.

3.2. Weighted estimation methodology

In section 2, we observed that the probability of having full information on the key endogenous variables is a decreasing function of household vehicle ownership. Since the number of vehicles

⁴The 3SLS is actually identical to the OLS estimation by each equation under the assumption of a diagonal covariance matrix (i.e., error correlations are zero).

in the household is closely related to the endogenous variables in our model, this means that the estimation sample is effectively stratified on the endogenous variables. This non-randomness of missing data may cause biased results and inferences.

There are two approaches to correct for the bias caused by this non-random missing data problem: the structural approach⁵ and the weighting approach. Following Brownstone and Golob (2009), we use the weighted estimation methodology. The weighted estimation is always inefficient, but it does not rely on functional form assumptions that are hard to justify. It also allows easy implementation of allowing for error heteroskedasticity.

The weights are defined to compensate for the different probabilities of having complete energy information. Since the probabilities of having full information for households owning more vehicles are lower, the households with more vehicles must be weighted more. The weights are calculated as the inverse probability of having full information. For example, since 93.5% have full information on the key endogenous variables for households owning one vehicle, the inverse probability (the weight) is 1.07. Likewise, since 29.6% have full information for households owning 5 or more vehicles, the weight is 3.38. The weighted estimator is defined as

$$\arg \min_{A,B} \sum_i w_i [(I - A)y_i - BX_i]' \Omega^{-1} [(I - A)y_i - BX_i], \quad (3)$$

where w_i denotes the weights. The covariance of the weighted estimator above is given by

$$\begin{aligned} V &= \Psi^{-1} \Lambda \Psi^{-1}, \\ \Psi &= -E \left[\frac{\partial^2 w_i L_i(\theta; y_i, X_i)}{\partial \theta \partial \theta'} \right], \\ \Lambda &= E \left[\left(\frac{\partial w_i L_i(\theta; y_i, X_i)}{\partial \theta} \right) \left(\frac{\partial w_i L_i(\theta; y_i, X_i)}{\partial \theta'} \right) \right], \end{aligned} \quad (4)$$

where L_i is the log likelihood function, and θ is the parameter vector.

Usual computer programs provide the weighted estimation solution. However, these programs use Ψ^{-1} to estimate the covariance of the estimator, and this is clearly biased.

⁵ The structural approach proposed by Heckman (1979) specifies a separate binomial probit model of whether the household has complete energy information. Then, the specified equation is added to the original structural system and the equation system is estimated. However, Heckman's method is known to be very sensitive to model specifications because it relies on a joint normality assumption among error terms, which is very strong assumption.

Following Brownstone and Golob (2009), we use the wild bootstrap method (Horowitz (2002)) to generate standard errors for the weighted estimates. This bootstrap works by taking the vector of estimated residuals, denoted by e_i , for each observation and multiplying by $(1-\sqrt{5})/2$ with probability $(1+\sqrt{5})/(2\sqrt{5})$ and by $(1+\sqrt{5})/2$ with probability $1-(1+\sqrt{5})/(2\sqrt{5})$. This implies that across the bootstrap repetitions the residuals will have mean equal to e_i and covariance equal to $e_i e_i'$, which is the same approximation used to derive White heteroskedastic-consistent standard errors. This bootstrap procedure will yield consistent standard errors even if the errors in the model are actually heteroskedastic. We used 200 bootstrap iterations to generate standard errors for the weighted estimates.

As it is mentioned earlier in this section, the structural model is imposing restrictions on the coefficients and the covariance matrix. One drawback of using weighted estimation is that since they are not equivalent to maximum likelihood estimation, standard likelihood ratio tests for over-identifying restrictions cannot be used. Therefore, we implemented a bootstrap test for overidentifying restrictions (including the restrictions on the covariance matrix) by bootstrapping the variance of the difference between the restricted and the unrestricted reduced forms. The reduced form is given by

$$y_i = CX_i + \mu_i, \quad (5)$$

where C are the reduced form coefficients. The over-identifying (or structural) restrictions are given by

$$\begin{aligned} C &= (I - A)^{-1} B \\ Cov(\mu_i) &= (I - A)^{-1'} \Omega (I - A)^{-1}. \end{aligned} \quad (6)$$

Our test statistic then is given by

$$(C_R - C_U)' \Sigma^{-1} (C_R - C_U), \quad (7)$$

where C_R are the restricted reduced form coefficients, C_U are the unrestricted reduced form coefficients, and Σ is the bootstrap variance estimate of $(C_R - C_U)$. If the restrictions are correct, then this statistic follows a Chi-squared distribution with degrees of freedom equal to the number

of restrictions. The final model presented in the next section cannot reject the null hypothesis that the overidentifying restrictions on the empirical specification (i.e., the restrictions on the coefficient matrices A and B and the restriction on the diagonal residual matrix) are correct.

We also implemented a Hausman (1978) test to check whether the assumption of a diagonal covariance matrix is acceptable in cases that all of the over-identifying restrictions (including the restriction that error correlations are zero) did not pass the specification test described above. The Hausman test compares the 3SLS estimates without imposing the restriction of zero error correlations and the estimates from the restriction. Although most model specifications passed the Hausman test⁶, we decided to be conservative and present the model that passes both the bootstrap over-identification test and the Hausman test.

We implemented another Hausman test of the null hypothesis that the weights are exogenous. This test compares weighted estimates and standard maximum likelihood estimates ignoring the weights. When applied to our final model, the test statistic does not reject the null hypothesis that weights are exogenous at any usual confidence level,⁷ and this result is fairly robust to changes in model specification. The coefficients are not much different between the weighted and unweighted estimates. Because unweighted estimation (i.e., standard maximum likelihood estimation) is fully efficient, standard errors for unweighted estimation are reduced. Also, both weighted and unweighted estimates pass the over-identification describes above. We therefore decided to present the unweighted estimation result as the final result.

Finally, we implemented formal tests of whether the results from the national and the California subsample are systematically different. Since we used unweighted estimation, standard tests for maximum likelihood estimation are valid. We used the Hausman test to compare the coefficients between the two cases. The unrestricted model is the model using the data that contains both national and California subsample, which is composed of 6613 observations. The restricted model is the model with the same specification, but using the data containing only 2079 California subsample. This test statistic rejected the null hypothesis of no systematic differences.⁸ As another test, we included interaction terms with a dummy variable

⁶ In the final model presented, P-value (prob.>chi2) is 0.9572 when the intercepts are included in comparison. Otherwise, P-value is 0.9535.

⁷ P-value (prob.>chi2) is 0.5774 when the intercepts are included in comparison. Otherwise, P-value is 0.9798.

⁸ P-value is 0.003.

for being in the California subsample for all variables in the model, and the null hypothesis is that all coefficients for these interaction terms are zero. We dropped the Extensive Rail MSAs dummy so that this dummy does not affect the result since California doesn't have any MSA of this variable. This test statistic also rejected the null that all coefficients are zero, which implies that the result of California is not the same as that of the national sample.

4. Results and interpretation

4.1. Estimation results

The structural equation model was estimated using the unweighted OLS (i.e., 3SLS with the restriction of a diagonal covariance matrix), and heteroskedasticity-robust standard errors are estimated using the wild bootstrap method described in the previous section. The best model uses housing density at the census block level as the measure of urban spatial structure, but the models using other six land use density variables⁹ also produce similar results. The urban/rural dimension variable is used to control for geographic heterogeneity. Note that the model is estimated under the assumption that the structural errors are uncorrelated. The best model using housing density and the urban/rural dimension variable passes the bootstrap test for over-identifying restrictions including the restriction on the residual correlation matrix at any usual level of confidence.

We also used the MSA category variable¹⁰ and the census region variable¹¹ to control for geographic heterogeneity. The estimated impacts of residential density on vehicle usage are not much different from the results of the model using the urban/rural dimension variable, but the influences of these geographic control variables are quantitatively less than the influences of the

⁹ 2001 NHTS provides seven levels of density variables: population and housing density at the census and block level, percentage of renter occupied-housing units at both the block group and tract group level, and jobs per square mile at the tract level.

¹⁰ This variable is categorized as 1=MSA of 1 million or more with rail, 2=MSA of 1 million or more and not in 1, 3=MSA less than 1 million, and 4=Not in MSA.

¹¹ The census region variable is defined by the Census Bureau and it divides states into four groups: northeast, mid-west, south, and west.

urban/rural dimension. In addition, the models using these geographic control variables do not pass the bootstrap test for over-identifying restrictions, unlike the model using the urban/rural dimension. Thus, we present the results of the best model using housing density and the urban/rural dimension variable. The results for the structural model are given in Table 6.

Table 7 gives the restricted reduced form coefficients corresponding to the structural model in Table 6. The restricted reduced form coefficients are calculated from the formula

$C_R = (I - A)^{-1}B$ in equation (6), where A and B are the structural model coefficients. The reduced form coefficients give the total effects of the exogenous variables on the endogenous variable. For example, controlling for all of the other exogenous variables in the model households living in New York, Chicago, Boston, Washington, and Philadelphia drive 2530 fewer miles and consume 134 gallons less fuel each year.

Table 6: Structural model coefficients

Explanatory variables	Endogenous variables		
	Household fuel consumption per year in gallons	Total mileage per year for all household vehicles	Dwelling units per sq. mile in units of 1000 - census block group
Dwelling units per sq. mile in units of 1000 - census block group	-6.1 (-2.18)	-1341 (-7.02)	
Total mileage per year for all household vehicles	0.0441 (43.18)		
Annual household income in units of \$10000	7.8 (6.56)	639 (9.45)	-0.021 (-5.17)
Number of children	8.0 (2.18)	128 (0.42)	-0.045 (-2.50)
Number of workers	-24.4 (-0.37)		0.056 (1.83)
1-worker household	25.1 (0.37)	3950 (6.37)	
2-worker household	67.7 (0.50)	6603 (6.96)	
3-worker household	60.8 (0.28)	11510 (6.41)	
Number of drivers	22.6 (3.20)	7811 (3.37)	-0.067 (-0.97)
1-driver household		-1835 (-0.79)	-0.316 (-3.24)
2-driver household		-2126 (-0.45)	-0.532 (-3.47)
3-driver household		-3240 (-0.43)	-0.716 (-2.94)
Respondent has college degree	-22.2 (-2.90)		
Respondent has postgraduate degree	-52.8 (-5.00)		
Respondent is retired		-3011 (-3.28)	-0.154 (-2.68)
Single-person household		-1997 (-2.23)	-0.098 (-1.33)
Race is Asian	-41.6 (-3.15)	-2305 (-2.56)	0.471 (3.35)
Race is Hispanic		-1153 (-0.87)	0.397 (3.67)
Race is Black			0.118 (1.51)
Race is mixed White or Hispanic			0.195 (1.80)
Extensive Rail MSAs ²⁾	-18.3 (-2.46)	-1724 (-2.74)	0.601 (10.59)
Household resides in Second City ³⁾	-71.2 (-5.30)	-2348 (-2.75)	1.865 (36.98)
Household resides in Suburban	-73.6 (-5.71)	-1913 (-2.13)	1.597 (36.31)
Household resides in Town	-44.1 (-3.69)	-1330 (-1.63)	0.497 (16.50)
Household resides in Urban	-68.8 (-3.78)	-2777 (-2.42)	3.914 (56.44)

Notes: 1) The bootstrapped t-statistics are in parentheses. 2) Extensive Rail MSAs include New York, Chicago, Boston, Washington, and Philadelphia. 3) The urban/rural dimension variable divides urban form into 5 categories: urban, second city, suburban, town, and rural. The classification is based on the contextual density measure (density in the context of its surrounding area)

Table 7: Restricted reduce-form coefficients

Explanatory variables	Endogenous variables		
	Household fuel consumption per year in gallons	Total mileage per year for all household vehicles	Dwelling units per sq. mile in units of 1000 - census block group
Annual household income (\$10000)	37.4 (11.56)	668 (9.75)	-0.021 (-5.17)
Number of children	16.6 (1.22)	188 (0.62)	-0.045 (-2.50)
Number of workers	-28.1 (-0.42)	-76 (-1.71)	0.056 (1.83)
1-worker household	199.4 (2.81)	3950 (6.37)	
2-worker household	359.2 (2.57)	6603 (6.96)	
3-worker household	568.9 (2.43)	11510 (6.41)	
Number of drivers	371.7 (3.66)	7901 (3.43)	-0.067 (-0.97)
1-driver household	-60.4 (-0.59)	-1411 (-0.60)	-0.316 (-3.24)
2-driver household	-59.1 (-0.29)	-1412 (-0.30)	-0.532 (-3.47)
3-driver household	-96.3 (-0.29)	-2280 (-0.30)	-0.716 (-2.94)
Respondent has college degree	-22.2 (-2.90)		
Respondent has postgraduate degree	-52.8 (-5.00)		
Respondent is retired	-122.8 (-2.95)	-2804 (-3.05)	-0.154 (-2.68)
Single-person household	-81.8 (-2.01)	-1866 (-2.07)	-0.098 (-1.33)
Race is Asian	-174.0 (-4.22)	-2936 (-3.11)	0.471 (3.35)
Race is Hispanic	-76.8 (-1.32)	-1686 (-1.28)	0.397 (3.67)
Race is Black	-7.7 (-1.43)	-158 (-1.44)	0.118 (1.51)
Race is mixed White or Hispanic	-12.8 (-1.75)	-262 (-1.74)	0.195 (1.80)
Extensive Rail MSAs ³⁾	-133.6 (-4.47)	-2530 (-4.01)	0.601 (10.59)
Household resides in Second City ⁴⁾	-296.5 (-8.10)	-4849 (-6.25)	1.865 (36.98)
Household resides in Suburban	-262.2 (-6.58)	-4054 (-4.76)	1.597 (36.31)
Household resides in Town	-135.2 (-3.29)	-1996 (-2.42)	0.497 (16.50)
Household resides in Urban	-446.7 (-10.73)	-8025 (-8.91)	3.914 (56.44)

Notes: 1) The bootstrapped t-statistics are in parentheses. 2) The restricted reduced form coefficients are calculated from the equation (6), where A and B are the structural model coefficients. 3) Extensive Rail MSAs include New York, Chicago, Boston, Washington, and Philadelphia. 4) The urban/rural dimension variable divides urban form into 5 categories: urban, second city, suburban, town, and rural. The classification is based on the contextual density measure (density in the context of its surrounding area).

4.2. Interpretation of the results

4.2.1. Relationship among three endogenous variables

The model implies that if two households are identical in all aspects measured by the socio-demographics variables and residing in the same category of urban/rural dimension, but one household is located in a residential area that is 1000 housing units per square mile denser (roughly 50% of the sample average), the household in the denser area will drive 1341 (6.9%) miles per year less than the household in the less dense area. This yields an arc elasticity of $-.08$, which is somewhat higher than the overall $-.04$ elasticity in the meta-analysis performed by Ewing and Cervero (2010) and within the $-.05$ -- $-.12$ range reported in Transportation Research Board (2009). This is the net effect of vehicle ownership level and trip patterns. Also note that this is the partial effect of residential density within a category defined by the urban/rural dimension variable.

We can compare this result with that in California subsample analysis, which uses the same specification including the urban/rural dimension and imposing the same restrictions on the coefficient matrices as in this study. The California subsample is the data containing 2,079 observations, which was used in Brownstone and Golob (2009). In national sample, the coefficient of residential density on household annual mileage is greater than the California subsample result of -780 (3.6%) in response to a 1000 housing units per square mile density change (roughly 40% of the sample average). The Chow test that compares the coefficient from the model using 4534 national data (excluding California) and the coefficient from the model using 2079 California subsample cannot reject the null of no systematic difference at 5% confidence level (P-value is 0.0657). Note however that since the density in California is higher, the similar arc elasticity of VMT with respect to density for the California subsample is $-.09$ which is slightly higher than we found for our national sample.

The household in a 1000 housing units per square mile denser area will consume 65 (7.0%) fewer gallons of fuel, and this effect of residential density on fuel usage is decomposed into two paths of influence. The mileage difference of 1341 miles leads to a difference of 59 gallons, which is calculated by multiplying 0.0441 gallons per mile (from the coefficient of mileage on fuel consumption in Table 6) by 1341 mileage difference. However, there is an additional direct effect of density on fuel consumption of 6 gallons per 1,000 housing units per square mile. This

is due to the relationship between residential density and fleet fuel economy, a result of vehicle type choice; people in denser area choose more fuel efficient vehicles.

The total impact of density on fuel consumption in national data (65 gallons) is slightly greater than the California subsample result of 47 (4.5%) fewer gallons of fuel in denser area. However, the direct effect of density on fuel consumption (6 gallons) is quantitatively smaller than the California subsample result of 14 gallons and it accounts for only 9% (6/65) of the total impact of residential density on fuel consumption. It is smaller than the California subsample result of 29% (14/47). However, the Chow test does not reject the null hypothesis of no systematic difference between the two coefficients (the direct impact of density on fuel usage) at any usual confidence level (P-value is 0.1043).

4.2.2. Urban/rural dimension and rail transit dummies

The results above suggest that residential density has only a modest effect on mileage and fuel consumption (a 50% increase in housing density gives rise to only about 7% decrease in mileage and fuel consumption). However, this is not necessarily the case if both residential density and the urban/rural dimension change simultaneously. This is the total impact of urban/rural dimension on mileage and fuel consumption, and it is given in Table 7 as the reduced-form coefficients.

Since the urban/rural dimension is highly correlated with residential density, the total effect of the urban/rural dimension on mileage is due to both a direct effect and an effect channeled through residential density. From Table 6, because an urban area is 3914 (roughly 200% of the sample average) housing units per square mile denser than rural area, there are 5249 (27% of the sample average) fewer annual miles traveled per household. There is an additional direct effect of 2777 (14%) fewer miles for a household living in urban area compared to a household in a rural area. In total, the household residing in urban area drives 8025 (41%) miles less than the household residing in a rural area, and this is presented as the coefficient in Table 7.¹² The household living in a suburban area drives 3971 (21%) miles more than a household in an urban

¹² The indirect effect (5249 miles) is calculated by multiplying the density difference (3.914) and the effect of density on mileage (1341) (i.e., $3.914 \times 1341 = 5249$). The total effect is just the sum of the indirect and the direct effect (i.e., $5249 + 2777 = 8025$).

area, and this total effect is decomposed into an indirect effect of 3107(16%) miles (calculated by multiplying 2.317 density difference by 1341 mileage per density difference) and a direct effect of 864 (5%) miles.

Note that the urban/rural dimension is defined from the contextual density measure (density in the context of surrounding area). For example, even when a certain region is dense, if its surrounding area is not dense, then the contextual density measure may indicate that the region is a less dense area. The coefficients of the urban/rural dimension in the mileage equation, which is the direct effect of the urban/rural dimension, capture the influences of moving a household from rural to other dimensions holding residential density fixed. The direct effect captures the change in density of its surrounding area holding density of the region fixed. We can also interpret it as if the development around the region has occurred. From Table 6, we can see that the effects of moving a household from rural to other areas (holding density of the region fixed) are statistically significant except for the change from the rural area to the town area.¹³

The effect of the urban/rural dimension variable on fuel consumption is the sum of the direct effect, the effect channeled through mileage, and the effect channeled through residential density. The effect channeled through mileage is obtained by multiplying the total effect of the urban/rural dimension on mileage by the direct effect of mileage on fuel consumption. The effect channeled through residential density is from multiplying the effect of the urban/rural dimension on density by the direct effect of density on fuel consumption. From Table 7, the household living in an urban area consumes 447 (48% of the sample average) fewer gallons of fuel compared to a household living in a rural area. This total effect is decomposed of 69 gallons of direct effect, 354 gallons of effect channeled through mileage, and 24 gallons of effect channeled through residential density.

The effect of rail transit on mileage and fuel consumption can also be decomposed into direct and indirect effects. Surprisingly, the effect of rail transit on mileage is not statistically significant. The indirect effect channeled through residential density does exist, but the magnitude is not large and statistically insignificant. Therefore, the specifications that include this rail transit dummy are not presented. Among the MSAs which have rail transit several large

¹³ But, the effect of moving a household from a non-rural to another non-rural area (holding density fixed) is not significant.

old cities, including New York, Chicago, Boston, Washington, and Philadelphia, were separated into a new dummy variable and labeled as “Extensive Rail MSAs.” This Extensive Rail MSAs dummy has significant influences on both mileage and fuel consumption, unlike the rail transit dummy. From Table 7, a household living in one of these large old cities is predicted to drive 2530 (13% of the sample average) miles less and consume 134 (14%) fewer gallons of fuel than the other household.

Simulation using the urban/rural dimension

The total effect of the urban/rural dimension can be interpreted as the impact of a program that alters urban form from changing the initial category of the urban/rural dimension into another category. Table 8 reports the total effect of every category change of the urban/rural dimension. Note that the discrete category changes are accompanied by the associated density changes, which are the coefficients in the density equation in Table 6. The same calculation methods are applied to all category changes as described above, but the percentage changes are calculated based on the mean of each category. Since standard errors are presented, we can see the statistical significances of the effect of moving households.

When we move a household from a rural to an urban area, the effect is to reduce annual household mileage by 34%. The household living in an urban area is predicted to drive 18% fewer miles than the household living in a suburban area. When we move the household from a rural to a suburban area, the household will drive less by 17%. All these effects of category changes are statistically significant except for the change from suburban to second city category. All these percentage changes are higher when the households move from denser areas to less dense areas because of the low base effects when percentages are calculated. Percentage change in fuel consumption is slightly higher than percentage change in mileage in most dimensions reflecting the tendency toward more fuel efficient vehicle choices in denser area.

The effects of this program is very similar to the results of Bento et al. (2005), where the authors implemented a similar simulation that moves a household from a city which has the same characteristics as Georgia to a city which has the same characteristics as Boston. In Bento et al., the simultaneous changes in all characteristics of the cities influenced the household travel behavior significantly, which is consistent with our results.

The effects of this program using California data are much smaller than the results using national data. The effects of most category changes are only about half of those of national data. In California, the effects of category changes are frequently statistically insignificant, unlike the results from the national data. The effects of this program in California data is given in Table 9.

Note that while we compare the magnitudes of the effects of density and urban/rural dummies between California and other regions, we cannot say much about the reasons for these differences. The different effects of density may be due to the differences in urban spatial structure that are not captured by the density variable or by the urban/rural dimension dummies. This conjecture is consistent with Bento et al. (2005), which shows how travel behavior interacts with various attributes of urban forms that vary by regions. But, there could be other factors such as fuel prices and availability of mass transit that also differ by regions.

Table 8: The influences of a program that moves a household across the urban/rural dimension (4992 national sample)

A. Changes in household annual mileage										
From	Urban		Second City		Suburban		Town		Rural	
To	Mileage change	% Change	Mileage change	% Change	Mileage change	% Change	Mileage change	% Change	Mileage change	% Change
Urban			-3176 (881)	-19.71%	-3971 (847)	-17.92%	-6029 (857)	-26.11%	-8025 (900)	-33.89%
Second city	3176 (881)	31.45%			-795 (819)	-3.59%	-2852 (763)	-12.35%	-4849 (775)	-20.48%
Suburban	3971 (847)	39.32%	795 (819)	4.93%			-2058 (744)	-8.91%	-4054 (852)	-17.12%
Town	6029 (857)	59.70%	2852 (763)	17.70%	2058 (744)	9.29%			-1996 (824)	-8.43%
Rural	8025 (900)	79.46%	4849 (775)	30.08%	4054 (852)	18.30%	1996 (824)	8.65%		

B. Changes in household annual fuel consumption										
From	Urban		Second City		Suburban		Town		Rural	
To	Fuel usage change	% Change	Fuel usage change	% Change	Fuel usage change	% Change	Fuel usage change	% Change	Fuel usage change	% Change
Urban			-150 (37.5)	-19.59%	-184 (38.3)	-17.49%	-312 (39.0)	-27.62%	-447 (41.7)	-37.40%
Second city	150 (37.6)	31.82%			-34 (35.9)	-3.25%	-161 (35.1)	-14.30%	-296 (36.6)	-24.82%
Suburban	184 (38.3)	39.08%	34 (35.9)	4.46%			-127 (34.3)	-11.26%	-262 (39.8)	-21.95%
Town	312 (39.0)	65.99%	161 (35.1)	21.03%	127 (34.3)	12.05%			-135 (41.0)	-11.32%
Rural	447 (41.7)	94.62%	296 (36.6)	38.65%	262 (39.8)	24.87%	135 (41.0)	11.99%		

Notes: 1) The calculated changes are the sum of direct and indirect effects of the urban/rural dimension. The direct effects are the coefficients of the urban/rural dimension in the mileage equation in Table 6. The indirect effects are calculated by multiplying the direct effect of the urban/rural dimension and the associated density changes (from the first row of Table 2), which are the coefficients of the urban/rural dimension in the density equation in Table 6. 2) The effect of the urban/rural dimension variable on fuel consumption is the sum of the direct effect, the effect channeled through mileage, and the effect channeled through residential density. The effect channeled through mileage is obtained by multiplying the total effect of the urban/rural dimension on mileage by the direct effect of mileage on fuel consumption. The effect channeled through residential density is obtained by multiplying the effect of the urban/rural dimension on density by the direct effect of density on fuel consumption. 3) Percentage changes are achieved from dividing the calculated changes by the mean of each category of the urban/rural dimension. 4) Standard errors are reported in parenthesis.

Table 9: The influences of a program that moves a household across the urban/rural dimension (2079 California subsample)

A. Changes in household annual mileage										
From	Urban		Second City		Suburban		Town		Rural	
To	Mileage change	% Change	Mileage change	% Change	Mileage change	% Change	Mileage change	% Change	Mileage change	% Change
Urban			-2483 (995)	-11.76%	-2036 (917)	-8.46%	-3541 (1376)	-14.09%	-3414 (1863)	-14.78%
Second city	2483 (995)	13.18%			447 (1063)	1.86%	-1059 (1413)	-4.21%	-932 (1949)	-4.03%
Suburban	2036 (917)	10.81%	-447 (1063)	-2.12%			-1505 (1390)	-5.99%	-1378 (1866)	-5.97%
Town	3541 (1376)	18.80%	1059 (1413)	5.01%	1505 (1390)	6.26%			127 (2199)	0.55%
Rural	3414 (1863)	18.12%	932 (1949)	4.41%	1378 (1866)	5.73%	-127 (2199)	-0.51%		

B. Changes in household annual fuel consumption										
From	Urban		Second City		Suburban		Town		Rural	
To	Fuel usage change	% Change	Fuel usage change	% Change	Fuel usage change	% Change	Fuel usage change	% Change	Fuel usage change	% Change
Urban			-152 (48.2)	-14.93%	-101 (38.8)	-8.89%	-218 (63.3)	-17.66%	-224 (86.6)	-19.59%
Second city	152 (48.2)	17.38%			51 (50.3)	4.47%	-66 (69.0)	-5.36%	-72 (93.2)	-6.32%
Suburban	101 (38.8)	11.57%	-51 (50.3)	-4.99%			-117 (63.1)	-9.47%	-123 (84.9)	-10.76%
Town	218 (63.3)	24.95%	66 (69.0)	6.50%	117 (63.1)	10.29%			-6 (100.0)	-0.55%
Rural	224 (86.6)	25.67%	72 (93.2)	7.12%	123 (84.9)	10.84%	6 (100.0)	0.51%		

Notes: 1) The calculated changes are the sum of direct and indirect effects of the urban/rural dimension. The direct effects are the coefficients of the urban/rural dimension in the mileage equation in Table 6. The indirect effects are calculated by multiplying the direct effect of the urban/rural dimension and the associated density changes, which are the coefficients of the urban/rural dimension in the density equation in Table 6. 2) The effect of the urban/rural dimension variable on fuel consumption is the sum of the direct effect, the effect channeled through mileage, and the effect channeled through residential density. The effect channeled through mileage is obtained by multiplying the total effect of the urban/rural dimension on mileage by the direct effect of mileage on fuel consumption. The effect channeled through residential density is obtained by multiplying the effect of the urban/rural dimension on density by the direct effect of density on fuel consumption. 3) Percentage changes are achieved from dividing the calculated changes by the mean of each category of the urban/rural dimension. 4) Standard errors are reported in parenthesis.

4.2.3. Socio-demographic variables

The number of drivers has a strong influence on household travel behavior. The total effects on each of the three endogenous variables are nonlinear, as captured by up to four variables: a continuous “number of drivers” variable, and dummy variables for one-driver, two-driver and three-or-more-driver households. However, the variables for one-driver, two-driver and three-or-more-driver households are all insignificant, indicating that the influence of drivers per household on annual vehicle usage exhibits a fairly linear pattern. The effect of the number of drivers is smaller than that of the number of workers. Unlike the total effect of the number of drivers, the influence of the number of workers exhibits a non-linear pattern.

The model predicts that fuel usage increases with income because a higher income translates into a choice of lower density residential location, a greater total driving distance, and a lower overall fuel economy of the household fleet. All other things equal, if one household earns \$10000 higher income annually, the household with higher income will consume 37 (4% of the sample average) more gallons of fuel per year (see Table 7).

The influence of the number of children on annual mileage and fuel consumption is statistically insignificant, unlike that in Brownstone and Golob (2009). Households headed by a respondent with a college degree tend to have a vehicle fleet with lower fuel economy than their less educated counterparts. This effect is accentuated if the household is headed by a respondent with a postgraduate degree. Retired households tend to drive less than other counterpart households. The total effect of a single-person household on household annual mileage and fuel consumption is statistically significant, implying that single-person households tend to drive less than non-single-person households. The effects of races are frequently insignificant except that Asian households tend to utilize vehicles less than White or other races.

5. Robustness checks

In addition to the internal checks for the validity of the overidentifying restrictions, we also carried out a number of out of sample tests using observations from the NHTS data not included in our estimation sample. We drew 20 samples of size 4992 from the 16358 observations not in the original estimation sample, and for each of these samples we carried out Hausman tests for

the null hypothesis that the coefficients were equal across the two samples (similar to methods we used to test the California subsample). This null hypothesis was rejected for 13 of the samples, but it was not rejected for the other 7. Of course we could go back and change the model specification so that there were no rejections, but that would just be equivalent to doing standard model specification searches on the entire sample. Since it is very rare for applied econometrics researchers to use a holdout sample to check for overfitting, it is hard to say whether our out of sample test results indicate serious problems.

The other check we carried out was to fit our model specification to data from the newly-released 2009 NHTS data. This work is described in Hong (2011), and he finds quantitatively similar conclusion to those given in this paper. The 2009 NHTS only collected a single odometer reading from each household vehicle, so the resulting miles traveled variable is less reliable than in the 2001 data. The 2009 NHTS data also do not clearly identify the national sample, so it is difficult to judge the reliability of the weights given in these data. For these reasons we prefer to rely on the 2001 NHTS data for our results, although it is comforting that our specification yields similar results on the 2009 data.

6. Conclusion and further research

This paper measures the impact of residential density on vehicle usage and fuel consumption following the same methodology as Brownstone and Golob (2009), but using national level data from the 2001 NHTS. To account for residential self-selection effects, a simultaneous equation model is specified and OLS (3SLS with the restriction of a diagonal covariance matrix) is used for estimation. A weighting estimation method is used to correct for the bias caused by non-random missing data, but unweighted estimates are presented as the final result since we cannot reject the null hypothesis that the missing data process doesn't bias the results. After controlling for various socio-demographic and geographic locations, the model cannot reject the null hypothesis that there are no significant self-selection effect, which is consistent with Bhat and Guo (2007) and Brownstone and Golob (2009).

We find that residential density has a statistically significant but economically modest influence on household travel behavior, which is similar to other previous studies (Bento et al.

(2005), Brownstone and Golob (2009)). However, this is not necessarily the case if the contextual density measure (also defined by the urban/rural dimension) as well as residential density changes simultaneously. A program that alters both the urban/rural dimension and residential density affects household vehicle usage significantly.

Compared to the California subsample analysis, which uses the same specification as in this study, the impact of residential density on mileage and fuel consumption is slightly higher. The direct impact of residential density on fuel consumption is statistically significant, but smaller than the California subsample result. We also find that the urban/rural dimension does not have an explanatory power in explaining household vehicle usage in California data, which is different from the result using national data.

This research can be usefully extended in several directions. First, there might be some ways to overcome the limitations from using residential density as the measure of urban sprawl. Considering that the phenomenon of urban sprawl has a dynamic aspect, new measures can be devised using the information on urban development of the U.S since World War II. Although the urban/rural dimension was devised to consider the density surrounding the area, the definition of each category is ad-hoc. To overcome this problem, adjunct geographic location information can be merged into the NHTS dataset to provide more information about the households' neighborhood characteristics.

Second, we can test the model with different selection nodes other than household mileage traveled and fuel consumption. This paper assumes that household mileage is just the summation of miles drive by all household vehicles. However, the decision of mileage traveled can be a different choice node from purchasing and owning cars. In that case, the number of cars can be also regarded as another endogenous variable. The daily travel diary data set, which is included in the NHTS data, can be also used for further analysis of travel behavior. If it is possible to obtain information on accessibility to public transportation for those households in major metropolitan areas, a model can be developed to jointly determine public transit accessibility along with residential density and transportation energy use.

REFERENCES

- Badoe, D. and E. Miller, 2000. Transportation-land use interaction: empirical findings in North America, and their implications for modeling. *Transportation Research D*, 5, pp. 235-263.
- Bento, A.M., M.L. Cropper, A.M. Mobarak and K. Vinha, 2005. The impact of urban spatial structure on travel demand in the United States. *Review of Economics and Statistics*, 87, pp. 466-478.
- Bhat, C.R. and J.Y. Guo, 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research B*, 41, pp. 506-526.
- Boarnet, M. and R. Crane, 2001. The influence of land use on travel behavior: specification and estimation strategies. *Transportation Research Part A*, 35, pp. 823-845.
- Brownstone, D and T. F. Golob, 2009. The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, 65, pp. 91-98.
- Ewing, R. and R. Cervero, 2010. Travel and the built environment: a meta-analysis. *Journal of the American Planning Association*, 76: 3, pp. 265-294.
- Hausman, J.A., 1978. Specification tests in econometrics. *Econometrica*, 46, pp. 1251-1271.
- Heckman, J.J., 1979. Sample selection bias as a specification error. *Econometrica*, 47, pp. 153-161.
- Hong, H., 2011. Measuring the effect of land use density on vehicle and fuel usage. Manuscript. Department of Economics, UC Irvine, March.
- Horowitz, J. L., 2002. The bootstrap, in E. E. Leamer and J.J. Heckman (Eds.), *Handbook of Econometrics*, Vol. 5, Elsevier, Amsterdam, Chapter 52.
- Miller, D. R. and K. Hodge, 1994. A population density approach to incorporating an urban-rural dimension into small area lifestyle clusters. Paper presented at the Annual Meeting of the Population Association of America, Miami, Florida
- Newman, P. and J. Kenworthy, 1999. Costs of automobile dependence: global survey of cities. *Transportation Research Record*, 1670, pp. 17-26.

Salon, D. 2009. Neighborhoods, cars, and commuting in New York City: A discrete choice approach. *Transportation Research Part A*,43, pp. 180-196.

Transportation Research Board Committee for the Study on the Relationships Among Development Patterns, Vehicle Miles Traveled, and Energy Consumption, 2009. *Driving and the Built Environment: The Effects of Compact Development on Motorized Travel, Energy Use, and CO2 Emissions*. Special Report 298, National Research Council, Washington, D.C. Accessed at <http://onlinepubs.trb.org/Onlinepubs/sr/sr298.pdf> .