Journal of Urban Economics 65 (2009) 91–98

Contents lists available at ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

The impact of residential density on vehicle usage and energy consumption ☆

David Brownstone [a,*], Thomas F. Golob [b]

[a] Department of Economics, 3151 Social Science Plaza, University of California Irvine, Irvine, CA 92697-5100, USA
[b] Institute of Transportation Studies, University of California Irvine, Irvine, CA 92697-3600, USA

ARTICLE INFO

Article history:
Received 26 September 2008
Available online 7 October 2008

JEL classification:
C30
D12
L92
Q58
R14
R41

Keywords:
Residential density
Vehicle use
Vehicle fuel consumption
Simultaneous equations
Self-selection

ABSTRACT

We specify and estimate a joint model of residential density, vehicle use, and fuel consumption that accounts for both self selection effects and missing data that are related to the endogenous variables. Our model is estimated on the California subsample of the 2001 U.S. National Household Travel Survey (NHTS). Comparing two California households that are similar in all respects except residential density, a lower density of 1000 housing units per square mile (roughly 40% of the weighted sample average) implies an increase of 1200 miles driven per year (4.8%) and 65 more gallons of fuel used per household (5.5%). This total effect of residential density on fuel usage is decomposed into two paths of influence. Increased mileage leads to a difference of 45 gallons, but there is an additional direct effect of density through lower fleet fuel economy of 20 gallons per year, a result of vehicle type choice.

1. Introduction and background

This paper measures the relationship between residential density, household vehicle use, and household vehicle fuel use. It contributes to a large literature on the impact and/or desirability of low-density suburban development, frequently called "urban sprawl," that has dominated development in the U.S. since World War II. Increased vehicle usage associated with suburbanization or urban sprawl has been linked to increasing global warming, emissions, and other problems (see Pickrell and Schimek, 1999, and Kahn, 2000). Urban sprawl is not simply low density, but also involves scattered development, commercial strip development, or large expanses of single-use development. Nevertheless, density is highly correlated with almost all measures of urban sprawl and is the measure used most frequently in this literature (see reviews by Ewing and Cervero, 2001, and Badoe and Miller, 2000). When used alone as an indicator of sprawl, density should therefore be interpreted as a proxy for access to employment, shopping, and other travel destinations. The main reason density is used so frequently is that it is one of the few indicators of sprawl that is consistently measured across space and time, and it is readily available in most relevant data sets.

Aggregate studies examining the bivariate relationship between vehicle miles traveled and density find a large significant inverse effect (see Newman and Kenworthy, 1999). These studies are flawed because they do not account for the possibility of residential self-selection, which is the tendency for those households that prefer non-private vehicle travel to locate in dense areas with more transit and shorter trip distances. Many studies use disaggregate household data to attempt to control for observable differences between households living in low and high density areas. One of the best of these is Bento et al. (2005), which used the 1990 National Personal Transportation Study to build disaggregate models of number of vehicles per household and vehicle miles traveled (VMT) per vehicle. They supplemented the density measures in the data with road density, rail and bus transit supply, population centrality, city shape, jobs-housing balance, population density, land area, and climate. Bento et al. (2005) found that the magnitudes of the impact of any of their built environment measures were frequently statistically insignificant and small in magnitude.

Although disaggregate studies that include a rich set of socioeconomic control variables (e.g. Bento et al., 2005) are less subject

☆ This is a revised version of Working Paper UCI-ITS-WP-05-1, Institute of Transportation Studies, and University of California Irvine. The authors gratefully acknowledge financial support from the University of California Energy Institute and the University of California Transportation Center. Kenneth Small, two anonymous referees, and especially Jan Brueckner provided many useful comments on earlier drafts, but the authors bear sole responsibility for any remaining errors.
* Corresponding author. Fax: +1 949 824 2182.
E-mail address: dbrownst@uci.edu (D. Brownstone).

0094-1190/$ – see front matter
doi:10.1016/j.jue.2008.09.002

to residential self-selection bias, it is still possible that residents of high density areas differ in some unobservable characteristics that influence their travel behavior. The only way to deal with this possibility is to build joint models of residential (or density) choice and travel behavior. One of the first studies to do this was Boarnet and Sarmiento (1998). They used the percentage of buildings built before 1945, the percentage of buildings built between 1945 and 1985, the percentage of residents more than 65 years old, and the percentage of foreign residents as instrumental variables for residential density, and they found no stable link between density and VMT.

Bhat and Guo (2007) use San Francisco Bay Area data to build an ambitious joint model of residential location and number of household vehicles. Their model allows for 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. Bhat and Guo find statistically significant but quantitatively small impacts of built environment measures (street block density, transit availability, and transit access time) on vehicle ownership.

We also directly model the joint choice of density and VMT to control for potential selectivity, and we also include a rich set of socioeconomic variables using the California subsample from the 2001 National Household Travel Study described in the next section of this paper. We chose to work with California because it has as much variation in the key variables as the U.S., but is relatively homogeneous in climate, fuel, and vehicle prices. Unlike previous studies we also explicitly model vehicle fuel consumption to account for the possibility that residents of high density neighborhoods choose smaller, more fuel efficient vehicles. This might be due to the relative difficulty of maneuvering and parking large vehicles in dense neighborhoods. Fang (2008) uses the same data to show that residents of dense neighborhoods choose fewer trucks and more small cars.

Unlike Bhat and Guo (2007) and Fang (2008) we do not explicitly model the number of vehicles or their type. This greatly simplifies the econometrics and allows us to easily deal with problems caused by non-random data selection as described in the next section. The third section describes our simultaneous equations model in which residential density, vehicle usage (VMT), and fuel consumption are joint endogenous variables. The fourth section describes the results, which are similar to previous studies in finding a statistically significant but quantitatively small impact of residential density. Even though our model allows for joint causality between the endogenous variables, our preferred model has density causing VMT (as in Bento et al., 2005) and fuel usage. The final section concludes and argues that the impacts of increased residential density are too small to make increasing density a relevant policy tool for trying to reduce VMT or greenhouse gas emissions from residential vehicles.

2. Data

2.1. 2001 National Household Travel Survey (NHTS)

The NHTS is a household-based travel survey conducted every five years by the U.S. Department of Transportation. There are 2583 California (CA) households in the 2001 NHTS sample, representing 9.9% of the total base sample of 26,038. The survey was conducted over a period of fourteen months ending in May 2002. Daily travel was collected using one-day trip diaries for all household members, and data on non-commuting trips of at least 50 miles to the furthest destination was collected for a four-week period. Household vehicles were defined as all vehicles generally available to household members, including motorcycles, mopeds, and recreational vehicles. Odometer readings were obtained at two dates, generally a few months apart, in order to provide accurate data on annual vehicle miles of travel. The 2001 NHTS is described in detail in exhibits, reports, and codebooks maintained on the NHTS website (ORNL, 2004).

2.2. Vehicle ownership and fuel usage

This study focuses on the energy used by all vehicles owned or leased by California households, including vehicles otherwise available to households for the general use of household members. The weighted frequencies from the NHTS show that 7.5% of California households have no vehicles, 33.8% have one vehicle, 35.0% have two, 15.2% have three, 5.4% have four, and 3.0% have five or more vehicles. As is usual in surveys of this type, households with the fewest numbers of vehicles are under-represented in the sample.

The procedures used to estimate annual fuel usage for each vehicle in the survey are reported in Schipper and Pinckney (2004). Reported and imputed odometer readings, together with fuel economy test results for each vehicle make, model and vintage, are adjusted for on-road shortfalls of vehicle dynamometer test results, seasonal variations, and relationships between total mileage and average trip lengths. The resulting annual fuel usage and annual miles traveled variables are much more accurate than those available in previous versions of the NHTS. Since annual mileage and exact vehicle make, model and vintage are needed to compute fuel usage, 2079 (80.5%) of all California NHTS households have full information on transportation fuel usage.

Since each household vehicle must be accounted for in order for full energy consumption information to be computed, the proportion of households with full information is a decreasing function of vehicle ownership level. It is difficult to collect odometer readings for households with many vehicles since the survey respondent may not have ready access to all of these vehicles when the survey firm calls to collect the data. Full energy information is available for the vast majority of 1- and 2-vehicle households (91 and 86% respectively), but less than half of all households with four or more vehicles have available energy consumption information. Since the number of vehicles is endogenous in our models, this means that the sample of households with complete energy information is not a random sample. We describe the econometric techniques we use to produce consistent estimates later in this paper.

2.3. Land use densities

The 2001 NHTS provides several measures of land use related to household location. Population per square mile and housing units per square mile are provided at the block group and tract level. Percentage of renter-occupied housing units is provided at both the block group and tract level, and jobs per square mile are provided the tract level. As expected, these seven land use variables are all highly correlated. The typical correlation between any two is above 0.7.

2.4. Vehicle usage and land use

As expected, there is a significant negative relationship between fuel usage and land use density. Each of the seven land use variables was tested, and the strongest relationships were found for dwelling units per square mile at the Census block group level. Consequently, we show only the results for the housing density variable, but the other six land use variables exhibit similar patterns. For (urban) densities greater than 50 housing units per square mile, both total annual mileage on all household vehicles and total fuel usage generally decline with increasing housing density, as shown in Table 1. The differences in means for both series

Table 1
Vehicle characteristics by residential density (weighted averages across estimation sample, 2079 observations).

Housing units per square mile in Census block group	<50	50–250	250–1K	1K–3K	3K–5K	>5K
Percentage households residing in density group	5	7	14	40	18	16
Annual fuel consumption in gallons	1200	1650	1230	1330	1030	690
Total annual mileage	27,900	33,200	26,600	29,800	23,900	16,900
Vehicles per household	1.9	2.2	1.9	1.9	1.7	1.4
Percentage households with at least one truck	85	78	73	68	61	41
Average number of drivers	1.8	2	1.8	1.8	1.7	1.45
Household income ($1K)	60	85	74	70	59	61

are statistically significant, and linear relationships cannot be rejected at the $p < 0.01$ level for either series. The slope of the curve is greater for fuel consumption, indicating that there is a positive relationship between effective vehicle fuel economy and urban density. Indeed, effective fuel economy, measured by the ratio of total mileage to total fuel consumption, ranges from a low of 19.7 miles per gallon for households located in areas with densities less than 50 housing units per square mile, to a high of 22.4 miles per gallon for households in areas with greater than 5000 housing units per square mile.

These relationships are caused in large part by differences in household vehicle ownership levels. As shown in Table 1, vehicles per household ranges from a high of 2.2 vehicles per household for households located in areas of 50–250 dwellings per square mile, to a low of 1.4 vehicles per household for those located in the highest density areas. The differences in fuel economy can be attributed to vehicle type choice differences involving size and power of cars and to the greater number of pickup trucks, vans, and SUVs in lower density areas. As shown in Table 1, the likelihood of owning one of these three types of trucks increases with decreasing density, and there is no reversal of the trend at the lowest density as there is for mileage and fuel usage. Since 1990 the fleet average fuel economy for light trucks was no less than 20.2 miles per gallon while the fleet average fuel economy for cars was no less than 27.5 miles per gallon (see NHTSA, 2008).

Of course, different types of households choose to live in areas of different residential density. Quite a few socioeconomic and demographic variables were found to describe choice of residential density in the model presented in the next section. Two of the variables that stand out are the number of household drivers and average household income. The last two rows of Table 1 show that households living in more dense neighborhoods have fewer drivers and lower income.

It is apparent that different types of households choose to live in areas defined by different residential densities. These households have different patterns of activity participation and travel, and choose to own or lease or otherwise have available different numbers and types of vehicles. To account for such selectivity effects of land use on vehicle fuel consumption, we specify and estimate a structural equation model that contains both density of land use and vehicle usage as endogenous variables.

3. Structural model of density of land use and fuel usage

3.1. Model specification

We estimate the effects of land use variables on fuel usage by specifying a simultaneous equation model with three endogenous variables and many exogenous variables. The three endogenous variables are: total annual miles driven by all household vehicles (M), total annual household fuel usage measured in gallons of gasoline equivalents per year (F), and housing units per square mile in Census block group described previously (D). Our preferred model is given by the equation system below, where A and B are coefficient matrices, X_i is a vector of exogenous household attributes, and ε is a vector of residuals with an unrestricted correlation structure. We need to put restrictions on the coefficient matrices to identify the system. We have chosen to identify our system primarily by restricting the A matrix to the recursive system shown below. We also impose enough restrictions on the B matrices to identify the system (see Table 3), but these restrictions are based on removing insignificant variables. The excluded variables from the B matrix are therefore weak instruments, and estimates are essentially unchanged when all exogenous variables are included in each equation.

In the context of our model, residential self-selection implies positive correlations between the structural errors (ε). We cannot reject the null hypothesis that all of the error correlations are zero, and this is consistent with other studies (e.g. Bhat and Guo, 2007) that condition on a rich set of socioeconomic variables. Note that this finding of no significant error correlations does not mean that there are not self-selection effects, but it does imply that the included socioeconomic variables capture these effects.

$$M_i = A_{1,3}D_i + B_1X_i + \varepsilon_{1,i},$$
$$F_i = A_{2,1}M_i + A_{2,3}D_i + B_2X_i + \varepsilon_{2,i},$$
$$D_i = B_3X_i + \varepsilon_{3,i}.$$

Our recursive model starts by assuming that the choice of residential density is only a function of exogenous household characteristics. This is equivalent to assuming that households first choose their residential location and then choose their vehicle holdings and driving patterns conditional on this choice. Density, which is a proxy for access to employment and other destinations, affects total miles driven, but after controlling for exogenous sociodemographic factors fuel usage is assumed to not directly affect total miles driven. Density also affects fuel usage since households in denser neighborhoods choose more fuel efficient vehicles. Density is postulated to affect fuel usage by both decreasing total miles driven and increasing vehicle fuel economy.

The exogenous variables in the structural equations model are designed to capture all sociodemographic characteristics related to choice of density and vehicle use (see the definitions in Table 2). In many cases we include indicator variables of common household types along with counts of children, workers, and drivers. This combination of count and indicator variables allows for flexible nonlinear impacts of key sociodemographic characteristics. A continuous variable was constructed for income by using the midpoints of the 10 categories used in the survey instrument, with $170,000 assumed for the top category, and $35,000 assigned for missing incomes. All of the overidentifying restrictions in our preferred model passed the specification tests described below. In particular, we could find no economically or statistically significant "backward" links from fuel usage to land use density. Note however that removing any of the exogenous characteristics from the model leads to rejection of the null hypothesis that the error correlations are all zero. This highlights the necessity of including a rich set of sociodemographic controls to avoid residential self-selection bias.

3.2. Weighting and estimation methodology

As discussed in the previous section, our estimation sample, which requires full energy information, is not a random sample of any population. The strongest factor causing missing energy information is the number of vehicles in the household, and this

Table 2
Descriptive statistics of the variables of the structural equation model (weighted sample, $N = 2079$).

Variable	Mean	Std. dev.*
Annual household fuel consumption in gallons	1173	1201
Total mileage per year for all household vehicles	25,018	28,486
Thousand dwelling units per sq. mile—Census block group	2.61	1.91
Annual household income in units of $10,000	7.08	5.66
Number of children in household	0.69	1.07
Number of workers in household	1.43	1.08
1-worker household	0.36	
2-worker household	0.31	
3-or-more-worker household	0.13	
Number of drivers in household	1.86	1.03
1-driver household	0.32	
2-driver household	0.46	
3-or-more-driver household	0.18	
Respondent has only college degree	0.53	
Respondent has postgraduate degree	0.15	
Respondent is retired	0.23	
Youngest child at least 16–21 and at least 2 adults not retired	0.05	
Single-person household not retired	0.14	
Race is Asian	0.07	
Race is Hispanic	0.11	
Race is Black	0.05	
Race is mixed White & Hispanic	0.06	

* Variables with missing std. dev. are dummy variables defined as $= 1$ if condition is true and $= 0$ otherwise.

is closely related to the endogenous variables in our model. This means that the estimation sample is effectively stratified on an endogenous variable, which implies that standard estimation methods will yield biased coefficient estimates and inferences. There are two basic approaches to getting valid estimates in this situation (see Wooldridge, 2002, Chapter 17): the "structural" approach and the weighting approach. The "structural" approach adds an explicit equation explaining whether a household has complete energy information and then estimates this equation together with the structural equations model described above. The weighting approach uses weighted estimation where the weights compensate for the different probabilities of having complete energy information. The weighting approach is almost always inefficient, but unlike the structural approach it doesn't rely on functional form assumptions that are hard to justify.

We began by trying the structural approach using Heckman's (1979) two-step estimation method. This method starts with a separate binomial probit model of whether the household has complete energy information. Under the assumption that all of the errors in the system are normally distributed, the Mill's ratio estimated from this probit equation can then be added to the substantive structural equations model to control for the bias caused by non-random sampling. When applied to our data this showed that there was no substantial bias. However, small changes in model specification led to strong rejections of the no bias hypothesis. A simple Ramsey test for the joint normality assumption can be carried out by adding the square and cubed Mill's ratio, and this test strongly rejected the joint normality assumption.

We therefore adopted the weighted estimation approach, and we estimated the weights so that the weighted distribution of the number of vehicles (categorized by 0, 1, 2, 3, 4, and 5 or more vehicles per household) in our sample of 2079 households with complete energy information matched the distribution in the entire sample of 2583 California households in the NHTS. The resulting weights range from 0.8 for the 0 vehicle households to 6.4 for the 5 or more vehicle households. Note that we did not use any additional exogenous socioeconomic information about the households to improve the weights since we directly control for these exogenous factors in our structural equations models. Adding these adjustments to the weights would reduce the efficiency of the weighted estimation methods, but it is important to adjust the weights when using the estimates for population projections and simulations.

Our structural model is:

$$y_i = Ay_i + Bx_i + \varepsilon_i,$$
$$Cov(\varepsilon_i) = \Omega.$$

The weighted estimator we use is defined by:

$$\min \sum w_i \big((I - A)y_i - Bx_i\big)' \Omega^{-1} \big((I - A)y_i - Bx_i\big).$$

Where the weights, w_i, are the inverse probability of selection. The covariance of the weighted estimator above is given by:

$$V = \Psi^{-1} \Lambda \Psi^{-1},$$
$$\Psi = -E\left(\frac{\partial^2 w_i L_i(\theta, x_i)}{\partial\theta\partial\theta'}\right),$$
$$\Lambda = E\left(\left(\frac{\partial w_i L_i(\theta, x_i)}{\partial\theta}\right)\left(\frac{\partial w_i L_i(\theta, x_n)}{\partial\theta'}\right)\right).$$

Once the weights are estimated, then most standard software for structural equations models can perform the weighted estimation. Unfortunately these softwares typically use Ψ^{-1} to estimate the covariance of the estimator, and this is clearly biased. We therefore use a "wild" bootstrap (Horowitz, 2002) to generate standard errors for our weighted estimates. This bootstrap works by taking the vector of estimated residuals, e_i, for each observation and multiplying by:

$$\frac{(1-\sqrt{5})}{2} \quad \text{with Probability} = \frac{(1+\sqrt{5})}{(2\sqrt{5})},$$
$$\frac{(1+\sqrt{5})}{2} \quad \text{with Probability} = 1 - \frac{(1+\sqrt{5})}{(2\sqrt{5})}.$$

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 has the advantage that it will yield consistent standard errors even if the errors in the model are heteroskedastic. We used 200 bootstrap iterations, although we checked our final results using 1000 bootstrap iterations, and the results were very stable. We found that the incorrect standard errors (Ψ^{-1}) were downward biased by from 10–1000%, and the weighted estimates are statistically and operationally significantly different from unweighted estimates in many specifications.

One drawback of using weighted estimations is that they are not equivalent to maximum likelihood, so standard likelihood ratio tests of overidentifying restrictions cannot be used. We implemented a bootstrap test for overidentifying restrictions (including the restrictions on the residual correlation matrix) by bootstrapping the difference between the restricted and unrestricted reduced forms for the various models we examined. The reduced form is given by:

$$y_i = Cx_i + \mu_i$$

and the overidentifying (or structural) restrictions are given by:

$$C = (I - A)^{-1} B,$$
$$Cov(\mu_i) = (I - A)^{-1'} \Omega (I - A)^{-1}.$$

Our test statistic is given by:

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

Table 3
Structural regression coefficients (bootstrapped t-statistics in parentheses).

Explanatory variable	Endogenous variable		
	Household fuel usage per year in gallons	Total mileage per year on 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	−64.7 (−6.15)	−1171 (−4.97)	
Total mileage per year on all household vehicles	0.0382 (17.3)		
Annual household income in units of $10,000	13.3 (4.41)	255 (1.04)	−0.017 (−1.99)
Number of children in household	40.0 (4.2)		−0.232 (−5.43)
Number of workers in household	−117 (−1.64)		0.180 (2.42)
1-worker household	97.3 (1.25)	8493 (1.88)	
2-worker household	252 (1.69)	13,316 (2.24)	
3-or-more-worker household	384 (1.54)	23,327 (2.11)	
Number of drivers in household	65.7 (3.35)	13,652 (3.64)	−0.139 (−0.77)
1-driver household		−4537 (−1.19)	−0.701 (−2.34)
2-driver household		−9977 (−1.3)	−1.013 (−2.42)
3-or-more-driver household		−8777 (−0.78)	−1.078 (−1.68)
Respondent has only college degree	−45.9 (−2.22)		
Respondent has postgraduate degree	−74.9 (−3.03)		
Respondent is retired	−40.0 (−1.43)	3729 (0.59)	−0.409 (−3.04)
Youngest child at least 16–21 and at least 2 adults not retired		−11,669 (−1.66)	−0.700 (−3)
Single-person household not retired			0.218 (1.37)
Race is Asian	−34.9 (−1.25)	−3286 (−1.38)	0.601 (3.11)
Race is Hispanic	−26.5 (−1.01)	−2655 (−0.86)	0.684 (4.24)
Race is Black			0.908 (4.89)
Race is mixed White & Hispanic			0.713 (3.87)

where C_R are the restricted reduced form estimates, C_U are the unrestricted reduced form estimates, 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. This test appears to work well since it ruled out many possible model specifications.

Finally, we also implemented a simple Hausman (1978) test for the null hypothesis that the weights are actually exogenous. This test compares the weighted estimates with standard maximum likelihood estimates ignoring the weights. When applied to our preferred model this test also does not reject the null hypothesis that the weights are exogenous, but, as with the "structural" Heckman test, this result is very sensitive to slight changes in model specification. We therefore decided to be conservative and use the weighted estimates for our empirical results. Although inefficient, they are consistent under the widest array of assumptions about the underlying data generation process.

4. Estimation results

The best model uses housing density at the Census block level, although the other six land use variables also produce acceptable models and similar results. The structural equation model was estimated using weighted three-stage least squares with bootstrapped standard errors as described in Section 3, and the results are given in Table 3. Note that the estimates in Table 3 are computed under the assumption that the structural errors are uncorrelated. The overidentifying restrictions for this model cannot be rejected at any usual level of confidence. Table 4 gives the restricted reduced form estimates corresponding to the structural model in Table 3. The reduced form gives the total impact of the exogenous variables on endogenous variables. Note that the exclusion restrictions imposed on the structural model in Table 3 imply different exclusion restrictions on the restricted reduced form in Table 4 due to the nonlinear relationship between the two models.

The squared multiple correlations for the structural equations are 0.11 for housing density, 0.37 for annual mileage, and 0.95 for annual fuel usage. For the reduced-form equations, the squared multiple correlations are 0.11 for housing density (same as the structural R^2 because there are no endogenous variable effects on housing density), 0.37 for annual mileage, and 0.42 for fuel usage.

4.1. Interpretation of results

4.1.1. The effects of land use density

The model implies that, if two households are identical in all aspects measured by the exogenous variables in the model, but one household is located in a residential area that is 1000 housing units per square mile more dense, the household in the denser area will drive 1171 miles per year less than the household in the

Table 4
Reduced form coefficients (bootstrapped t-statistics in parentheses).

Exogenous variable	Endogenous variable		
	Household fuel usage per year in gallons	Total mileage per year on all household vehicles	Dwelling units per sq. mile in units of 1000—Census block group
Annual household income in units of $10,000	24.2	276	−0.017
	(2.92)	(1.12)	(−1.99)
Number of children in household	55.0	271	−0.232
	(5.12)	(3.51)	(−5.43)
Number of workers in household	−129	−211	0.180
	(−1.79)	(−1.91)	(2.42)
1-worker household	422	8493	
	(2.77)	(1.88)	
2-worker household	761	13,316	
	(3.42)	(2.24)	
3-or-more-worker household	1274	23,327	
	(2.93)	(2.11)	
Number of drivers in household	596	13,815	−0.139
	(4.10)	(3.59)	(−0.77)
1-driver household	−128	−3716	−0.701
	(−0.86)	(−0.96)	(−2.34)
2-driver household	−315	−8792	−1.013
	(−1.07)	(−1.12)	(−2.42)
3-or-more-driver household	−265	−7515	−1.078
	(−0.59)	(−0.65)	(−1.68)
Respondent has only college degree	−45.9		
	(−2.22)		
Respondent has postgraduate degree	−74.9		
	(−3.03)		
Respondent is retired	129	4208	−0.409
	(0.60)	(0.67)	(−3.04)
Youngest child at least 16–21 and at least 2 adults not retired	−400	−10,850	−0.700
	(−1.60)	(−1.55)	(−3.00)
Single-person household not retired	−14.1	−256	0.218
	(−1.31)	(−1.30)	(1.37)
Race is Asian	−199	−3989	0.601
	(−2.17)	(−1.64)	(3.11)
Race is Hispanic	−172	−3456	0.684
	(−1.54)	(−1.11)	(4.24)
Race is Black	−58.7	−1063	0.908
	(−3.93)	(−3.51)	(4.89)
Race is mixed White & Hispanic	−46.1	−835	0.713
	(−3.14)	(−2.87)	(3.87)

less dense area. This is the net effect of vehicle ownership level and trip patterns. The household in the denser area will consume 64.7 fewer gallons of fuel, and this effect of residential density on fuel usage is decomposed into two paths of influence. The mileage difference of 1171 miles leads to a difference of 44.7 gallons (using 0.0382 gallons per mile, the estimated direct effect of mileage on fuel consumption, implying a fuel economy of 26.2 miles per gallon). However, there is an additional direct effect of density on fuel consumption of 20 gallons per 1000 housing units per square mile. This is due to the relationship between residential density and fleet fuel economy, a result of vehicle type choice.

4.1.2. Exogenous variable effects

4.1.2.1. Number of drivers As expected, the number of household drivers has a strong influence on household annual mileage and fuel consumption. However, the number of drivers also affects the choice of residential density. Thus, the total effect on mileage is due to both a direct effect and an effect channeled through residential density. In turn, the effect on fuel consumption is a sum of a direct effect, an effect channeled through mileage, and an effect channeled through residential density. 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.

Drivers per household has a negative diminishing marginal effect on choice of residential density. All else held constant, the model predicts that a household with one driver will locate in a residential area that is less dense by 840 dwelling units per square mile, when compared with a household with no drivers; a household with two drivers will locate in a residential area that is less dense by about 450 dwelling units per square mile, when compared with a household with one driver; and the difference in density between two- and three-driver households declines to about 200 dwelling units per square mile.

The influence of drivers per household on annual vehicle usage and fuel consumption does not exhibit such diminishing marginal effects, and the main nonlinearities involve the effects of more than two drivers. Based on the reduced form results in Table 4, adding the first driver in the household increases annual mileage by 10,100, and adding an additional driver leads to an additional 8700 miles per year. From two to three drivers per household the added mileage per year is 15,100 miles, and from three to four it is 13,800. The effects of the number of drivers on fuel usage follow the same trend, but the rates of increase per driver are slightly greater. This is due to an additional positive direct effect of the number of drivers on fuel usage, indicating a lowering of fleet fuel economy as a function of the number of drivers.

4.1.2.2. Number of workers There is a positive linear effect of the number of workers on residential density. Households with more workers tend to live in higher density areas, *ceteris paribus*. As in the previous case of household drivers, the total effects of number of workers on annual mileage and fuel usage are both nonlinear, each being captured by three variables: a continuous variable and dummy variables for one-worker, two-worker and three-or-more-worker households. However, in contrast to number of drivers, the greatest marginal effect for number of workers is the difference in mileage and fuel consumption attributed to the difference between two to three workers, which is significantly greater than the differences between one and two workers, and somewhat greater than the difference between zero and one worker. The model implies that increases in total household mileage are generally shorter for the second worker in the household and longer for the third worker, in comparison to the first worker. Fuel consumption per worker generally tracks annual mileage, with the exception that fuel consumption is more linear than mileage in the range of zero to two workers, implying that first workers generally use more fuel efficient vehicles.

4.1.2.3. Income The model predicts that fuel usage increases linearly with income, and this is caused by all three factors. Higher income translates into: (1) choice of lower density residential location, (2) greater total driving distances, independent of the greater distances caused by lower densities, and (3) lower overall fuel economy of the household fleet.

4.1.2.4. Number of children Fuel usage increases with number of children due to two factors. Larger families tend to choose lower residential density, which in turn increases total mileage. In addition, fuel economy decreases as a function of the number of children, due to increased likelihood of a least one van or SUV in the household fleet.

4.1.2.5. Education Only two education dummy variables were found to be significant. Households headed by a respondent with a college degree tend to have a vehicle fleet with greater overall lower fuel economy than their less educated counterparts. This effect is accentuated if the household is headed by a respondent with a postgraduate degree.

4.1.2.6. Life cycle effects Retired two-person households tend to live in lower-density residential areas. However, the positive influence of lower residential density on fuel consumption is partially offset by a vehicle fleet with higher fuel economy, probably due to a lower likelihood of vans, pickup trucks and SUVs.

Households with older children choose to live in lower density areas. In California, many children over sixteen years of age have driving licenses, so the effects of this variable on vehicle usage and fuel consumption should be combined with the household drivers variables. If an additional household driver is a child 16–21 years of age, the model predicts that the additional vehicle usage and fuel consumption will be less than if the driver is not such a child.

Finally, non-retired single-person households also tend to live in higher density areas. This translates into lower annual mileage and fuel consumption strictly through the direct effect of land use density.

4.1.2.7. Race and ethnicity Four race and ethnicity variables were determined to have significant effects on choice of residential density and mobility. Households which are solely Black, solely Asian, solely Hispanic, or mixed White and Hispanic, all tend to reside in higher-density areas, compared to other households, predominately solely White households. This leads to lower vehicle usage and fuel consumption for all of these groups. In addition, there are possible direct travel and fuel economy effects for Asian and Hispanic households, but these effects are not estimated with precision. Further research is needed to improve our understanding of these and other demographic influences on residential transportation fuel consumption.

5. Conclusions and directions for further research

We specified a simultaneous equation model that accounts for self selection effects in estimating the influence of residential density on household vehicle annual mileage and fuel consumption. This model was estimated using a method that corrects for missing data that is non-random and related to the endogenous variables. Once we included a complete set of sociodemographic control variables, we could not reject the hypothesis that there are not significant self-selection effects (similar to Bhat and Guo, 2007). We find that density directly influences vehicle usage, and both density and usage influence fuel consumption. Comparing two households that are similar in all respects except residential density, a lower density of 1000 (roughly 40% of the mean value) housing units per square mile implies a positive difference of almost 1200 miles per year (4.8%) and about 65 more gallons of fuel per household (5.5%). This total effect of residential density on fuel usage is decomposed into to two paths of influence. Increased mileage leads to a difference of 45 gallons, but there is an additional direct effect of density through lower fleet fuel economy of 20 gallons per year, a result of vehicle type choice.

Unfortunately for those wishing to use land use planning to control residential vehicle use, it is very difficult to increase the density of an established urban area by 40%. Downs (2004, Chapter 12) shows that increasing the density of an existing metropolitan area by 40% requires extreme densities of new and infill development. Bryan et al. (2007) have recently developed a consistent historical database of U.S. city and regional densities. These data show that only 30 out of 456 cities increased population density more than 40% between 1950 and 1990, and the median city in this sample decreased population density by 36%. The cities that did increase population density by more than 40% are similar to Santa Ana, California. They experienced large increases in low-income immigrants into very tight housing markets. The increase in densities in these cities was largely accommodated by cramming more people into the existing housing stock. Of course, increasing dwelling unit density is even harder than increasing population density.

As expected, the most important exogenous influences are number of household drivers and number of workers, but education and income also are significant. Isolating the effects of number of workers on fuel consumption allows the development of models aimed at forecasting the effects of employment levels on residential transportation energy consumption. There are also demographic, race, and ethnicity effects, as retired households are more likely to live in less dense residential areas, and singles and non-White households are more likely to live in denser areas.

This research can be usefully extended in a number of directions. Adjunct geographic location information can be merged into the NHTS dataset to provide more information about the households' neighborhood characteristics. For those households in major metropolitan areas it might be possible to obtain information on accessibility to public transportation. An expanded model can then be developed to jointly determine public transit accessibility along with residential density and transportation energy use.

Detailed geographic information can also be utilized to empirically examine the claim that balancing the number of residences and jobs within a community will reduce residential transportation fuel use. Tract-level Census data could be used to develop measures of "jobs-housing imbalance" for each of the NHTS California sample members and then test whether these measures have any significant impact on vehicle use and fuel use.

The present method for handling the endogenous sample selection caused by missing energy information also invites improvement. Ideally both the structural and weighting methods should yield the same quantitative results. The structural method should yield more efficient estimates if the equations explaining the missing data process are correctly specified. The problem is likely due to the joint normality assumption required by standard structural methods. Bhat and Eluru (2008) have developed a promising new methodology using copulas can be used to relax this assumption.

Finally, the present research concentrates on California, using only that portion of the NHTS national sample. This work can be expanded to the national level, both as a check on the stability of the models and to empirically examine the claim that California driving behavior has unique characteristics that cannot be captured by standard socioeconomic measures.

References

Badoe, D., Miller, E., 2000. Transportation—land use interaction: Empirical findings in North America, and their implications for modeling. Transportation Research D 5, 235–263.

Bento, A.M., Cropper, M.L., Mobarak, A.M., Vinha, K., 2005. The impact of urban spatial structure on travel demand in the United States. Review of Economics and Statistics 87, 466–478.

Bhat, C.R., Eluru, N., 2008. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Technical Paper. Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin (June).

Bhat, C.R., Guo, J.Y., 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. Transportation Research B 41, 506–526.

Boarnet, M., Sarmiento, S., 1998. Can land-use policy really affect travel behavior? A study of the link between non-work travel and land-use characteristics. Urban Studies 35, 1155–1169.

Bryan, K.A., Minton, B.D., Sarte, P.G., 2007. The evolution of city population density in the United States. Federal Reserve Bank of Richmond Economic Quarterly 93, 341–360. Data accessed from http://www.richmondfed.org/research/research_economists/files/urbandensitycode.zip, on August 25, 2008.

Downs, A., 2004. Still stuck in traffic: Coping with peak-hour traffic congestion. The Brookings Institution, Washington, DC.

Ewing, R., Cervero, R., 2001. Travel and the built environment: A synthesis. Transportation Research Record 1780, 87–113.

Fang, A., 2008. A discrete-continuous model of households' vehicle choice and usage, with an application to the effects of residential density. Transportation Research B 42, 736–758.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.

Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.

Horowitz, J.L., 2002. The bootstrap. In: Leamer, E.E., Heckman, J.J. (Eds.), Handbook of Econometrics, vol. 5. Elsevier, Amsterdam (Chapter 52).

Kahn, M.E., 2000. The environmental impact of suburbanization. Journal of Policy Analysis and Management 19, 569–586.

Newman, P., Kenworthy, J., 1999. Costs of automobile dependence: Global survey of cities. Transportation Research Record 1670, 17–26.

NHTSA, 2008. CAFÉ overview—frequently asked questions. U.S. National Highway Transportation Safety Agency website, http://www.nhtsa.dot.gov/CARS/rules/CAFE/overview.htm (accessed September 13, 2008).

ORNL, 2004. The 2001 National Household Travel Survey. Oak Ridge National Laboratory for the U.S. Department of Transportation, http://nhts.ornl.gov/publications.shtml (accessed September 14, 2008).

Pickrell, D., Schimek, P., 1999. Growth in motor vehicle ownership and use: Evidence from the National Personal Transportation Survey. Journal of Transportation and Statistics 2, 1–17.

Schipper, M.A., Pinckney, D.S., 2004. Supplementing the 2001 National Household Travel Survey with energy-related data. Presented at the 2004 Annual meeting of the Transportation Research Board, January 11–15, Washington, DC.

Wooldridge, J.A., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.