The Impact of Residential Density on Vehicle Usage and Energy Consumption*

June 22, 2005

Thomas F. Golob
Institute of Transportation Studies
University of California Irvine
tgolob@uci.edu

David Brownstone (*corresponding author*)
Department of Economics
3151 Social Science Plaza
University of California Irvine
Irvine, California, 92697-5100
dbrownst@uci.edu
Tel: 949-824-6231
Fax: 949-824-2182

* This is a revised version of Working Paper UCI-ITS-WP-05-1, Institute of transportation Studies, 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 provided many useful comments on an earlier draft, but the authors bear sole responsibility for any remaining errors.

The Impact of Residential Density on Vehicle Usage and Energy Consumption

Thomas F. Golob
David Brownstone
University of California Irvine

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 Transportation Survey (NHTS). Comparing two California households that are similar in all respects except residential density, a lower density of 1,000 housing units per square mile implies a positive difference of almost 1,200 miles per year and about 65 more gallons of fuel per household. 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.

Keywords: residential density, vehicle use, vehicle fuel consumption, simultaneous equations, self-selection.

Background

Understanding total residential transportation energy usage is vital for the planning of conservation measures and the evaluation of incentives and mandates aimed at improving vehicle fuel efficiency. The annual vehicular fuel consumption of households is clearly the outcome of complex decisions that involve the number of vehicles the household owns or otherwise has available (including company cars), the makes, models and vintages of these vehicles, allocation of vehicles and activities among drivers, and choices of activity sites, modes of transportation, and the chaining and combining of activities. Lifestyle clearly plays a major role in all of these decisions, as do numerous demographic and socioeconomic factors, including age, race, ethnicity, education, and income. The spatial location of the residence is a determining factor for trip lengths, travel speeds, and the availability and level of service of public transportation and non-motorized modes.

Lower density urban areas are usually indicative of decentralization and an extended aggregation of urban communities, which is sometimes referred to as urban sprawl. Urban sprawl is a contentious issue among social scientists and planners, and the debate is sometimes emotional. On the one hand, sprawl is viewed as the growth of cities in an unplanned or otherwise wasteful manner, resulting in the segregation of land uses, a loss of farm and open lands, degradation of the urban core, and an automobile-dependent transportation system, with accompanying increases in traffic congestion, energy consumption, and air pollution (Ewing, 1997). On the other hand, it is viewed as the inevitable outcome of a market preference for low-density settlements and for automobile travel in preference to mass transit (Gordon and Richardson, 1997). It has been pointed out that urban sprawl is not synonymous with all low densities, but rather with scattered development, commercial strip development, or large expanses of low density or single-use development. Whatever definition is used for sprawl, the issue we address here is the degree to which residential density affects residential vehicle usage and energy consumption. We view this as a needed contribution to the debate swirling around urban land use density.

Studies of the effects land use density (or other measures of urban form) on vehicle usage can be divided into aggregate and disaggregate studies. Aggregate studies use spatially defined averages for all variables; observations usually being for cities or metropolitan areas, but also for zones or neighborhoods within cities. Disaggregated studies use household observations of vehicle usage and either city-wide, zonal, or neighborhood averages for urban form variables. However, almost all of these studies have ignored selectivity problems regarding the residential density variables. Residential location choices are inextricably interrelated with household vehicle ownership and type choices, and the levels of usage of all household vehicles. Persons choose their residential location on the bases of preferred lifestyles, available housing stock type and cost, the socioeconomic characteristics of their neighbors, demand for educational and recreational facilities, locations of other activity sites, security concerns, and their tolerances and preferences for various modes of transportation. Many of these community and neighborhood characteristics are related to residential density. Boarnet and Crane (2001) point out that people who dislike driving might both drive less and choose to live in a high density, mixed use neighborhood that supports transportation alternatives other than driving. Handy (1996) concludes that accounting for self-selection with respect to residential neighborhood is a critical issue in studying the broader problem of relationships between urban various measures of urban form and travel.

Newman and Kenworthy (1989a; 1989b) is a widely quoted aggregate study that finds a negative correlation between density and residential energy use for an international sample of cities. Aggregate studies need to control not only for socioeconomic and demographic differences among households in each area, but also the differences in transportation infrastructure, and the cultural, political historical, and economic differences among the areas. As pointed out by Gomez-Ibañez (1991), Newman and Kenworthy fail to control for such effects and uses suspect data. Steiner (1994) and Handy (1996) both review many other studies and conclude that, by masking within-area variations in both urban form and travel behavior, aggregate studies are generally not capable of uncovering true relationships between density and travel.

Disaggregate studies have also not effectively accounted for the simultaneity of residential location choice on the effects of residential density of vehicle usage. One device recently used in to account for selectivity in choice of residential location is to use city- or metropolitan-area-wide data on urban form together with disaggregate travel data (Levinson and Kumar, 1997; Bento, *et al.*, 2003). Unfortunately, the use of city-wide measures of urban form does not improve upon the problem, encountered in aggregate studies, of ignoring potentially important influences on travel of differences in urban form at the zonal and neighborhood level.

We adopt a more direct approach to the problem of selectivity bias in disaggregate studies. The approach is to apply a simultaneous equations model in which residential density, vehicle usage, and fuel consumption are joint endogenous variables. In this way we can model socioeconomic and demographic effects on each of these three endogenous variables, while simultaneously capturing the direct effects of residential location on the vehicle usage and energy consumption.

Data

2001 National Household Transportation Survey (NHTS)

The NHTS is a household-based travel survey conducted every five years by the U.S. Department of Transportation. Prior to 2001, the portion of the NHTS focusing on local trips was known as the National Personal Transportation Survey (NPTS) and the long-distance travel portion of the survey was called the American Travel Survey. There are 2,583 California (CA) households in the 2001 NHTS sample, representing 9.9% of the total base sample of 26,038. (The 2001 NHTS survey also contains nine add-on samples for specific geographical regions, all of which are outside of California.) 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 mile 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 recreation vehicles. Odometer readings were obtained at two dates, generally a few months apart, in order to provide 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).

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. Regarding vehicle availability (generally meaning vehicle ownership or leasing), 5.3% of CA households in this sample have no vehicles, 28.3% have one vehicle, 39.2% have two vehicles, 16.7% have three vehicles, 6.8% have four vehicles, and 3.6% have five or more vehicles. The weighted breakdown, according to the NHTS weights for all household is: 7.5% no vehicles, 33.8% one vehicle, 35.0% two, 15.2% three, 5.4% four, and 3.0% 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, in terms of gallons of gasoline, or gasoline equivalent gallons for alternative fuel vehicle, are reported in Schipper and Pinckney (2004). These procedures used reported and imputed odometer readings, together with fuel economy test results for each vehicle make, model and vintage, and involved adjustments for on-road shortfalls of vehicle dynamometer test results, seasonal variations, and relationships between total mileage and average trip lengths. These adjustments rely heavily on results from previous Residential Transportation Energy Consumption Surveys (RTECS) conducted by the Energy Information Administration (EIA). The RTECS was abandoned in 1994, and the intention is to use the 2001 NHTS to provide alternative information on residential fuel usage.

Of the 2445 CA households with vehicles (unweighted sample), 1941 or 79.4% have full information required for calculating annual fuel usage for all vehicles, or residential transportation energy consumption. The necessary data are annual mileage and vehicle make, model and vintage for all vehicles. The breakdown of the weighted CA household sample by vehicle ownership and whether or not energy consumption data is available is shown in Figure 1. By definition, the 195 households (138 households in the unweighted sample) without vehicles have zero residential transportation fuel usage. Consequently, 2079, or 80.5% of all California households have full information on transportation fuel usage. Since each household vehicle must be accounted for in order for full energy consumption information, the proportion of households with full information is a decreasing function of vehicle ownership level. Full energy information is available for the vast majority of 1- and 2-vehicles 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.

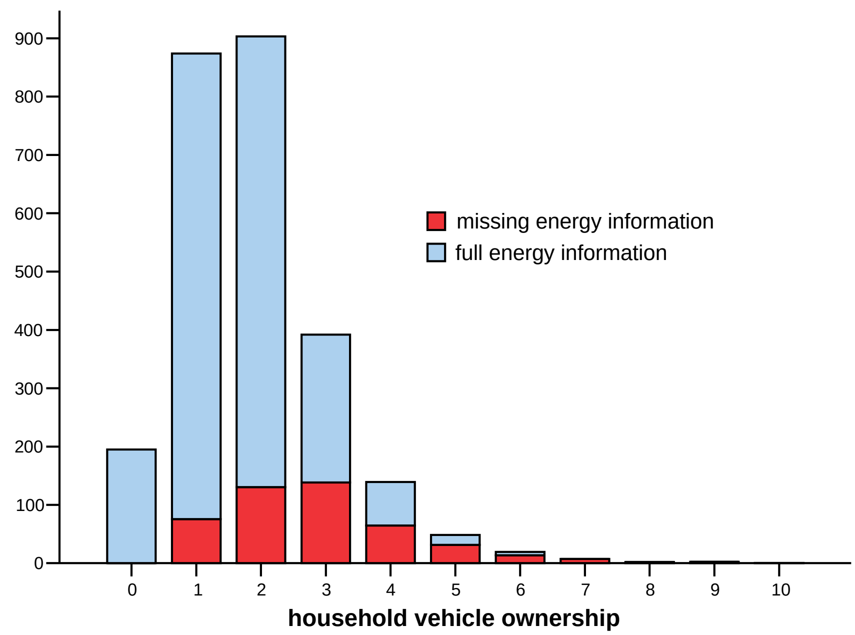

Figure 1 California 2001 NHTS Household Sample by Number of Vehicles, Broken Down by Availability of Energy Consumption Information for All Vehicles (N = 2583, weighted).

Mean fuel consumption for California households in the 2001 NHTS sample is 1,034 gallons, costing $1,485. It should be remembered that this sample is biased due to the greater likelihood of missing energy data for households with greater numbers of vehicles. Mean fuel consumption per vehicle was 562 gallons per year in 2001.

Land Use Densities

The 2001 NHTS provides several measures of land use related to household location. These data, supplied by the marketing information resources firm Claritas, are computed from 2000 Census data at both the census tract and block group levels. Two of these land use measures are categorical. Population per square mile is provided in eight categories both at the block group and track level, and in our models the categories are assigned the mid-point values shown in Table 1. The largest concentration of California households is in the 4-10,000 per square mile category, but all categories are fairly well represented. Housing units per square mile are provided in six categories, as shown in Table 2. More than forty-one percent of California households in the NHTS sample are in 1-3,000 housing units per square mile category.

Table 1 Population Per Square Mile for California Households with Full Energy Information (N = 2079)

		Block Group Level		Tract Level	
Category	Value	Frequency	Percent	Frequency	Percent
< 100	50	94	4.5	116	5.6
100 - 500	300	100	4.8	123	5.9
500 - 1,000	750	83	4.0	94	4.5
1 – 2,000	1,500	165	7.9	173	8.3
2 – 4,000	3,000	270	13.0	303	14.6
4 - 10,000	7,000	825	39.7	809	38.9
10 - 25,000	17,000	428	20.6	373	17.9
> 25,000	30,000	114	5.5	88	4.2
Totals		2079	100.0	2079	100.0

Three continuous measures of land use intensity are also available. Percentage of renter-occupied housing units is provided at both the block group and tract level. Jobs per square mile are appropriately measured at the tract level. Correlations among the seven land use variables are listed in Table 3.

Table 2 Housing Units Per Square Mile for California Households with Full energy Information (N= 2079)

		Block Group Level		Tract Level	
Category	Value	Frequency	Percent	Frequency	Percent
< 50	25	101	4.9	131	6.3
50 - 250	150	125	6.0	143	6.9
250 - 1,000	700	293	14.0	326	15.7
1 - 3,000	2,000	835	40.2	854	41.1
3 - 5,000	4,000	370	17.8	335	16.1
> 5,000	6,000	355	17.1	290	13.9
Totals		2079	100.0	2079	100.0

Table 3 Correlations Among Land Use Variables for California Households with Full Energy Information (N=2079)

	Pop. per sq. mi. tract level	H. units per sq. mi. block group	H. units per sq. mi. tract level	Percent renter-occ. block group	Percent renter-occ. tract level	Jobs per sq. mi. tract level
Pop. per sq. mi. block group	.840	.857	.783	.582	.587	.791
Pop. per sq. mi. tract level		.840	.757	.519	.588	.881
H. units per sq. mi. block group			.847	.616	.601	.703
H. units per sq. mi. tract level				.550	.626	.792
Percent renter-occ. block group					.882	.502
Percent renter-occ. tract level						.559

Vehicle Ownership, Type Choice, and Fuel Usage

Total annual residential vehicular energy consumption is graphed as a function of household income category in Figure 2. The differences in means displayed in Figure 2 are significantly significant. There is a similar pattern for all households and households with vehicles, but the rate of increase in fuel consumption as a function of income is less for households with vehicles and less than $50,000 income. While it is not instructive to estimate and plot regression curves, due to the uncertainty in locating the mean of the top income class (more than 18% of households fall into the $100,000 and above category), total gasoline consumption per year appears to increase at a decreasing rate with income, especially for incomes greater than $50,000. The mean gasoline consumption for households with missing income data (7.4% of all households) is most similar to that of households in the $30-40,000 range.

One important question is the extent to which full energy information is related to household variables separate from vehicle ownership. Tests were conducted to measure the statistical significance of the relationships between the dummy variable indicating energy data availability and various key household variables. Since sample composition is the issue, these analyses were applied using the unweighted sample of households with vehicles (N = 2443), with an overall level of energy data availability of 79.5%.

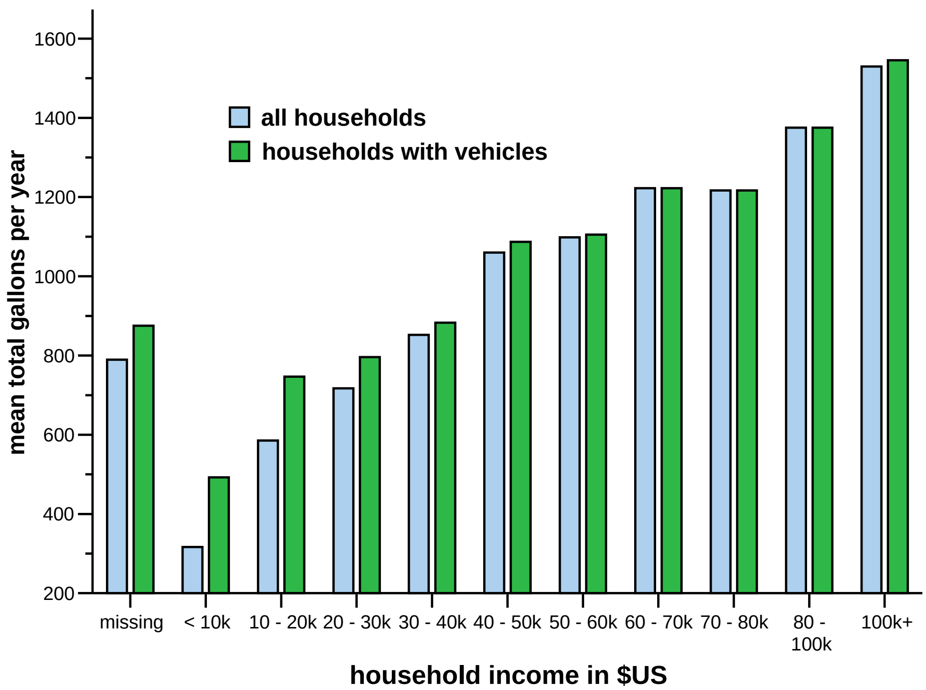

Figure 2 Annual Fuel Consumption in Gasoline-equivalent Gallons by Income for All Households (N= 2079) and for Households with Vehicles (N= 1943)

Due to the strong relationship between data availability and vehicle ownership level, we can expect that all household variables strongly correlated with vehicle ownership will also be related to energy data availability. Indeed, significant ordinal correlations were found between a dummy variable measuring data availability and the following household variables: household size, number of adults, number of workers, and number of drivers.

Variables strongly related to car ownership level should also be related to energy data availability. These include respondent education, housing type and home ownership. Education level displays a significant ordinal correlation with energy data availability. With respect to housing type, full data is available for 85.8% of apartment and condominium dwellers, 81.5% of duplex, row- and townhouse dwellers, but only 74.8% of households in single-family detached houses. With respect to home ownership, full data is available for 82.1% of renters, but only 75.8% of homeowners.

There is no significant relationship between energy data availability and the following variables: respondent race, household location within California (four metropolitan regions and the rest of the state), and household annual income. The income distribution is shown in Figure 3, where zero-car ownership is included as a separate category. The split between full energy information and missing energy information for households with vehicles, is weakly related to income (X^2 = 18.69 with 10 degrees of freedom, corresponding to p = .044).

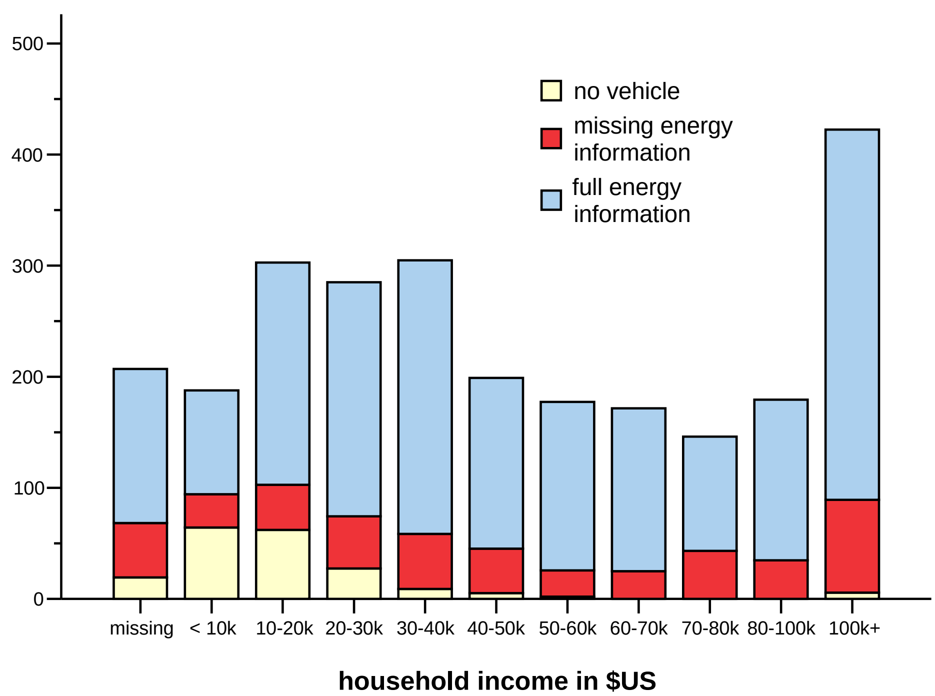

Figure 3 Whether Full Energy Consumption Information is Available for All Vehicles by Household Annual Income Category (N = 2583)

Regarding mileage and fuel consumption as a function of the number of vehicles available to a household, fuel consumption per vehicle varies more than mileage per vehicle, as shown in Figure 4. The differences in mean fuel consumption per vehicle are statistically significant (at the p =.02 level), but the differences in annual mileage are not significant. This is apparently due to the mix of vehicle types in multi-vehicle households. As shown in Figure 5, most (about fifty-nine percent) of the vehicles available to California households in 2001 were cars, followed by pickup trucks, SUVs and vans. As shown in Figure 6, most cars are in two-vehicle households (37.1%), followed by almost equal numbers in single-vehicle and three-vehicle households (22.7% and 22.5% respectively). Most vans, SUVs and pickup trucks are also in two-vehicle households, but all three types of trucks are more prevalent in three-vehicle households than in one-vehicle households. Pickup trucks are more prevalent in four-vehicle households (15.3%) than in single-vehicle households (7.4%).

Overall, SUVs are driven most, followed by pickup trucks and vans, then cars (Figure 7). The differences among vehicle types in terms of fuel consumption are accentuated, due to lower fuel economy for SUVs and vans. Both sets of means graphed in Figure 7 are significantly different at the $p < .001$ level.

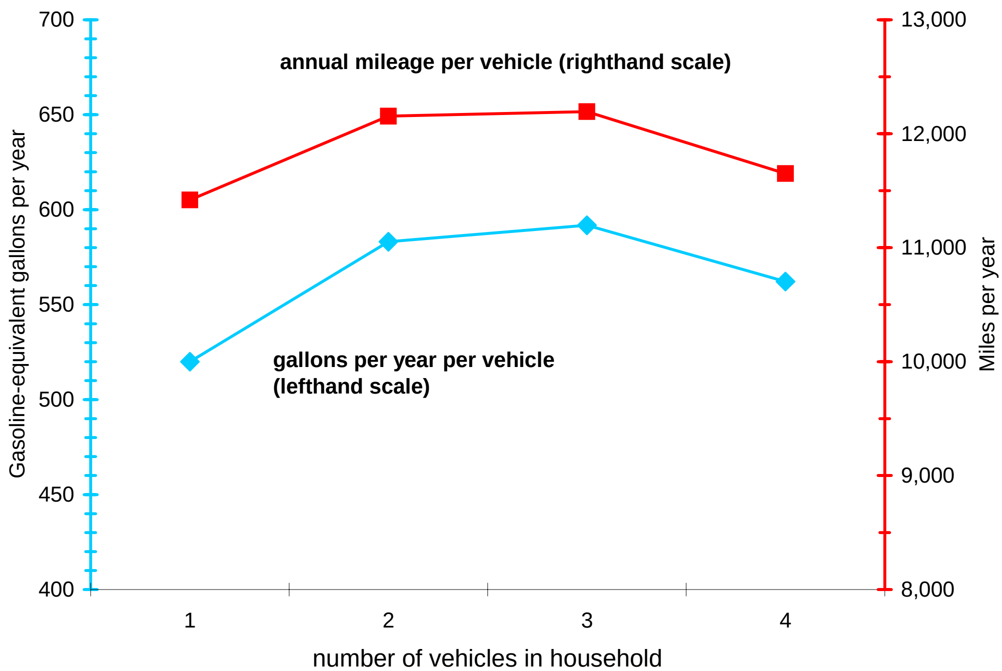

Figure 4 Annual Fuel Consumption and Mileage by Number of Vehicles for Households with One to Four Vehicles (N= 1918)

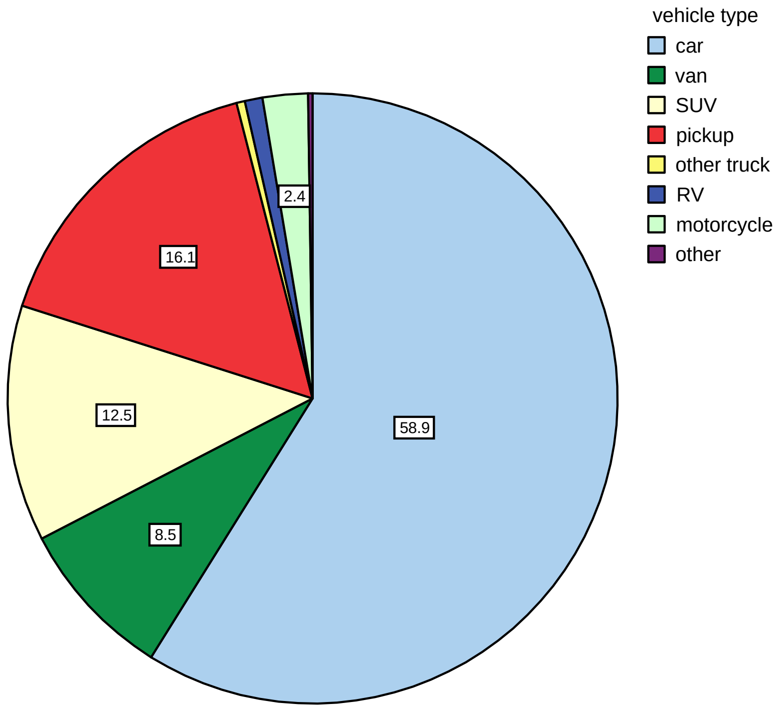

Figure 5 Breakdown of All Vehicles in CA Households by Vehicle Type (N = 4863 vehicles in 2583 weighted households)

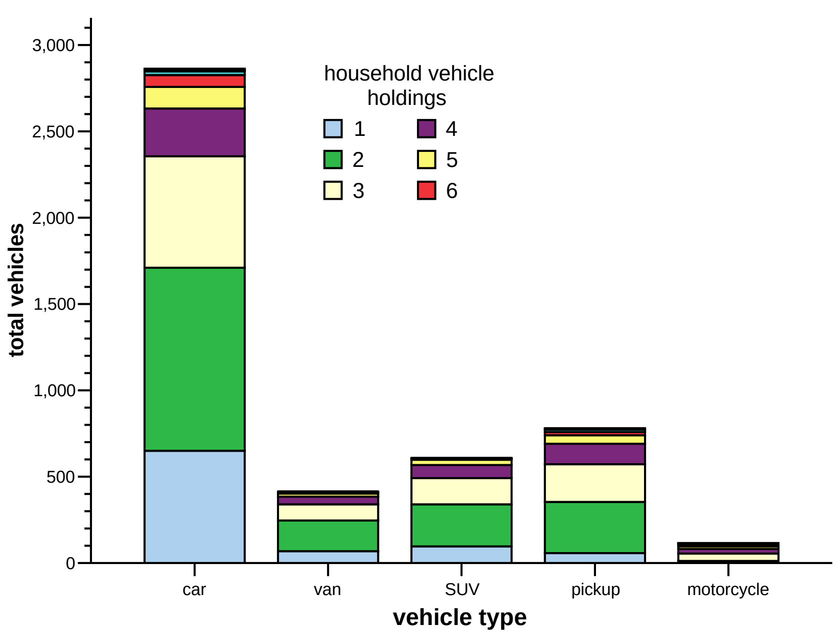

Figure 6 Distribution of Major Vehicle Types Across Households by Vehicle Holdings (N = 4782 vehicles in 2339 weighted households)

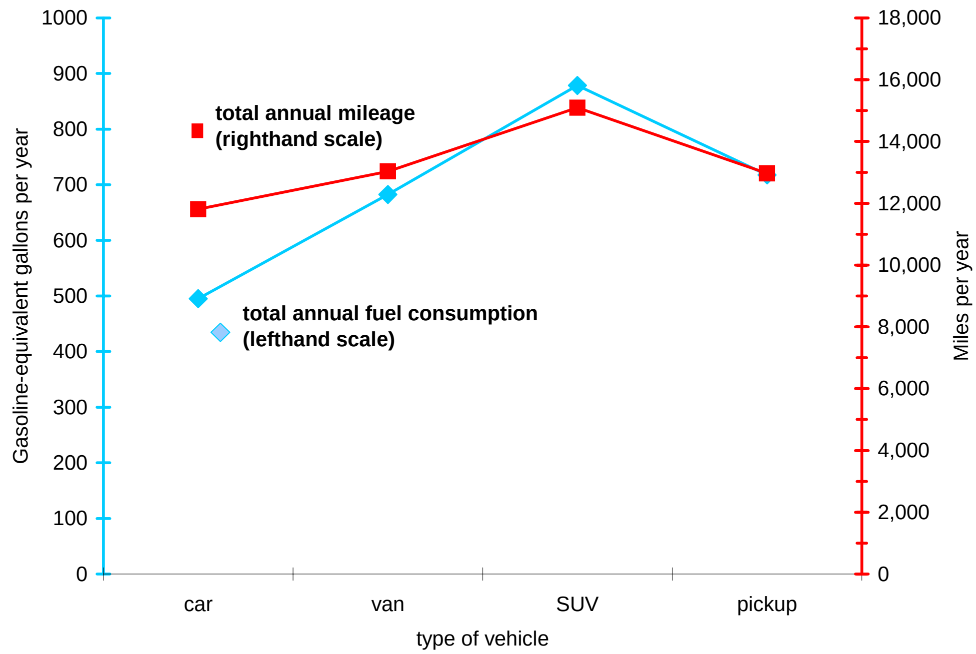

Figure 7 Mean Total Annual Fuel Consumption and Mileage by Vehicle Type for all vehicles with available energy information in CA households (N = 4215 vehicles in 2199 weighted households).

Vehicle Usage and Land Use

As expected, there is a significant 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 displayed in Figure 8. The differences in means for both series in Figure are statistically significant, and linear relationships cannot be rejected at the $p<.01$ 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 5,000 housing units per square mile.

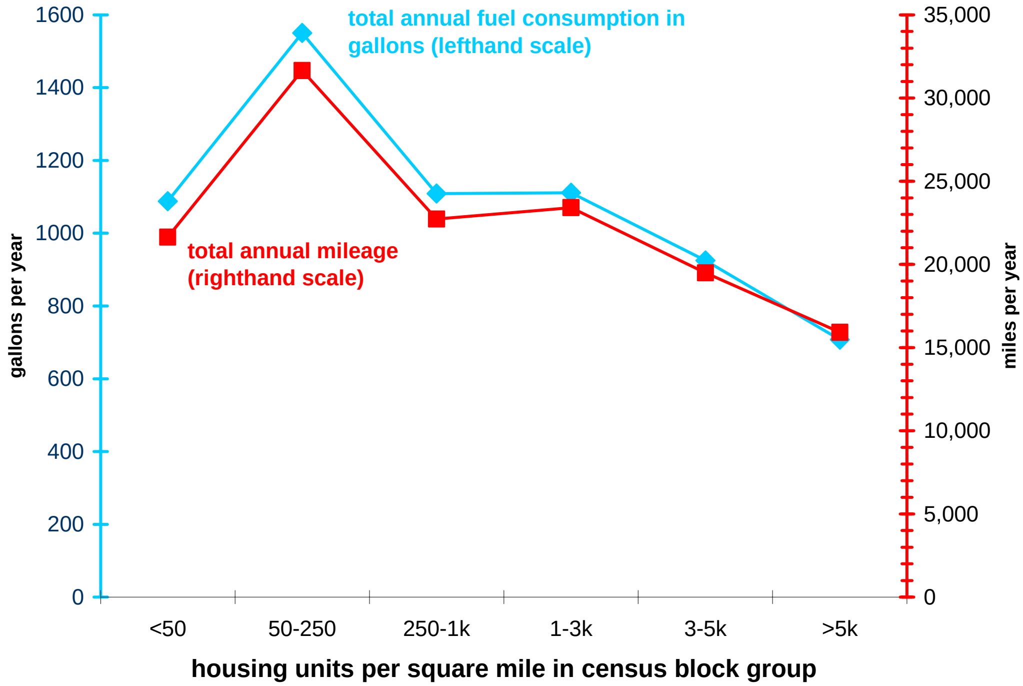

Figure 8 Annual Total Mileage and Fuel Consumption by Residential Density in Terms of Housing Units per Square Mile in Census Block Group

These relationships are caused in large part by differences in household vehicle ownership levels. As shown in Figure 9, vehicle ranges from a high of 2.24 vehicles per household for households located in areas of 50-250 dwellings per square mile, to a low of 1.36 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 Figure 9, the likelihood of owning one of these three types of trucks increases with decreasing density, and the reversal of the trend at the lowest density is less pronounced than it is for mileage and fuel usage.

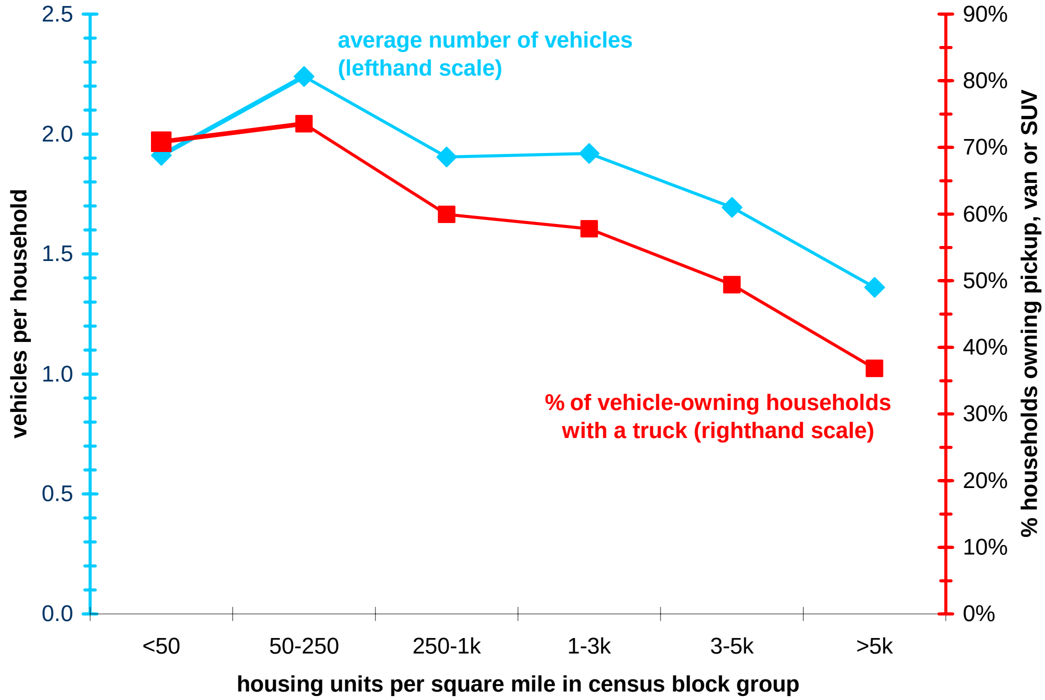

Figure 9 Vehicle Ownership and Percentage of Households With Trucks by Residential Density in Terms of Housing Units per Square Mile in Census Block Group

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 number of household drivers and average household income (Figure 10).

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.

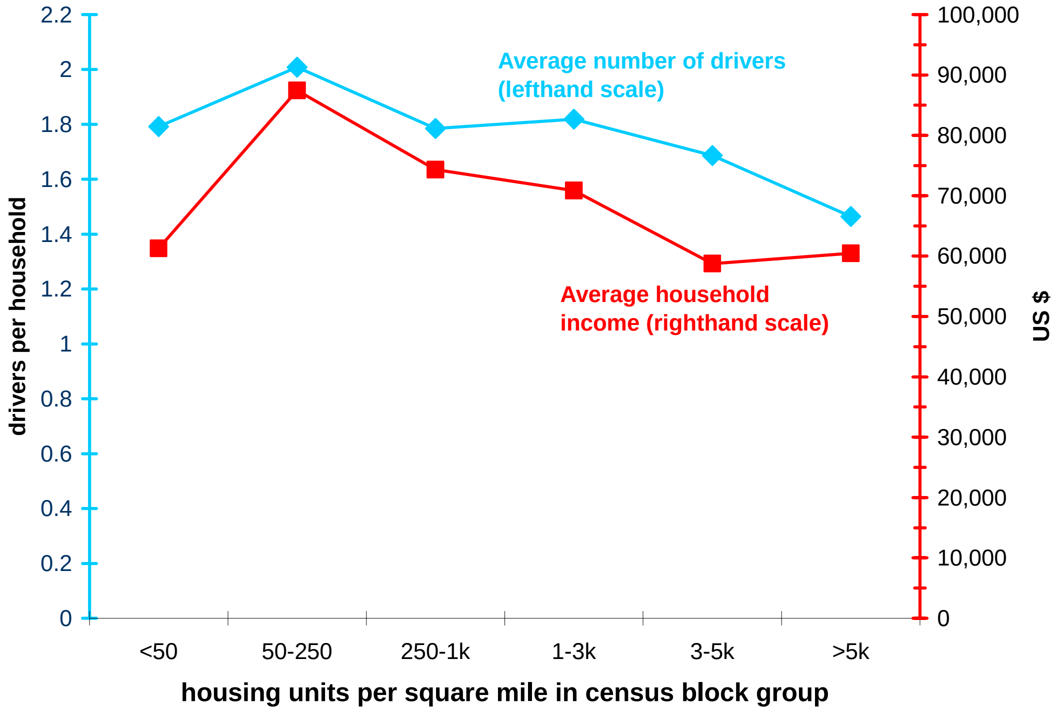

Figure 10 Drivers per Household and Household Income by Residential Density in Terms of Housing Units per Square Mile in Census Block Group

Structural Model of Density of land Use and Fuel Usage

Model Specification

We estimate the effects of land use variables on fuel usage by specifying a structural (or simultaneous) equation model (SEM) with three endogenous variables and many exogenous variables. The three endogenous variables are: (1) housing units per square mile in census block group described previously, (2) total annual miles driven on all household vehicles, and (3) total annual household fuel usage. The postulated structural relationships involving these endogenous variables are shown in the flow (path) diagram of Figure 11. There are three direct effects. First, land use density directly affects annual mileage, because households in lower density areas will choose to have more vehicles, controlling for socioeconomic and demographic differences, and miles driver per vehicle will be greater due to the separation of households and activity sites. Second, Land use also directly affects fuel usage in that households that choose to live in less dense areas also choose to own vehicles with lower fuel economy. As previously shown, fuel economy decreases with decreasing residential density, partly due to the penetration of small trucks (pickup trucks, SUVs, and vans). Third and finally, there is the most important direct link from mileage to fuel usage. The inverse of the coefficient for this direct effect is the effective fuel economy captured by the model. These endogenous effects define a recursive model, so there are no identification problems in the absence of error term correlations.

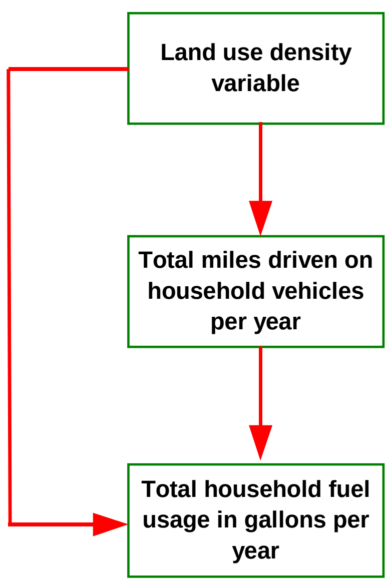

Figure 11 Flow Diagram for the Endogenous Effects in the Structural Equations Model.

The exogenous variables in the structural equations model are all those that were found to describe one or more of the endogenous variables. There are nineteen of these, as described in Table 4. A continuous variable was constructed for income by using the midpoints of the categories shown in Figure 3, 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 useage to land use density.

Weighting and Estimation Methodology

As illustrated in Figure 1, the sample with 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 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 .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 model is:

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

The weighted estimator we us is defined by:

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

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

$$V = \Psi^{-1} \Lambda \Psi^{-1}$$
$$\Psi = -E\left(\partial^2 w_\mathrm{i}\, \mathrm{L_i}(\theta, \mathrm{x_i}) \big/ \partial\theta\partial\theta' \right)$$
$$\Lambda = E\left(\left(\partial w_\mathrm{i} \mathrm{L_i}(\theta, \mathrm{x_i}) \big/ \partial\theta \right) \left(\partial w_\mathrm{i} \mathrm{L_i}(\theta, \mathrm{x_n}) \big/ \partial\theta' \right) \right)$$

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:

$$\left(1 - \sqrt{5}\right)/2 \text{ with Probability} = \left(1 + \sqrt{5}\right)/\left(2\sqrt{5}\right)$$
$$\left(1 + \sqrt{5}\right)/2 \text{ with Probability} = 1 - \left(1 + \sqrt{5}\right)/\left(2\sqrt{5}\right)$$

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 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:

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

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.

Estimation Results

The best model is that using 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 above. The overidentifying restrictions for this model cannot be rejected at any usual level of confidence.

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.

Table 4 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	25018	28486
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
Dummy: 1-worker household	0.36	0.48
Dummy: 2-worker household	0.31	0.46
Dummy: 3-or-more-worker household	0.13	0.34
Number of drivers in household	1.86	1.03
Dummy: 1-driver household	0.32	0.47
Dummy: 2-driver household	0.46	0.50
Dummy: 3-or-more-driver household	0.18	0.38
Dummy: respondent has only college degree	0.53	0.50
Dummy: respondent has postgraduate degree	0.15	0.35
Dummy: respondent is retired	0.23	0.42
Dummy: youngest child at least 16-21 and at least 2 adults not retired	0.05	0.23
Dummy: single-person household not retired	0.14	0.35
Dummy: race is Asian	0.07	0.26
Dummy: race is Hispanic	0.11	0.31
Dummy: race is Black	0.05	0.22
Dummy: race is mixed White & Hispanic	0.06	0.23

The estimated direct effects among the endogenous variable are listed in Table 5. The postulated effects of residential density on both mileage and fuel usage are highly significant. The total effects of each endogenous variables on the other are listed in Table 6. These results are interpreted in the next section.

Table 5 Structural Coefficients (Direct Effects) Among the Endogenous Variables (t-statistic in parentheses)

Influenced endogenous variable	Causal endogenous variable	
	Dwelling units per sq. mile in units of 1,000 – census block group	Total mileage per year on all household vehicles
Total mileage per year on all household vehicles	-1171 (-4.97)	
Household fuel usage per year in gallons	-20 (-5.12)	0.0382 (17.3)

Table 6 Total Effects Among the Endogenous Variables (t-statistic in parentheses)

Influenced endogenous variable	Causal endogenous variable	
	Dwelling units per sq. mile in units of 1,000 – census block group	Total mileage per year on all household vehicles
Total mileage per year on all household vehicles	-1171 (-4.97)	
Household fuel usage per year in gallons	-64.7 (-6.15)	0.0382 (17.3)

The direct (regression) effects of the exogenous variables on the endogenous variables are listed in Table 7. The implications of these results concerning the direct regression effects on the endogenous variables are explored in the next section.

Finally, the total effects of the exogenous variables on the endogenous variable are listed in Table 8. These are known as the coefficients of the reduced-form equations. The effects of the socioeconomic variables that explain land use density are translated to mileage and fuel usage by the direct effects between the endogenous variables. These total effects are interpreted together with direct effects in next.

Table 7 Structural Regression Coefficients (Direct Effects of the Exogenous Variables on the Endogenous Variables) (bootstrapped t-statistics in parentheses)

Exogenous variable	Endogenous variable: Household fuel usage per year in gallons	Endogenous variable: Total mileage per year on all household vehicles	Endogenous variable: Dwelling units per sq. mile in units of 1,000 - census block group
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)
Dummy: 1-worker household	97.3 (1.25)	8493 (1.88)	
Dummy: 2-worker household	252 (1.69)	13316 (2.24)	
Dummy: 3-or-more-worker household	384 (1.54)	23327 (2.11)	
Number of drivers in household	65.7 (3.35)	13652 (3.64)	-0.139 (-0.77)
Dummy: 1-driver household		-4537 (-1.19)	-0.701 (-2.34)
Dummy: 2-driver household		-9977 (-1.3)	-1.013 (-2.42)
Dummy: 3-or-more-driver household		-8777 (-0.78)	-1.078 (-1.68)
Dummy: respondent has only college degree	-45.9 (-2.22)		
Dummy: respondent has postgraduate degree	-74.9 (-3.03)		
Dummy: respondent is retired	-40.0 (-1.43)	3729 (0.59)	-0.409 (-3.04)
Dummy: youngest child at least 16-21 and at least 2 adults not retired		-11669 (-1.66)	-0.700 (-3)
Dummy: single-person household not retired			0.218 (1.37)
Dummy: race is Asian	-34.9 (-1.25)	-3286 (-1.38)	0.601 (3.11)
Dummy: race is Hispanic	-26.5 (-1.01)	-2655 (-0.86)	0.684 (4.24)
Dummy: race is Black			0.908 (4.89)
Dummy: race is mixed White & Hispanic			0.713 (3.87)

Table 8 Reduced Form Coefficients (bootstrapped t-statistics in parentheses)

	Endogenous variable		
Exogenous variable	Household fuel usage per year in gallons	Total mileage per year on all household vehicles	Dwelling units per sq. mile in units of 1,000 - census block group
Annual household income in units of $10,000	24.2 (2.92)	276 (1.12)	-0.017 (-1.99)
Number of children in household	55.0 (5.12)	271 (3.51)	-0.232 (-5.43)
Number of workers in household	-129 (-1.79)	-211 (-1.91)	0.180 (2.42)
Dummy: 1-worker household	422 (2.77)	8493 (1.88)	
Dummy: 2-worker household	761 (3.42)	13316 (2.24)	
Dummy: 3-or-more-worker household	1274 (2.93)	23327 (2.11)	
Number of drivers in household	596 (4.10)	13815 (3.59)	-0.139 (-.77)
Dummy: 1-driver household	-128 (-.86)	-3716 (-.96)	-0.701 (-2.34)
Dummy: 2-driver household	-315 (-1.07)	-8792 (-1.12)	-1.013 (-2.42)
Dummy: 3-or-more-driver household	-265 (-.59)	-7515 (-.65)	-1.078 (-1.68)
Dummy: respondent has only college degree	-45.9 (-2.22)		
Dummy: respondent has postgraduate degree	-74.9 (-3.03)		
Dummy: respondent is retired	129 (.60)	4208 (.67)	-0.409 (-3.04)
Dummy: youngest child at least 16-21 and at least 2 adults not retired	-400 (-1.60)	-10850 (-1.55)	-0.700 (-3.00)
Dummy: single-person household not retired	-14.1 (-1.31)	-256 (-1.30)	0.218 (1.37)
Dummy: race is Asian	-199 (-2.17)	-3989 (-1.64)	0.601 (3.11)
Dummy: race is Hispanic	-172 (-1.54)	-3456 (-1.11)	0.684 (4.24)
Dummy: race is Black	-58.7 (-3.93)	-1063 (-3.51)	0.908 (4.89)
Dummy: race is mixed White & Hispanic	-46.1 (-3.14)	-835 (-2.87)	0.713 (3.87)

Interpretation of Results

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 that is 1,000 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 less dense area. (referencing the direct and total effects given in Tables 5 and 6). This is the net effect of vehicle ownership level and trip patterns. Regarding annual fuel consumption, the household is the denser area will consume 64.7 fewer gallons of fuel, and this effect of residential density on fuel usage is decomposed into to 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 1,000 housing units per square mile (Table 5). This is due to the relationship between residential density and fleet fuel economy, a result of vehicle type choice.

Exogenous Variable Effects

Number of Drivers

As expected, the number of household drivers is the strongest influence of household annual mileage and fuel consumption. However, number of drivers also affects 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 about 700 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 influences 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. For the first driver in the household, the increase in annual mileage is 10,100, and in going from one to two drivers, the increase is 8,700 miles per year. From two to three 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.

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 commuting distances 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.

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.

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.

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 there counterparts. This effect is accentuated if the household is headed by a respondent with a postgraduate degree.

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.

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 due to the inefficiency of the estimation method used to assure unbiased coefficient estimates, 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.

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. Results showed that residential density directly influences vehicle usage, and both density and usage influence fuel consumption. All of these effects are large in magnitude and precisely estimated: Comparing two households that are similar in all respects except residential density, a lower density of 1,000 housing units per square mile implies a positive difference of almost 1,200 miles per year and about 65 more gallons of fuel per household. 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.

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 and 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 model can also be extended to include explicit endogenous indicators for different vehicle types. This will allow us to more clearly identify the impacts of switching from cars to vans and light trucks. The present results indicate that certain sociodemographic groups are likely to drive more fuel-efficient vehicles, and new research might identify how this is related to substituting away from trucks and SUVs, as opposed to choosing newer or more efficient vehicles for the same body types.

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. This calls for experiments with more flexible functional forms and semi-parametric estimation methods to find a better structural model.

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

Bento, A.M., M.L. Cropper, A.M. Mobarak and K. Vinha (2003). The impact of urban spatial structure on travel demand in the United States. World Bank Policy Research Working Paper 3007. The World Bank, Washington, DC. http://econ.worldbank.org/files/24989_wps3007.pdf. Accessed June 2004.

Boarnet, M. and R. Crane (2001). The influence of land use on travel behavior: specification and estimation strategies. *Transportation Research Part A: Policy and Practice,* 35, *823-845.*

Ewing, R. (1997). Is Los Angeles Style Sprawl Desirable? *Journal of the American Planning Association*, 63: 107-126.

Levinson, D.M. and A. Kumar (1997). Density and the journey to work. *Growth and Change*, 28, 147-172.

Gomez-Ibañez, J. (1991). A Global View of Automobile Dependence (book review). *Journal of the American Planning Association,* 57: 342-346.

Gordon, P. and H.W. Richardson (1997). Are Compact Cities a Desirable Planning Goal? *Journal of the American Planning Association*, 63: 95-106.

Handy, S. (1996). Methodologies for exploring the link between urban form and travel behavior. *Transportation Research Part D: Transport and the Environment*, 35, *823-845*

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 E. E. Leamer and J.J. Heckman, eds., *Handbook of Econometrics, Vol. 5*, Elsevier, Amsterdam, Chapter 52.

Newman, W.G. and J.R. Kenworthy (1989a). Gasoline consumption and cities. *Journal of the American Planning Association,* 55: 24-35.

Newman, W.G. and J.R. Kenworthy (1989b). *A Global View of Automobile Dependence.* Brookfield, VT: Gower.

ORNL (2004). *The 2001 National Household Transportation Survey*. Oak Ridge National Laboratory for the U.S. Department of Transportation. http://nhts.ornl.gov/2001/index.shtml

Schipper, M.A. and D.S. Pinckney (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.

Steiner, R.L. (1994). Residential density and travel patterns: Review of the literature. *Transportation Research Record 1465*: 37-43.

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