How much can vehicle travel be reduced through land-use policies in California?: An application of a Modified Two Part Model with Instrumental Variables *

David R. Heres Del Valle
(PhD Candidate)

Department of Agricultural and Resource Economics
University of California, Davis

June 1, 2009

Abstract

The California Global Warming Solutions Act of 2006 requires year 2020 greenhouse gas (GHG) emissions in the state to be reduced back to 1990 levels. The recently approved Proposed Scoping Plan expects emissions reductions through land-use policies based on a 4% decrease in car travel demand by 2020. It is not clear, however, how large the increase in residential density must be in order to achieve such reductions. Furthermore, results from different strands of the literature are not yet conclusive about the magnitude of the impact of land-use variables on the amount of car travel and thus on GHG. This study aims to contribute to the ongoing debate by implementing a modified two-part model (M2PM) with instrumental variables (IV), a procedure that respectively takes into account the large mass of observations with zero car travel, and the possibility of residential self-selection, both of which could otherwise bias the estimates. The analysis takes advantage of a large dataset on travel patterns and socio-economic characteristics of more than 7,000 households across the 58 counties in the state of California. The study calculates the impacts of residential density and jobs/housing balance on vehicle miles traveled (VMT), and includes measures for the supply of public transportation in one set of the estimations. Results derived from our dataset and model specifications show that VMT elasticities

*This working paper is part of the author's dissertation research. Please do not cite without permission. The study has benefited from comments from Colin Cameron, Deb Niemeier, Colin Vance, and Jim Wilen. All remaining errors and omissions are solely the responsibility of the author.

with respect to residential density are considerably larger for the instrumented M2PM specifications than those from ordinary least squares, linear IV, or non-instrumented M2PM, revealing potential bias from estimations based on the latter approaches. Although our elasticities are larger than others found in the recent econometric literature, the impact of residential density on VMT would not be as large as previously suggested from other recent agency reports unless very large increases in residential density take place. On the other hand, recent estimates of the elasticity of VMT with respect to the price of gasoline imply that moderate increases in the price of gasoline would suffice to reduce travel by the required 4%. The comparisons suggest that although pricing measures and land-use policies should not necessarily be seen as mutually exclusive options, the former might be more effective in reducing the externalities derived from automobile use.

1 Introduction

In 2006, the state of California passed the Assembly Bill 32 - California Global Warming Solutions Act of 2006. The document, the first ambitiously addressing climate change at a large scale in the United States of America (US), requires year 2020 greenhouse gas (GHG) emissions in the state to be reduced back to 1990 levels. The requirement translates into an elimination of 173 metric tons of carbon dioxide equivalent (MMTCO2e) compared to the business-as-usual scenario. [1] In December of 2008, an scoping plan including the mitigation strategies that will attain the reductions was approved (CARB, 2008).

The reduction of greenhouse gases (GHG), the leading objective of climate policy, has a particularly complicated task in the transportation sector. In the case of California, this sector contributes with 38% of all GHG emissions (CARB, 2007b). Transportation emissions can mainly be reduced by reducing the carbon content of fuels, improving motor efficiency, and importantly, by reducing travel. Throughout the building process of an scoping plan, land-use policies to reduce GHG from the transportation sector stood out for its lack of rigor in the calculation of potential vehicle miles traveled (VMT) reductions. The recently approved Proposed Scoping Plan (CARB, 2008) expects reductions of 5 MMTCO2e through land-use policies which follow after assuming a drop by 4% in car travel demand by 2020. It is not clear, however, from CARB's report how large the increase in residential density must be in order to achieve the 4% reduction in VMT. [2]

[1]This figure is based on the preliminary projected 600 MMTCO2e for 2020 and the 427 MMTCO2e 1990 level officially adopted in December 2007 (CARB, 2007b).

[2]The 4% reduction is the median value of the impacts of combined land use and transit improvements over a 10-year period in the studies surveyed in Rodier (2008). The studies surveyed largely differ in their methodology, location and magnitudes of assumed improvements.

The estimated GHG reductions to be achieved through such policies in earlier drafts such as CAT (2007) were driven by presumed reductions of 10% to 30% in VMT. No further details were provided on such reductions in CAT's report, whereas ETAAC's (2008) 20%-30% estimate is based on JHK and Associates (1995) who compared mean VMT between communities or areas with different levels of residential density and other land-use characteristics (but which are also different in other respects that could be affecting the divergence.). Based on previous literature, Ewing et al. (2007), considered that a typical household in a compact development community would travel 30% vehicle miles less compared to a typical suburban household, while CEC's (2007) report on the links between land use and transportation also found similar figures.

Furthermore, in spite of the aforementioned impacts of land-use characteristics on VMT, there are a number of recent econometric studies that find little or no impact of land-use variables on the amount of travel and thus on GHG (Bento et al. (2005), Brownstone and Golob (2009), Fang (2008)). Our study aims to contribute to this debate by implementing a procedure that among other controls, takes into account the following two characteristics that could bias results otherwise: 1) large mass of observations with zero car travel, and 2) self-selection of the decision on where to locate given travel preferences. The analysis takes advantage of a large dataset on travel patterns and socio-economic characteristics of households (HH) in the 58 counties in the state of California. [3] This study considers two land-use-related variables (residential density, and jobs/housing balance), and includes a variable for the supply of public transportation in one set of the estimations. Some other variables such as road network design and density, population and employment centers, and availability of recreational areas, are relevant in terms of land-use policy, however these are commonly correlated with the two included in our study. Therefore, as noted in Brownstone and Golob (2009) residential density, the preferred land-use variable in this type of studies, should be interpreted as a proxy for the set of land-use characteristics that could have an impact on VMT.

Two-part models (2PM), originally motivated in Cragg (1971) and Duan et al. (1983) have been widely applied in health economics studies in order to correct for the bias resulting from a large mass of zero doctor visits observations (Dow and Norton, 2003). In these models, the first part estimates the probability of incurring in a certain activity (e.g. medical expenses, visits, or travel by car in our case). The second part of the model estimates the level of that activity conditional on its occurrence. When the dependent variable of the second part of the model is log-transformed to ensure its positivity, the presence of heteroskedasticity can severely bias the results (Duan et al. (1983), Manning and Mullahy (2001)). We follow one of the suggestions in Mullahy (1998) implementing a modified version of the 2PM approach. The modified 2PM (M2PM) ensures the positivity of the second part's outcome with a direct estimation of the

[3] Our final samples are more than twice as large as those used in previous studies for travel demand in California (Fang (2008); Brownstone and Golob (2009)).

dependent variable in levels (i.e., no log-transformation required).

On the other hand, residential self-selection, largely acknowledged in the literature is a particular case of the problem of endogeneity of regressors. Instrumental variables (IV), have been considered among a set of strategies aimed to correct for self-selection (Bhat and Guo (2007); Mokhtarian and Cao (2008)). This approach eliminates the *endogeneity* (i.e., correlation between the error term and the regressors) by finding so-called *instruments* that are correlated to the endogenous regressor (i.e., residential density or jobs/housing balance) but not correlated to the error term. The latter condition rules out the possibility of the instruments being regressors in the original equation; in other words, the instruments are variables that affect the dependent variable only indirectly through their impact on the endogenous regressor .

A number of important characteristics distinguish our study from other recent econometric studies. In contrast to the German study from Vance and Hedel (2007), that implemented an IV-2PM, ours also incorporates a modified version of the 2PM to avoid potential problems derived from heteroskedasticity. Both Fang (2008), and Brownstone and Golob (2009) base their estimations on the same California dataset. Estimates from the former do not address self-selection bias, while the latter identifies the coefficients of an structural equations model by choosing a specific recursive structure between residential choice, VMT and fuel consumption that relies on debatable exclusion restrictions.[4] Bento et al. (2005) include several city-wide measures of urban form in order to estimate the impacts of such variables on mode choice, travel demand and vehicle ownership in the US. Although the use of variables with a larger spatial scale diminishes the risks of them being correlated with the unobserved travel preferences, the impact of variability of such characteristics at smaller geographical delimitations would remain unknown. Nevertheless, we include a transit variable only available at the urban area level for one set of our specifications. [5]

At a glance, results derived from our dataset and model specifications show that VMT residential density elasticities are considerably larger for instrumented M2PM specifications than those from ordinary least squares, linear IV, or non-instrumented M2PM, revealing potential bias from estimations based on those approaches. Our elasticities are also larger than others found in the recent econometric literature. However, unless very large increases in residential density are enforced, the impact of residential density would not be as large as previously suggested from other recent reports such as CEC (2007), Ewing et al. (2007), CARB (2007a), and ETAAC (2008).

The next section extends on the sources and characteristics of our assembled dataset. The description of our estimation strategy is followed by the results for seven different specifications.

[4]The model in Brownstone and Golob implies that, for example, the number of children in a household affects the residential choice but not VMT

[5]Neither Fang (2008) nor Brownstone and Golob (2009) studies included transit variables.

Estimated impacts from changing land-use characteristics should be ultimately contrasted to other options such as fuel or travel taxation - a discussion section considering alternative policy paths precedes the conclusions.

2 Data

Data from several sources were used to build a dataset with the variables needed for our study. Travel data from the *2000-2001 California Statewide Household Travel Survey* were obtained from Caltrans. The survey included travel and socio-economic information from 17,040 households (HH) across 58 counties. Spatial information, available for most of the HH home addresses and their travel destinations, was exported to GIS software to calculate the distances traveled, and to retrieve the census tract, zip code tabulation area (ZCTA), and urban area where each HH was located. [6] Data on residential density and other characteristics at the census tract level were obtained from the Census 2000 Summary File 3. [7] The number of business establishments at the ZCTA level was obtained from the Census' County Business Patterns for the year 2000. Transit data, only available for urban areas, come from the National Transit Database of the Federal Transit Administration.

Our final sample (*finalall*) consists of 7,666 HH (of which 4,098 HH are located in urban areas - the *finalurb* sample). The largest loss in observations is due to the lack of complete information necessary to calculate travel distance. Other losses occurred due to the restriction of our analysis to weekday travel[8], the discarding of observations with inconsistent travel data (e.g., estimated speeds that do not match the reported travel mode), and observations dropped due to missing income data. Table A9 summarizes the causes leading to sample size reductions, and tables A7 and A8 respectively describe and summarize the variables used in in our estimations.

Descriptive statistics of the variables in A8 do not largely differ across the original, *finalall* and *finalurb* samples.[9] The median HH annual income falls in the category of $35,000 to $50,000 for the *finalall* and *finalurb* samples. In all samples, average HH size and number of vehicles are both about 2, while the average number of workers in a HH is roughly half its size. Across the

[6] Cartographic boundary files were obtained from the Census Bureau's TIGER geographic database. These files already included data on the area occupied by each tract and ZCTA.

[7] The number of housing units for each ZCTA was obtained from the Census' Gazetteer files.

[8] A small percentage (954 out of 17,040 HH) of the original sample was surveyed for two continuous weekday and weekend days.

[9] Table A8 shows the means of these variables for each sample. It is important to note that income categories for the original sample significantly differ from the two samples used in our study. This is so because households with income category equal to 9 (i.e., non-reported) were discarded in our final estimations. Also noticeable is the increase in the mean of residential density from the *finalall* to the *finalurb* samples. This is not surprising since urban areas are by definition more densely populated than rural ones (in terms of residential density, our original sample shows 2010.7 and 286.4 housing units per square mile respectively).

three samples, age of the oldest member of the HH is in average between 51 and 55 years, and 42% of the HH have a member with an education level higher or equal to a college degree. The average residential density is 1518, 1570, and 2491 housing units per square mile for the *original*, *finalall*, and *finalurb* samples respectively. In the *finalall* sample, residential density attains a minimum of 0.11 in a census tract in Inyo County, and a maximum of 66,173 corresponding to a census tract within the boundaries of San Francisco County. Furthermore, 72% of the HH in the *original* and *finalall* samples are located in urban areas.[10]

The number of business establishments is roughly 6% of the number of housing units across the three samples. In the *finalurb* sample, each urbanized area is served in average by 2759 route miles, and average transit density is nearly 6 miles. Importantly, on average, only 8% of units were built before 1940 with a maximum of 88% in an Alameda County census tract and, on average, 26% of the state's population is other than white (with a maximum of 98% and a minimum of 2% in census tracts in Los Angeles County and Riverside, respectively). About 25% of the HH in the final samples reported no car-trips in the surveyed day (i.e., dr=0.75), which motivates the implementation of estimation techniques addressing skewness and non-linearity associated with the dependent variable (*vmt*).

Means of the regional dummies show the number of HH from a specific region as a percentage of the total sample. The shares are practically identically between the *original* and *finalall* samples with the SCAG (20%), Rural (14%), and MTC (9%) regions being the most predominant, and each of the other 14 regions representing between 3% and 7% of the observations. However, in the *finalurb* sample, no observations fall within the rural regions 1, 6, and 9, drastically changing the spatial distribution of the sample. Nevertheless, our estimations do not require sample weighting measures since our interest is in finding structural-causal relationships, rather than in describing correlations between population variables. When the objective is the former and there is no endogenous stratification, the use of sample weights to obtain consistent estimates is no longer necessary (Cameron and Triverdi, 2005).

Figure 1 plots the mean daily HH VMT at different levels of residential density, mix, and transit.[11] The figure suggests that HH located in census tracts with more housing units per square mile tend to drive less.[12] The trajectories for the lines showing the relationship between VMT and *mix* and *transit* are not as smooth as that for residential density. Moreover, an expected negative correlation between *transit* and *vmt* is not clear from Figure 1. The following

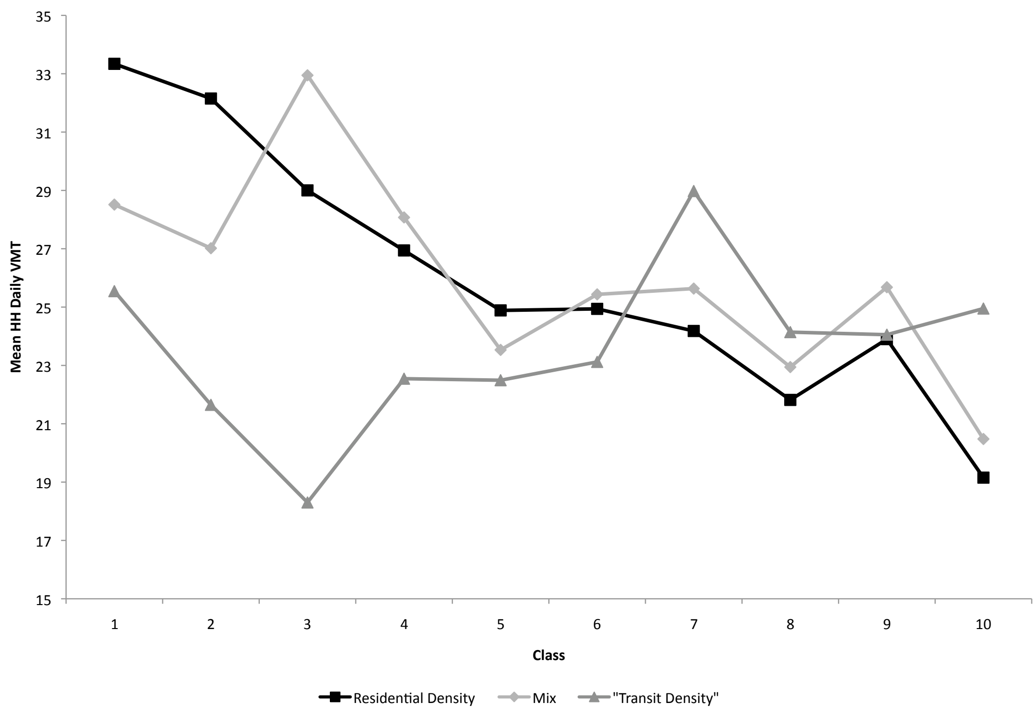
[10]Both *urbanized areas* and *urban clusters* are classified as *urban areas* according to the Census Bureau. As mentioned, transit data were only available for *urbanized areas* (i.e., neither *rural areas* nor *urban clusters* are included in the *finalurb* sample).

[11]The classification in horizontal axis is based on the percentiles for residential density, mix, and transit in our finalall sample (finalurb for transit).

[12]We also observe a lower mean daily HH VMT for HH located in *urban areas* (24.6) than those located in census tracts classified as *rural* (29.9).

section describes the methodology we adopted to obtain a more precise estimate of the impact of these land-use variables on *vmt*.

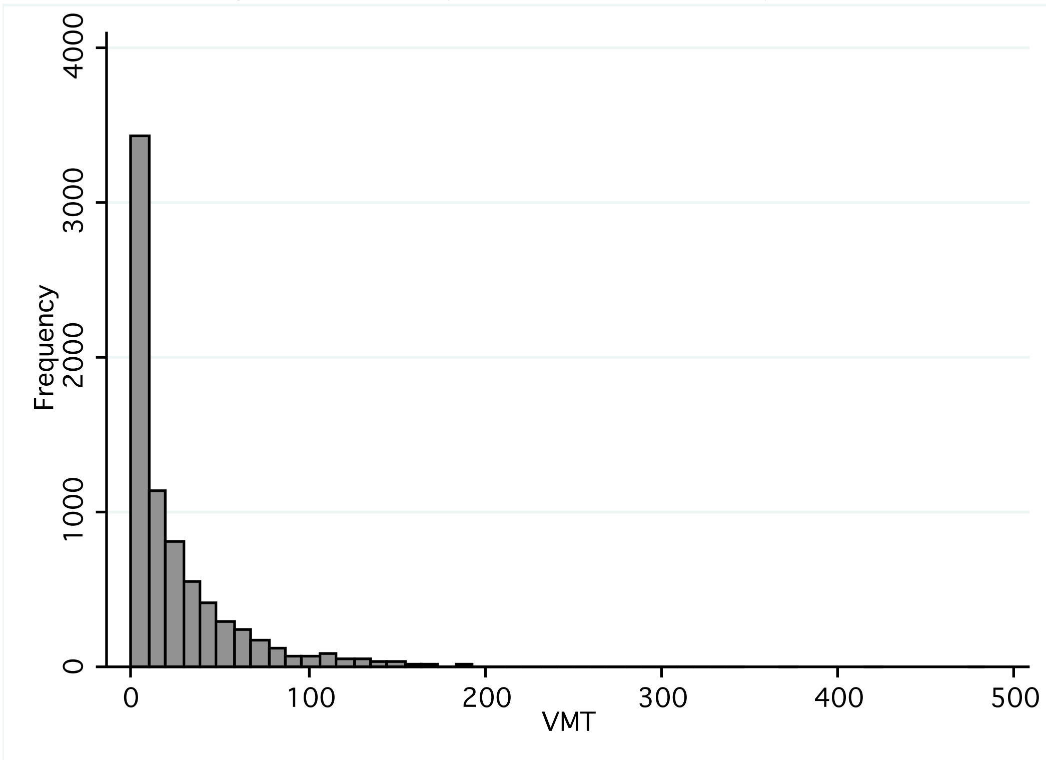
Figure 1: Mean daily VMT per HH for different residential density classes



3 Estimation Strategy

Figures 1 suggests that some correlation exists between residential density, land-use mix, and transit with HH VMT. However, in order to generate consistent causal statements between the land-use variables and *vmt*, we need to consider that 1) other socio-economic variables might affect the level of VMT that a HH chooses, 2) there exists a mass of zero-values as shown in Figure 2, which can result in biased estimates if regular ordinary least squares (OLS) are performed, and 3) residential location self-selection, largely acknowledged in the literature (Bhat and Guo (2007), Mokhtarian and Cao (2008)), could derive in biased estimates of the impact of *resdes* and *mix* on *vmt*.

Figure 2: Frequency distribution of HH daily VMT



We address problems 1 and 2 *via* the estimation of a multivariate two-part model. This approach has been widely implemented to estimate health expenditures (Dow and Norton, 2003), and Vance and Hedel (2007) applied this model to estimate the impact of urban form on VMT in Germany. Finally, in order to correct for potential self-selection bias, we employ instrumental variables. [13] Our selection of instruments (variables that explain *resdes* and *mix*, but not *vmt*) follows suggestions from previous studies (Boarnet and Sarmiento (1998); Vance and Hedel (2007)).

[13]Mokhtarian and Cao (2008) review a number of estimation strategies, including instrumental variables, in presence of self-selection in the context of residential location and travel behavior.

3.1 The Two-Part Model (2PM)

Problems arising from a large mass of zeroes in the dependent variable as shown in Figure 2 are similar to those from data censoring (i.e. inconsistent-biased estimates due to non-linearity in the true model). Originally, the 2PM arised as a relaxation to some of the assumptions in the *tobit* model (Cragg, 1971). Another estimator used in similar circumstances is the *heckit*, however, as noted by Duan et al. (1983), Dow and Norton (2003) and Vance and Hedel (2007), the 2PM might be more appropriate if interest lies in the actual outcome. This is, zero *vmt* observations are treated as real zeroes instead of as latent variables. Although the actual outcome can always be recovered from the *tobit* and *heckit* models, and thus a correct interpretation is possible, it involves further calculations and assumptions.[14] The 2PM incorporates the probability of observing *vmt*>0 yielding the following expected value of *vmt* for any of the n observations:

$$E[vmt|x] = Pr[vmt > 0|x] \times E[vmt|vmt > 0, x] \tag{1}$$

where x is a column vector that includes the k explanatory variables for a given observation.

The first part of the 2PM can be derived from underlying economic choice. An index function model or a random utility model would yield the following probability of choosing to drive (i.e. vmt>0 or dr=1):

$$Pr[dr = 1|x] = F(x'\beta_1) \tag{2}$$

where β_2 is a column vector with the k coefficients, and F is the cumulative density function (cdf) of the error portion of an index function (or the cdf of the difference of the errors from the random utilities associated with driving and not driving).

The result in equation (2) is valid for densities symmetric around zero as it is the case for the standard normal. The probit model is the natural candidate to estimate (2) and we assume F to be the cdf of the standard normal (Φ).

In most of studies implementing two-part models, the second part estimates a log-linear model in order to ensure positive values of the dependent variable. Vance and Hedel (2007), however, run this regression in levels (i.e., *vmt* as the dependent variable instead of ln(*vmt*)), probably due to the potential problems related to the retransformation from log to levels described in Manning and Mullahy (2001). In particular, if the errors are not homoskedastic-normally distributed, the retransformation $E[y|y > 0, x] = exp(x'\beta_2 + 0.5\sigma^2)$ from $E[ln(y)|y > 0, x] = x'\beta_2$ is no longer valid. A column vector with the k coefficients from the second part of the 2PM is represented by β_2. Solutions for failing the non-normality assumption such as Duan's (1983) *smearing estimator* are simple, however, if the errors are heteroskedastic, further steps and

[14]Duan et al. (1983), Dow and Norton (2003), andVance and Hedel (2007) discuss this choice in further detail.

assumptions are necessary.

Given the aforementioned considerations, we opted for an estimation framework that directly estimates the conditional mean of *vmt* without compromising the positivity of the outcome. Mullahy's (1998) Modified Two-Part Model (M2PM) specifies the second part as an exponential conditional mean (ECM): $E[vmt|dr = 1, x] = exp(x'\beta_2)$. It follows that the expected value of *vmt* for the whole sample (zeroes and positive *vmt*) is:

$$E[vmt|x] = \Phi(x'\beta_1)exp(x'\beta_2) \tag{3}$$

Marginal effects in the M2PM are not readily available from the coefficients as in the OLS case. For continuous variables we instead need to derive them through the following formula which is the partial derivative of $E[vmt|x]$ in equation (3) with respect to the variable of interest x_k. [15]

$$\frac{\partial E[vmt|x]}{\partial x_k} = \beta_{2k}\Phi(x'\beta_1)exp(x'\beta_2) + \beta_{1k}\phi(x'\beta_1)exp(x'\beta_2) \tag{4}$$

Where ϕ is the standard normal density (i.e. the derivative of the normal cdf (Φ)). For indicative variables the following formula provides the appropriate marginal effects. In (5), subscripts in the x vector indicate whether the indicative variable of interest (x_d) is evaluated at zero (x_0) or at one (x_1).

$$\frac{\partial E[vmt|x]}{\partial x_d} = \Phi(x_1'\beta_1)exp(x_1'\beta_2) - \Phi(x_0'\beta_1)exp(x_0'\beta_2) \tag{5}$$

Interest often lies in the elasticity of *vmt* with respect to x_k. In such cases the following is calculated:

$$\frac{\partial E[vmt|x]}{\partial x_k} \times \frac{x_k}{E[vmt|x]} = \left(\beta_{2k} + \beta_{1k}\frac{\phi(x'\beta_1)}{\Phi(x'\beta_1)}\right)x_k \tag{6}$$

3.2 Self-selection and the endogeneity of regressors

A common estimation problem found when estimating the relationship between travel decisions and the built environment is that unobserved travel preferences might lead to choose a particular residential location in the first place (*self-selection*). This is, lower levels of car travel in high residential density areas (or areas with more businesses) could be the result of HH residential

[15] If disturbances were not homoskedastic and we were to run a log-lin model, equations (4) and (6) would need also to include the partial derivative of σ^2 with respect to x_k, such that for instance, (4) would become: $\partial E[vmt|x]/\partial x_k = (\beta_{2k} + .5\partial\sigma^2(x)/\partial x_k)\Phi(x'\beta_1)exp(x'\beta_2 + .5\sigma^2(x)) + \beta_{1k}\phi(x'\beta_1)exp(x'\beta_2 + .5\sigma^2(x))$. With the M2PM we avoid the specification of the heteroskedastic disturbances, however we incur in further calculation costs due to the non-linearity of the estimators.

location choice given prior travel preferences. [16] When this is the case we cannot distinguish between the self-selection and the built environment induced effects. Self-selection is a special case of the endogeneity problem that violates the OLS assumption of explanatory variables being uncorrelated with the error term, yielding biased estimators except for those for variables not related to the unobservable (Cameron and Triverdi, 2005).

Our approach to correct for self-selection follows that from Boarnet and Sarmiento (1998) and Vance and Hedel (2007), introducing instrumental variables (IV) to purge the estimates from the effect of unobserved travel preferences on *resdes* (*mix*). In order to implement an IV regression we must find at least as many exogenous variables explaining the endogenous regressors but not *vmt*, as endogenous regressors we have. [17] Potential valid instruments for *resdes* and *mix* are: the percentage of units built before 1939 (*pre40*), the percentage of population other than white (*nonwhite*), and the percentage of family HH (*famHH*). [18] These variables are a) likely to be correlated with *resdes* (*mix*), but b) unlikely to affect travel behavior except through the indirect impact that those variables have on *resdes* (*mix*). Formal statistical tests are presented in the following section as a check for their validity.

4 Results

Table 1 shows marginal effects (*mfx*) and statistical tests for the seven Instrumental Variables M2PM (IVM2PM) specifications. Models (1)-(3) are based on the *finalall* sample, whereas models (4)-(7) include the transit variable from the only-urban sample. Most of the coefficients are statistically significant at the 5% level. Kleibergen-Paap (KP) tests the relevancy of the instruments, the null hypothesis being that the instruments do not explain the endogenous regressors. While a rejection of the KP test lends support to the set of instruments used, the overidentification (OID) test supports the validity of the instruments whenever we fail to reject the null. The specific OID tests, are the Amemiya-Lee-Newey (ALN) minimum χ^2 for the probit and the Hansen's J for the ECM model. [19]

Model (1) is our preferred model among those with no transit data. Although *mix* is statistically significant in (2), it no longer is when the model includes also *resdes*. Furthermore, the validity of the instruments is rejected in (2). Along the same lines, we do not consider

[16] We treat *transit* as exogenous given its large geographical scale.

[17] More formally, the instruments should be correlated to the endogenous regressor but not to the error term, which rules out the possibility of the instruments being regressors in the original equation.

[18] These variables were also obtained at the census tract level from Census 2000.

[19] KP, ALN, and Hansen's J tests statistics were calculated with the *overid* STATA routine (Baum et al., 2006). OID tests are aimed to test the validity of the instruments (i.e., not correlated with the error term from equation (2) or from the ECM model). Hence, the two statistical conditions for any set of variables to be instruments for the endogenous regressors, relevancy and validity, are tested through the KP and OID tests respectively.

Table 1: Estimation summary for seven IVM2PM specifications

	Marginal Effects						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
resdes	-0.00278**	-	-0.00283**	-0.00119**	-	-0.00138**	-
mix	-	-175.452**	20.57788	-	-70.06921	49.00128	-
trans	-	-	-	-0.31928	-0.54876**	-0.30185	-0.58528
age	-0.07879**	-0.06149**	-0.08089**	-0.07048**	-0.05463	-0.07328**	-0.05344*
work	10.92029**	10.74677**	10.83011**	9.30792**	8.96656**	9.21359**	8.76520**
hhsize	4.13738**	4.09305**	4.15348**	3.85420**	3.80152**	3.94764**	3.95993**
nveh	3.79045**	3.99964**	3.80579**	4.24787**	4.53670**	4.27705**	4.70518**
eduh	4.00194**	4.15600**	3.97911**	2.90637**	2.80409**	2.86086**	2.74366**
inc2	0.73750	0.65183	0.886171	0.83835	0.65542	1.18373	1.07018
inc3	4.45611**	4.60448**	4.55719**	2.47562	2.54854	2.69995	2.79421
inc4	8.94348**	9.05875**	9.02111**	8.17560**	8.16926**	8.42505**	8.49294**
inc5	12.10060**	12.66414**	12.20337**	10.45821**	10.83332**	10.62859**	11.05363**
inc6	14.52842**	15.61446**	14.56010**	12.52687**	13.28175**	12.55567**	13.42140**
inc7	18.20045**	20.21815**	18.17801**	15.64128**	17.00517**	15.43915**	16.80407**
inc8	12.72168**	15.81087**	12.35778**	10.66692**	12.73201**	9.81779**	11.86821**
ur	-0.81164	-2.60979**	-0.97036	-	-	-	-
Observations	7666	7666	7666	4098	4098	4098	4098
KP LM (I)	0.0000**	0.0000**	0.0000**	0.0000**	0.0000**	0.0000**	-
KP LM (II)	0.0000**	0.0000**	0.0000**	0.0000**	0.0000**	0.0000**	-
ALN min χ^2 *(I)*	0.1288	0.0000**	0.4077	0.1907	0.0000**	0.5193	-
Hansen's J (II)	0.3581	0.0000**	0.1845	0.9803	0.1107	0.9772	-
$Pseudo - R^2$ *(I)*	0.1957	0.1787	0.1839	0.1885	0.1781	0.1669	0.1905
$Pseudo - R^2$ *(II)*	0.1181	0.1247	0.1219	0.1198	0.1255	0.1179	0.1211

*Statistically significant at the 10% level, **significant at the 5% level. Standard errors for *mfx* calculated through 299 bootstrap replications. Regional dummy variables not shown in the table (reg1 was dropped to avoid matrix singularity). Models (4), (5), (6), and (7) have all observations with *ur*=1 and do not have observations for *reg1*, *reg6* and *reg9*. $Pseudo - R^2$, and p-values for KP LM, Amemiya-Lee-Newey (ALN) minimum-χ^2, and Hansen's J tests are reported for the Probit (I) and the ECM (II) models. Model (7) has no endogenous regressors. KP rank LM test, and Hansen's J (ALM for (I)) test are χ^2-distributed with (L-K+1) and (L-K) degrees of freedom respectively. Where L is the number of excluded instruments and K is the number of endogenous regressors.

either models (5) or (6) as the best. Residential density remains statistically significant across the different specifications and for that reason we report detailed estimations for (4), rather than (7) which includes transit as the only land-use variable. Finally, it can be seen that the $pseudo-R^2$ does not show much variation between the models. [20]

Tables 2 and 3 respectively show the results for our preferred models with and without public transportation data. Marginal effects were computed at the mean values of the regressors (i.e., $x = \bar{x}$) using the formulas in equation (4) and (5) for continuous and indicative binary variables respectively. The reported standard errors (in parenthesis) were obtained through bootstrap replications. Not reported here, average marginal effects (the sum of the *mfx* for each observation divided by n) yielded comparable results to those at the mean values.

4.1 Results for the *finalall* sample (no transit data)

Table 2 reports the first and second parts of the IVM2PM, as well as the marginal effects (*mfx*) of each variable on *vmt* in model (1). This model, which includes *resdes* as the only land-use variable, is our preferred model for the *finalall* sample.

All *mfx* on *vmt* are statistically significant and coefficients and *mfx* have the expected signs in all of the specifications (including the coefficients at each of the two parts of the IVM2PM). Higher levels of education in the HH, HH size, number of vehicles, income and percentage of workers in the HH, are all positively correlated with *vmt*. On the other hand, the age of the oldest member of the HH, urban HH, and the residential density of the census tract where the HH is located are all negatively correlated with *vmt*.

4.2 Results for the *finalurb* sample (with transit data)

The signs of the *mfx* for this model that includes transit are the same as those from table 2. One difference is that the *mfx* for *inc3* is also not statistically significant for this model. The coefficient for *eduh* in the probit part is not statistically significant and its *mfx* is lower if compared to that from table 3. Moreover, *mfx* as well as the first and second part coefficients for *transit* are all not statistically significant. The *mfx* of *resdes* is lower compared to that obtained in model (1), and its coefficient for the second part is only statistically significant at the 10% level. In fact, not considering the regional dummies, only the coefficients for *work*, *hhsize*, *nveh*, *eduh*, and the constant are statistically significant at the 5% level in the second part of this specification. The latter finding suggests that in urban areas, age, income levels, and residential

[20] Several goodness-of-fit measures have been proposed for non-linear estimators. So-called $Pseudo-R^2$ is one class. For both parts of the IVM2PM we use $R^2 = 1 - RSS/TSS$, where RSS is the residual sum of squares, and TSS is the total sum of squares. Amemiya (1981) refers to this measure as *Efron's* R^2 in the probit case. Other measures based on log-likelihood values cannot be computed from our results obtained thorugh STATA Newey's twostep *ivprobit*. STATA and Mata codes for all the calculations in the study are available upon request from the author.

Table 2: IVM2PM estimates (*finalall* sample)

	(1) IVM2PM					
variable	*I. dr=1,0*		*II. vmt(vmt>0)*		*vmt (vmt>=0)*	
	coeff	std. err.	coeff	std. err.	mfx	std. err.
resdes	-0.00009	0.00001**	-0.00009	0.00002**	-0.00278	0.00048**
age	-0.00511	0.00122**	-0.00161	0.00114	-0.07879	0.02679**
work	0.66599	0.05168**	0.23810	0.05114**	10.92029	1.16847**
hhsize	0.16332	0.01885**	0.12190	0.01372**	4.13738	0.33464**
nveh	0.18257	0.02252**	0.09995	0.01561**	3.79045	0.40112**
eduh	0.17120	0.04062**	0.11181	0.03542**	4.00194	0.89306**
inc2	0.31629	0.07389**	-0.13589	0.12632	0.73750	2.02374
inc3	0.58796	0.07781**	-0.04040	0.12405	4.45611	2.02300**
inc4	0.77544	0.08032**	0.10559	0.12531	8.94348	2.09714**
inc5	0.77565	0.08067**	0.22965	0.12471*	12.10060	2.10993**
inc6	0.77254	0.09416**	0.31653	0.12774**	14.52842	2.30103**
inc7	0.85291	0.10556**	0.41141	0.13015**	18.20045	2.43595**
inc8	0.55279	0.12638**	0.32775	0.13906**	12.72168	2.86719**
ur	0.20613	0.04764**	-0.11264	0.04250**	-0.81164	1.16229
cons	-0.97089	0.13290**	3.16144	0.16280**	-	-
KP LM	361.120	(0.0000)**	291.98	(0.0000)**	-	-
OID tests	4.100	(0.1288)	2.0537	(0.3581)	-	-
$Pseudo-R^2$	0.1957		0.1181		-	
Observations	7666		5796		7666	

*Statistically significant at the 10% level, **significant at the 5% level. Heteroskedasticity-robust standard errors reported for I and II. Standard errors for *mfx* calculated through 299 bootstrap replications. Regional dummy variables not shown in the table (*reg1* dropped to avoid matrix singularity). KP rank LM test, and OID tests are χ^2-dstributed with (L-K+1) and (L-K) degrees of freedom respectively (p-value in parenthesis). Where L is the number of excluded instruments and K is the number of endogenous regressors. The OID for I is the Amemiya-Lee-Newey minimum χ^2, while Hansen's J test is reported for II.

density do not affect the amount of vehicle travel once the decision to drive has been taken. As previously shown in table 1, both KP and OID tests provide confidence about the validity and relevance of the instruments.

4.3 Elasticities for all the models

Tables 1, 2, and 3, showed coefficients and *mfx* for the explanatory variables under different model specifications, however, of particular interest are the elasticities of *vmt* with respect to the land-use variables. These are calculated with the formula in equation (6) and reported in Table 4

Table 3: IVM2PM estimates (*finalurb* sample)

	(4) IVM2PM					
variable	*I. dr=1,0*		*II. vmt(vmt>0)*		*vmt (vmt>=0)*	
	coeff	std. err.	coeff	std. err.	mfx	std. err.
resdes	-0.00007	0.00001**	-0.00003	0.00002*	-0.00119	0.00038**
age	-0.00358	0.00161**	-0.00207	0.00145	-0.07048	0.03299**
work	0.64338	0.07046**	0.21456	0.06423**	9.30792	1.42643**
hhsize	0.17695	0.02600**	0.11998	0.01730**	3.85420	0.40045**
nveh	0.19180	0.03257**	0.13336	0.02360**	4.24787	0.59355**
eduh	0.05693	0.05748	0.11589	0.04801**	2.90637	1.26606**
inc2	0.27681	0.10990**	-0.10143	0.21295	0.83835	3.1751
inc3	0.50854	0.11302**	-0.09985	0.20801	2.47562	3.2073
inc4	0.77549	0.11668**	0.10844	0.20826	8.17560	3.35179**
inc5	0.76784	0.11651**	0.20886	0.20899	10.45821	3.26664**
inc6	0.78684	0.13292**	0.28487	0.21282	12.52687	3.34738**
inc7	0.81283	0.14331**	0.38898	0.21186*	15.64128	3.67576**
inc8	0.41498	0.16567**	0.34372	0.22029	10.66692	3.78058**
transit	-0.00514	0.01338	-0.01325	0.00946	-0.31928	0.2125
cons	-0.34642	0.20951**	2.75898	0.23571**	-	-
KP LM	275.750	(0.0000)**	216.4	(0.0000)**	-	-
OID tests	3.314	(0.1907)	0.039	(0.9803)	-	-
$Pseudo-R^2$	0.1885		0.1198		-	
Observations	4098		3130		4098	

*Statistically significant at the 10% level, **significant at the 5% level. Heteroskedasticity-robust standard errors reported for I and II. Standard errors for *mfx* calculated through 299 bootstrap replications. Regional dummy variables not shown in the table (*reg2* dropped to avoid matrix singularity). KP rank LM test, and OID tests are χ^2-dstributed with (L-K+1) and (L-K) degrees of freedom respectively (p-value in parenthesis). Where L is the number of excluded instruments and K is the number of endogenous regressors. The OID for I is the Amemiya-Lee-Newey minimum χ^2, while Hansen's J test is reported for II.

for each of the seven specifications. It is important to note that the *mfx* and elasticities for *resdes* under the IVM2PM models are considerably larger than the ones for the non-instrumented, and non-zero corrected specifications, potentially revealing the biasness of estimates from studies not correcting for these characteristics. [21] The elasticity for *mix* is also negative and seems to be larger than that for *resdes*, however, as mentioned before, the former is statistically significant only in model (2). The elasticity of *vmt* with respect to transit is also negative but only statistically significant in models where it is not combined with *resdes*. From this point, we

[21]Elasticities from the IVM2PM are about twice as large as those from OLS. Outputs for models not reported are available upon request from the author.

focus our attention to the impact of *resdes*, which is the variable showing statistically significant coefficients whenever it is included, and for which the instruments are valid and relevant as shown in tables 1 and 2.

Table 4: Elasticities

	Model						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
resdes	-0.1898 (0.0346)**	-	-0.1934 (0.0408)**	-0.1399 (0.0454)**	-	-0.1622 (0.0507)**	-
mix	-	-0.4442 (0.1099)**	0.0524 (0.1328)	-	-0.2118 (0.1341)	0.1486 (0.1515)	-
trans	-	-	-	-0.08611 (0.0575)	-0.1476 (0.0564)**	-0.0814 (0.0580)	-0.1569 (0.0562)**

*Statistically significant at the 10% level, **significant at the 5% level. Standard errors calculated through 299 bootstrap replications in parenthesis. See Table 1 for further statistics for each model specification.

Our results imply that, everything else equal, a 10% in residential density would reduce *vmt* by 1.9% (1.4% in the urban sample with transit data). This elasticity is larger than the reported in previous econometric studies for the US and specifically for California. However, the magnitude of this impact is still low considering reasonable ranges for policies aimed to increase residential density.[22] For instance, an intensive policy aimed to increase residential density by 25% in San Joaquin county would reduce daily HH *vmt* from 34.2 to 32.6. Similarly, in Sacramento, the reduction under the same policy would be of about 1 daily vehicle mile, from 24.5 to 23.4. Clearly, annualizing and adding up households increases the figure, however this type of policies should ultimately be contrasted to others potentially more effective and less invasive.[23]

Fang (2008) suggests that specifications that do not incorporate mechanisms for correcting the endogeneity of residential density provide an upper bound for the magnitude of its impact on VMT. However, as shown by our results and previous studies (Boarnet and Sarmiento (1998); Vance and Hedel (2007)), IV specifications yield a larger (in magnitude) impact of residential density on VMT. Importantly, aside from the logic of their choice, the selection of instruments appears to suffice statistical tests of their relevancy and validity.

The fact that not all the variation in residential density is used when IV are implemented

[22] This figure is considerably larger than the implied -0.12 elasticity in Brownstone and Golob (2009). In line with our non-instrumented estimations, Fang's (2008) elasticity, which does not directly address self-selection is only -0.024 for cars.

[23] Although there is recent evidence of a reduction in the short-run price elasticity of gasoline demand (Small and Van Dender (2007); Hughes et al. (2008)), the impact of gasoline price changes on VMT could still be important with increases in the tax on gas. The discussion in the next section extends on this.

might as well be playing a role in this result (i.e. the problem of heterogenous responses across the sample (Angrist and Krueger, 2001)). However, aside from the KP tests suggesting the relevance of the instruments, OLS regression of *resdes* on the set of instruments has a particularly large R^2 (.46 for *finalall* and .42 in finalur) indicating that not much variation in *resdes* is lost. Measurement error in the endogenous variable is a plausible cause for the inflated coefficient if we consider *resdes* as a proxy for a set of land-use characteristics affecting VMT. As mentioned in Vance and Hedel (2007), the higher in magnitude coefficients for *resdes* that result under IV are consistent with corrections of the attenuation bias. Further work would be required to identify the exact causes for the higher negative impact of residential density on VMT when IV estimation is implemented.

5 Discussion

In California, average gasoline prices reached a 5-year minimum of \$1.73 per gallon in the second week of December 2008. This occurred after hitting historical peaks of \$4.58 per gallon six months earlier and averaging \$3.29 in the last two years. [24] The demand for gasoline and for travel, however, have shown to be rather irresponsive to changes in gasoline price. Moreover, recent studies show that that the responsiveness has been reduced in later years (Small and Van Dender (2007); Hughes et al. (2008)). Nevertheless, as shown in table 5, pricing policies such as gasoline taxes might be more effective in reducing VMT. [25]

According to Parry and Small (2005), and implicitly in Small and Van Dender (2007) and Brons et al. (2008), travel adjustments represent between 30% and 70% of the total impact of gasoline price on gasoline demand. Hughes et al. (2008) estimate short-run price elasticities of gasoline demand obtaining a range of -0.033 to -0.077 similar to those values from Small and Van Dender (2007). [26] The latter also estimates long-run gas price elasticities for both gasoline demand and VMT. The VMT gas price elasticities for Hughes et al. (2008) reported in table 5 are calculated considering that the impact of gasoline price on VMT is, as in Small and Van Dender (2007), 32% of that on gasoline demand. Midpoints (MP), upper bound (UP) and lower bound (LO) estimates for short-run and long-run VMT gas price elasticities are presented in table 5 for four recent studies, clearly showing a larger impact in the long run. Given the

[24] Average prices for the US show similar trajectories but with slightly lower prices at each point in time (\$1.61, \$3.97, and \$2.83 respectively.)

[25] A somewhat related and more direct intervention would be *pay-as-you-drive* schemes.

[26] As part of their optimal gas tax calculations for California, Lin and Prince (2009) used an estimated -0.065 short-run VMT elasticity with respect to the price of gasoline. This estimate is lower than those reported in Brons et al. (2008) but larger than those from Small and Van Dender (2007), and Hughes et al. (2008). However the model in Lin and Prince (2009) did not control for potential recent shifts in the relationship between gasoline price and demand for travel.

nature of land-use changes, relevant comparisons of the impacts of residential density to those from gasoline price on VMT require long-run estimates of the impact of the latter.

Table 5: VMT gasoline price elasticities

	VMT gas price elasticities					
	Short Run			Long Run		
	MP	LO	UP	MP	LO	UP
Brons et al. (2008)	-0.1100	-0.1000	-0.1200	-0.2950	-0.2900	-0.3000
Hughes et al. (2008)	-0.0176	-0.0106	-0.0246	NA	NA	NA
Parry and Small (2005)	NA	NA	NA	-0.2200	-0.0600	-0.5400
Small and Van Dender (2007)	-0.0330	-0.0216	-0.0452	-0.1500	-0.1066	-0.2200

The studies from Hughes et al. (2008) and Small and Van Dender (2007) estimate models using US data. We only consider the estimates from the Three Stage Least Squares specifications of the latter (i.e., OLS not considered). Parry and Small (2005) consider values from the existing literature, while Brons et al. (2008) estimate elasticities from a meta-analysis of existing worldwide estimates. Our upper and lower bounds for the latter study consider both their own and those estimates they report from previous studies.

Table 6: Alternative policies for a 4% reduction in VMT

	MP	LO	UP
Residential density increase	24.6%	20.7%	28.6%
Gasoline price increase	20.1%	13.6%	26.7%

Percentages are based on elasticities from tables 4 and 5

Table 6 shows MP, UP, and LO estimates of the required (separate) increases in the price of gasoline and residential density that would induce the 4% VMT reductions from land-use assumed in CARB (2008). [27] For residential density increases we use the smallest and largest in magnitude estimates from table 4. Upper bound increases from both policy instruments are similar, however the lower bound increase required for gas price is only two thirds of that for residential density. Midpoint estimates from table 6 show that independent increases of 24.6% or 20.1% would be respectively needed in residential density or gas price to achieve the 4% VMT reduction by 2020. [28] Given the similar sizes of the impacts of each policy, technical and political considerations could become more relevant in the design of GHG mitigation strategies.

[27] Lower (upper) bound for gas price increases are calculated with the maximum (minimum) of the studies' long-run midpoints in table 5.

[28] Among the long-run estimates in Table 5, only Small and Van Dender (2007) considered recent shifts in elasticities (*via* the utilization of 1997-2001 averages for one set of their calculations). In order to generate another long-run estimate accounting for recent shifts in the size of the impact, consider multiplying the midpoint short-run estimates in Hughes et al. (2008) by 4.5; which is roughly the proportion of long-run to short-run estimates in Small and Van Dender (2007). This calculation derives in a lower bound long-run VMT elasticity estimate with respect to

Finally, one must also acknowledge that urban form and public transportation policies could facilitate transition to a less intensive use of the car in the presence of gas or travel pricing policies. In the end, although pricing measures and land-use policies should not necessarily be seen as mutually exclusive options, the impact of each should be individually assessed.

6 Conclusions

The results of this study are based on estimates derived from an econometric procedure that takes into account the large mass of zero car-travel observations and residential self-selection. A M2PM with IV was implemented to analyze data from more than 7,000 HH in California. The results from a number of specifications showed that the elasticities of VMT with respect to residential density are considerably larger for instrumented M2PM specifications than those from models that consider none or only one of the two correction mechanisms, revealing potential bias from estimations based on those approaches. Our elasticities are also larger than others found in the recent econometric literature. However, unless very large increases in residential density are enforced, the impact of residential density would not be as large as previously suggested from other recent agency-based reports. Our results imply that, everything else equal, a 10% increase in residential density would reduce VMT by roughly 2%.

The importance of reducing travel not only lies on its potential to reduce GHG, but also on the impact that this would have in terms of improving local air quality, and road traffic and safety. Estimated externalities from such car travel side-effects can be large and optimal taxes for the US have been estimated at about twice its current size (Parry and Small, 2005). Although the demand for travel is inelastic to gasoline price changes, it could prove to be more feasible and effective to introduce pricing policies coping with the externalities derived from suboptimal car travel. Our estimates indicate that the 4% reduction in VMT to achieve the GHG reductions from land-use policies in California would require increasing residential density by almost 25%. On the other hand, even though recent gasoline (and thus travel) demand price elasticity estimates show a decline, they imply that a 20% increase on top of the price of gasoline would suffice to reduce travel by those same amounts. Importantly, even though pricing policies might seem more suitable than modifying the built environment to reduce travel, one should not rule out the latter group of strategies on these grounds, since they could facilitate the transition to new travel behavior in a world with higher gasoline prices.

gas price of only -0.08 that would shift the upper bound gas increase requirement in Table 6 to 50%. On the other hand, a very small VMT residential density elasticity such as that from Fang (2008) would require residential density increases of 133% to achieve a 4% reduction in VMT. If we consider the -0.12 % elasticity in Brownstone and Golob (2009) the same target would require a policy enforcing a 33% increase in residential density.

References

Amemiya, T. "Qualitative response models: A survey." *Journal of Economic Literature* 1483–1536.

Angrist, J.D., and A.B. Krueger. "Instrumental variables and the search for identification: from supply and demand to natural experiments." *Journal of Economic Perspectives* 15, 4: (2001) 69–85.

Baum, C.F., M.E. Schaffer, S. Stillman, and V. Wiggins. "overid: Stata module to calculate tests of overidentifying restrictions after ivreg, ivreg2, ivprobit, ivtobit, reg3.", 2006. URL `http://ideas.repec.org/c/boc/bocode/s396802.html`.

Bento, A., M. Cropper, A. Mushfiq, and K. Vinha. "The effects of urban spatial structure on travel demand in the United States." *The Review of Economics and Statistics* 87, 3: (2005) 466–478.

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

Boarnet, M., and S. Sarmiento. "Can land-use policy really affect travel behavior? A study of the link between non-work travel and land-use characteristics." *Urban Studies* 35(7): (1998) 1155–1169.

Brons, M., P. Nijkamp, E. Pels, and P. Rietveld. "A meta-analysis of the price elasticity of gasoline demand. A SUR approach." *Energy Economics* 30: (2008) 2105–2122.

Brownstone, D., and T. Golob. "The impact of residential density on vehicle usage and energy consumption." *Journal of Urban Economics* 65: (2009) 91–98.

Cameron, C., and P. Triverdi. *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005.

CARB. "Expanded list of early action measures to reduce greenhouse gas emissions in California recommended for board consideration." California Air Resources Board, California Environmental Protection Agency, 2007a.

———. "Staff Report: 1990 Greenhouse Gas Emissions Level and 2020 Emissions Limit." California Air Resources Board, California Environmental Protection Agency, 2007b.

———. “Climate change proposed scoping plan: A framework for change.” California Air Resources Board, 2008.

CAT. “Updated Macroeconomic Analysis of Climate Strategies Presented in the March 2006 Climate Action Team Report: Final Report and Attachments.” Climate Action Team, California Environmental Protection Agency, 2007.

CEC. “The role of land use in meeting California’s energy and climate change goals.” California Energy Commission, 2007.

Cragg, J.G. “Some statistical models for limited dependen variables with application to the demand for durable goods.” *Econometrica* 39, 5: (1971) 829–844.

Dow, W., and E. Norton. “Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions.” *Health Services & Outcomes Research Methodology* 4: (2003) 5–18.

Duan, N., W. Manning, C. Morris, and J. Newhouse. “A comparison of alternative models for the demand of medical care.” *Journal of Business and Economic Statistics* 1, 2: (1983) 115–126.

ETAAC. “Recommendations of the Economic and Technology Advancement Advisory Committee. Final Report, A Report to the California Air Resources Board.” Economic and Technology Advancement Advisory Committee, 2008.

Ewing, R., K. Bartholomew, S. Winkelman, J. Walters, and D. Chen. “Growing Cooler: The Evidence on Urban Development and Climate Change.” Urban Land Institute, 2007.

Fang, H. “A discrete-continuous model of households’ vehicle choice and usage, with an application to the effects of residential density.” *Transportation Research Part B: Methodological* 42, 9: (2008) 736–758.

Hughes, J.E., C.R. Knittel, and D. Sperling. “Evidence of a shift in the short-run price elasticity of gasoline demand.” *Energy Journal* 29, 1: (2008) 93–114.

JHK, and Associates. “Transportation-Related Land Use Strategies to Minimize Motor Vehicle Emission.” California Air Resources Board, Sacramento, 1995.

Lin, Cynthia C.-Y., and Lea Prince. “The optimal gas tax for California.”, 2009. Working Paper.

Manning, W., and J. Mullahy. “Estimating log models: to transform or not to transform?” *Journal of Health Economics* 20: (2001) 461–494.

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

Mullahy, J. "Much ado about two: reconsidering retransformation and the two-part model in health econometrics." *Journal of Health Economics* 17: (1998) 247–281.

Parry, I.W., and K.A. Small. "Does Britain or the United States have the right gasoline tax?" *American Economic Review* 95, 4: (2005) 1276–1289.

Rodier, Caroline. "A Review of the International Modeling Literature: Transit, Land Use, and Auto Pricing Strategies to Reduce Vehicle Miles Traveled and Greenhouse Gas Emissions." UC Berkeley, Transportation Sustainability Research Center, 2008.

Small, K.A., and K. Van Dender. "Fuel efficiency and motor vehicle travel: The declining rebound effects." *The Energy Journal* 28, 1: (2007) 25–51.

Vance, C., and R. Hedel. "The impact of urban form on automobile travel: disentangling causation from correlation." *Transportation* 34: (2007) 575–588.

Appendix

Table A7: Variables description

Variable	Description
dr	Binary variable equal to 1 if a HH member traveled by car (as a driver) on the day surveyed, and equal to 0 otherwise
vmt	Total vehicle miles traveled (as drivers) by HH members on the day surveyed
resdes	Housing units per square mile in the census tract where the HH address is located
age	Age of the oldest HH member
work	Percentage of HH members who work
hhsize	Number of HH members
nveh	Number of vehicles in the HH
eduh	Binary variable indicating whether the highest level of school education of a HH member is at least a college degree or not.
inc1	Binary variable equal to 1 if the total annual income of the HH is less than \$10,000, and equal to 0 otherwise
inc2	Binary variable equal to 1 if the total annual income of the HH is greater than \$10,000 but less than \$25,000, and equal to 0 otherwise
inc3	Binary variable equal to 1 if the total annual income of the HH is greater than \$35,000 but less than \$55,000, and equal to 0 otherwise

Continued on next page

Table A7 - continued from previous page

Variable	Description
inc4	Binary variable equal to 1 if the total annual income of the HH is greater than \$35,000 but less than \$50,000, and equal to 0 otherwise
inc5	Binary variable equal to 1 if the total annual income of the HH is greater than \$50,000 but less than \$75,000, and equal to 0 otherwise
inc6	Binary variable equal to 1 if the total annual income of the HH is greater than \$75,000 but less than \$100,000, and equal to 0 otherwise
inc7	Binary variable equal to 1 if the total annual income of the HH is greater than \$100,000 but less than\$150,000, and equal to 0 otherwise
inc8	Binary variable equal to 1 if the total annual income of the HH is greater than \$150,000 ,and equal to 0 otherwise
inc9	Binary variable equal to 1 if the HH did not report total annual income, and equal to 0 otherwise
ur	Binary variable equal to 1 if the HH home address is located in an urban area, and equal to 0 otherwise
reg1	Binary variable equal to 1 if the HH home address is located in the Western Slope/Sierra Nevada Region (Amador, Calaveras, Mariposa and Tuolumne counties) . Equals 0 otherwise
reg2	Binary variable equal to 1 if the HH home address is located in the AMBAG Region (Monterey, San Benito and Santa Cruz counties) . Equals 0 otherwise
reg3	Binary variable equal to 1 if the HH home address is located in the MTC Region (Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano and Sonoma counties) . Equals 0 otherwise
reg4	Binary variable equal to 1 if the HH home address is located in the SACOG Region (El Dorado, El Placer, Sacramento, Sutter, Yolo, and Yuba counties) . Equals 0 otherwise
reg5	Binary variable equal to 1 if the HH home address is located in the SCAG Region (Imperial, Los Angeles, Orange, Riverside, San Bernardino, and Ventura counties) . Equals 0 otherwise
reg6	Binary variable equal to 1 if the HH home address is located in the Rural Region (Alpine, Colusa, Del Norte, Glenn, Humboldt, Inyo, Kings, Lake, Lassen, Madera, Mendocino, Modoc, Mono, Nevada, Plumas, Sierra, Siskiyou, Tehama, and Trinity counties) . Equals 0 otherwise
reg7	Binary variable equal to 1 if the HH home address is located in Butte county. Equals 0 otherwise
reg8	Binary variable equal to 1 if the HH home address is located in Fresno county. Equals 0 otherwise
reg9	Binary variable equal to 1 if the HH home address is located in Kern county. Equals 0 otherwise
reg10	Binary variable equal to 1 if the HH home address is located in Merced county. Equals 0 otherwise
reg11	Binary variable equal to 1 if the HH home address is located in San Diego county. Equals 0 otherwise
reg12	Binary variable equal to 1 if the HH home address is located in San Joaquin county. Equals 0 otherwise
reg13	Binary variable equal to 1 if the HH home address is located in San Luis Obispo county. Equals 0 otherwise
reg14	Binary variable equal to 1 if the HH home address is located in Santa Barbara county. Equals 0 otherwise
reg15	Binary variable equal to 1 if the HH home address is located in Shasta county. Equals 0 otherwise
reg16	Binary variable equal to 1 if the HH home address is located in Stanislaus county. Equals 0 otherwise

Continued on next page

Table A7 - continued from previous page

Variable	Description
reg17	Binary variable equal to 1 if the HH home address is located in Tulare county. Equals 0 otherwise
uaarea	Area in square miles of the urban area where the HH home address is located
prinocrmi	Transit route miles of transit systems primarily serving the urban area where the HH home address is located (ferry boats and commuter rail excluded)
trans	Transit miles density in the urban area where the HH home address is located (prinocrmi/uaarea)
mix	Ratio of business establishments to housing units in the zip code tabulation area where the HH home address is located

Table A8: Comparison between original and final samples

	original	*finalall*	*finalurb*
dr	0.8163	0.7561	0.7638
resdes*	1518.561	1570.346	2491.349
age	54.7443	53.5316	51.7816
work	0.4965	0.4999	0.5356
hhsize	2.356	2.0974	2.0593
nveh	1.9684	1.8315	1.723
eduh	0.412	0.4211	0.4344
inc1	0.043	0.0582	0.0483
inc2	0.142	0.1743	0.152
inc3	0.1317	0.1544	0.1496
inc4	0.139	0.1687	0.1684
inc5	0.1989	0.2211	0.2318
inc6	0.1086	0.1123	0.1208
inc7	0.0744	0.0783	0.0915
inc8	0.0342	0.0327	0.0376
inc9**	0.1283	0	0
ur*	0.7146	0.7206	1
reg1	0.04	0.0446	0
reg2	0.0509	0.0532	0.0512
reg3	0.0964	0.0862	0.1247
reg4	0.0576	0.0566	0.08
reg5	0.1986	0.2002	0.2599
reg6	0.143	0.1338	0
reg7	0.032	0.036	0.0317
reg8	0.0362	0.0356	0.049
reg9	0.0337	0.0344	0
reg10	0.0292	0.0331	0.0395
reg11	0.0697	0.0686	0.1191
reg12	0.0339	0.034	0.0432
reg13	0.038	0.0394	0.0163
reg14	0.0479	0.0484	0.0798

Continued on next page

Table A8 - continued from previous page

	original	*finalall*	*finalurb*
reg15	0.03	0.0334	0.0425
reg16	0.0315	0.0322	0.0403
reg17	0.0315	0.03	0.0227
mix*	0.0587	0.0586	0.0644
uaarea	NA	NA	445.6325
princocrmi	NA	NA	2759.798
trans	NA	NA	5.7254
Observations	17040	7666	4098

Notes: *The number of observations for *resdes*, *mix*, and *ur* is 17,014 in the *original* sample because 26 HH did not report their street address. **When observations with *inc9*=1 are removed from the *original* sample, the number of observations drops to 14,854 and the income dummies are more similar to those from *finalall* and*finalurb*

Table A9: Causes leading to sample size reduction

HH in Original Sample (original)	17040
HH with speed violations and non precise geo identification of destinations^{i}	5702
HH with missing trip information^{i}	246
HH surveyed for two days^{i}	512
HH with further speed violations for all modes^{ii}	1302
HH with the age of a member not reported	343
HH in which maximum age is smaller than 17	5
HH with the education level of a member not reported	267
HH with no income reported	995
HH located in ZCTA where the number of business establishments is greater than the number of housing units	2
HH in Final All Sample (finalall)	7666
HH not located in urbanized areas (i.e., located in rural areas or urban clusters)	3098
HH in urbanized areas not primarily served by transit systems	470
HH in Final Urban/Transit Sample (finalurb)	4098

Notes: *finalall* sample results after subtracting preceding rows from *original*, while *finalurb* after subtracting the two rows after *finalall*. The deletion procedure was performed according to the ordering of this table. For instance, with respect to the original sample more than 2000 HH did not report their incomes (compared to the 995 showed here, which results after other deletions took place). i) Flags included in the original dataset considered. ii) Speed flags constructed based on the following speeds (allowing a 15 minute extra trip duration for the upper bound): vehicle, motorcycle, and transit trips for which calculated speeds were more than 70 miles per hour (mph), or less than 1 mph; bicycle trips with mph>20; and walking trips with mph>10. Also included in this category are trips for which no duration or mode were reported.