Transportation Research Part D 47 (2016) 136–148

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Small area estimation of vehicle ownership and use

Yangwen Liu [a], Cinzia Cirillo [b,*]

[a] Neustar, Inc., New York, USA
[b] Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20740, USA

ARTICLE INFO

Article history:
Available online 23 May 2016

Keywords:
Vehicle ownership
VMT
Discrete–continuous model
Transferability
NHTS
ACS

ABSTRACT

A national model of vehicle ownership and use is developed for the USA. Decisions about the number of cars owned by households and the annual miles traveled are jointly modeled using a discrete–continuous probit model, which has been estimated on the 2009 National Household Travel Survey (NHTS) data. The model system covers four Census Regions (Northeast, Midwest, South and West) and three area types (urbanized area, urban clusters and rural). Models' estimates have been applied to data extracted from the American Community Survey (ACS) to forecast household vehicle demand at county level. Results show that the national models are transferable to small areas with different geographical and socio-demographic characteristics.

Introduction

In the transportation and energy related literature a consistent number of studies are dedicated to the vehicle ownership problem; applications exist for rural and urban areas (Dargay and Vythoulkas, 1999), large cities (Potoglou and Kanaroglou, 2008) and metropolitan areas (Potoglou and Susilo, 2008), regional (Cirillo and Liu, 2013) and national level (Hensher and Ton, 2002), and for developed (Bhat and Pulugurta, 1998b) and developing countries (Dargay et al., 2007; Li et al., 2010). The majority of these studies are based on data extracted from household travel surveys or dedicated data collections of stated preference type. Vehicle ownership models developed for the U.S.A. are mainly based on the National Household Travel Survey (NHTS), which was last conducted in 2009. The NHTS data contains a wealth of nation's daily travel information, including the number of cars in the households, their type and vintage and the annual miles traveled. However, the NHTS is designed at the national level, and the sample size is not large enough to produce design-based (direct survey weighted) estimates with acceptable precision at the state or finer level (e.g. county, municipality). In general, the NHTS data are not recommended for analysis of categories smaller than the combination of Census division, Metropolitan Statistical Area (MSA) size, and the availability of rail. According to Hu et al. (2007) extrapolating NHTS data within small geographic areas could risk developing and subsequently using unreliable estimates. Even though some States participate to the add-on program that collects supplementary sample for the States participating, many lack the necessary resources to collect local data and do not have the technical capabilities to estimate their own models.

Statistical techniques, known as Small Area Estimation (SAE), have been developed and used to obtain estimates for cases where the number of area-specific sample observations is not big enough to produce reliable direct estimates (Rao, 2003). The term "area" in SAE refers to any subpopulation or domain of interest, such as geographical domains (e.g. state or county), sociodemographic groups (e.g. income, race, age), land use characteristics (e.g. density) (Vaish et al., 2010). These techniques

* Corresponding author.
E-mail addresses: yangwen.liu@gmail.com (Y. Liu), ccirillo@umd.edu (C. Cirillo).

http://dx.doi.org/10.1016/j.trd.2016.05.002
1361-9209/

have been used for a number of applications and the demand for such data small areas has greatly increased during the last two decades. This increase is due to the usefulness of these data in government policy and program development, allocation of various funds and regional planning (Hidiroglou, 2007). For example, Statistics Canada applied them to obtain estimation of health statistics, of average weekly earnings, of under-coverage in the census, and of unemployment rates (Hidiroglou, 2007). The World Bank has developed a technique that combines information from household surveys (which contain comprehensive information) and censuses (which allow fine disaggregation). This statistical inference method is used to create local welfare estimates and detailed poverty maps; the derived information on poverty is sufficiently disaggregated to capture heterogeneity (Elbers et al., 2003). Examples of major small area estimation programs in USA include: the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program (see the SAIPE web site at http://www.census.gov/hhes/www/saipe/ for more information); the Bureau of Labor Statistics' Local Area Unemployment Statistics (LAUS) program (see the LAUS web site at http://www.bls.gov/lau/); the National Agricultural Statistics Service's County Estimates Program, which produces county estimates of crop yield (USDA 2007, see also at http://www.nass.usda.gov/); and the estimates of substance abuse in states and metropolitan areas (refer to http://www.samhsa.gov/ for details) (Rahman, 2008).

In transportation, very few studies develop comprehensive model systems for small area estimation of vehicle ownership and/or vehicle miles traveled. The barriers include the difficulties to capture demand levels for different population segments and different land use, and the limited data availability for small geographical areas. SAE techniques have been applied by Vaish et al. (2010) to produce small area estimates of the percentage of persons among different age groups having high daily person miles of travel. In Vaish et al. (2010) the authors use NHTS to estimate the model parameters, then the model is applied to predict person level probabilities. The desired small-area-level estimates are obtained by aggregating the person-level predicted probabilities using the appropriate population count projections for each of the 50 States in the USA.

In the context of transferability of transportation related model predictors, Hu et al. (2007) combined the 2001 NHTS data and 2000 census data to provide estimates of regional or local travel, including vehicle trips (VT), vehicle miles of travel (VMT), person trips (PT), and person miles of travel (PMT) by trip purpose and a number of demographics (Hu et al., 2007). In their report, transferability refers to the process of using statistical analysis on survey data sampled at one level (in the case of NHTS, the Census Division-MSA Size-Rail level) to estimate travel statistics at finer levels, such as state or local. Specifically, data was considered "transferable" in the case that estimates resulting from the transferability process were statistically valid (Hu et al., 2007).

This study develops a model system that estimates jointly vehicle ownership and use on national survey data (NHTS 2009). Twelve sub-models are estimated for respectively four Census Regions (Northeast, Midwest, South and West; Fig. 1) and three area types (urbanized area, urban clusters and rural; Fig. 2). Regions and area types were determined

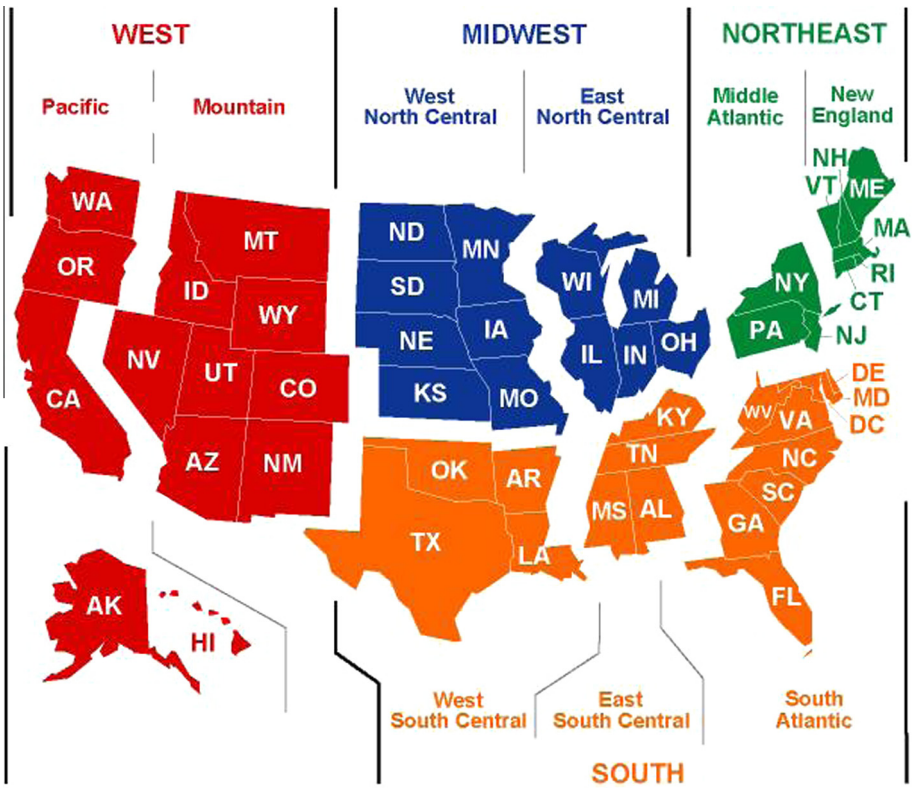


Fig. 1. United States regions (Census Bureau).

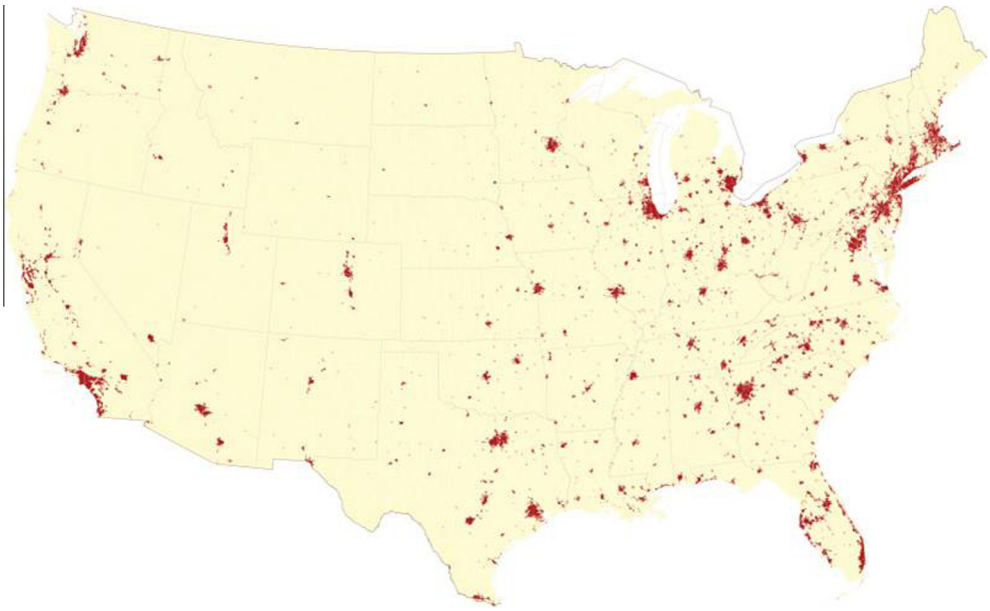


Fig. 2. United States urban area (Census Bureau).

according to the U.S. Census definitions. The models' predictors are then applied to calculate household predicted probabilities at county level using the American Community Survey (ACS) Public Use Microdata Sample (PUMS). Predicted and observed values for the number of vehicles in the household and annual vehicle miles traveled are finally compared to validate the results and the methodology.

Methodology

The methodology of small area estimation developed in this paper consists of three steps.

- Step 1 – We first estimate a joint discrete–continuous choice model of car ownership (number of vehicles owned by the household) and use (annual vehicle miles traveled) using a set of explanatory variables that are common to both NHTS and the ACS (e.g. household size, number of workers, housing and residential density, income and driving cost). This model system is estimated at the Census Region geographical level, for which the travel survey data is representative.
- Step 2 – Using the estimated coefficients from this model system we predict the level of car ownership and use for every household in the census. Basically, the model coefficients are used to "predict" the variables of interest for each household in the ACS on the basis of the explanatory variables that are common to the census and the survey.
- Step 3 – The household-unit data estimated in Step 2 are then used to calculate aggregate average vehicle ownership and annual vehicle miles traveled in the small areas (counties) selected for this study.

Discrete–continuous models

The discrete–continuous model system estimates jointly household's decisions on vehicle ownership (discrete choice) and usage (continuous choice). The integrated model uses a multinomial probit for the discrete decision variable and a regression for the continuous independent variable (Liu et al., 2014). Several studies have shown that unordered model structures are superior in terms of goodness of fit to ordered models for the car ownership problem (Bhat and Pulugurta, 1998a; Anowar et al., 2014). In particular, the analysis conducted by Cirillo et al. (in press) demonstrates that the same holds for discrete continuous models of car ownership and use, although model elasticities result to be similar.

In this context, the discrete problem concerns the forecast of the number of vehicles in a household (Y) using a set of predictors. Suppose there are $k+1$ $(0, 1, \ldots, k)$ vehicle ownership levels, the utility for each level consists of one observed part (systematic utility) and one unobserved part (error term):

$$
\begin{aligned}
U_0 &= \epsilon_0 \\
U_1 &= V_1 + \epsilon_1 \\
U_2 &= V_2 + \epsilon_2 \\
&\cdots \\
U_k &= V_k + \epsilon_k
\end{aligned}
$$

where U_k is the utility of having k vehicles; X are the explanatory variables associated with the household, the vehicles, and the land use; βs are the corresponding parameters to be estimated.

In the utility maximization theory, the household is assumed to be rational and to choose the alternative of vehicle ownership level that maximizes his utility. In this case, we adopt multinomial probit model for the vehicle holding decisions and therefore the error terms follow a multivariate normal distribution with full, unrestricted covariance matrix. The likelihood function can be expressed as follow:

$$P(Y = y|X, \beta, \lambda) = \int_{\mathbb{R}^{k+1}} \mathbb{I}(X_y^T \beta_y + \epsilon_y > X_j^T \beta_j + \epsilon_j \quad \forall j \neq y) \phi(\epsilon) d\epsilon \tag{1}$$

where

$X = (X_1, \ldots, X_k)$
$\beta = (\beta_1, \ldots, \beta_k)$
$\epsilon = (\epsilon_0, \epsilon_1, \ldots, \epsilon_k)$
$\Sigma :=$ Covariance of the error term

The functional indicator ($\mathbb{I}()$) ensures that the observed choice is indeed the one with the biggest utility. The subscript y indicates the predictors and coefficients of the chosen alternative and the subscript j indicate the other alternatives.

Since only differences in utility matter, the choice probability can be equivalently expressed as (k)-dimensional integrals over the differences between the errors. Suppose we differentiate against alternative y, the alternative for which we are calculating the probability:

$$\tilde{\epsilon}_{jy} = \epsilon_j - \epsilon_y \tag{2}$$

$$\tilde{V}_{jy} = (X_j^T \beta_j + J_j \lambda) - (X_y^T \beta_y + J_y \lambda) \tag{3}$$

$$\tilde{\epsilon}_y = \langle \tilde{\epsilon}_{1y}, \ldots, \tilde{\epsilon}_{ky} \rangle \tag{4}$$

where the "…" is over all alternatives except y.

Then the choice probability can be expressed as follows:

$$P(Y = y|X, \beta, \lambda) = \int_{\mathbb{R}^k} \mathbb{I}(\tilde{V}_{jy} + \tilde{\epsilon}_{jy} < 0 \quad \forall j \neq y) \phi(\tilde{\epsilon}_y) d\tilde{\epsilon}_y \tag{5}$$

which is a (k)-dimensional integral over all possible values of the error differences. The probit has been normalized to ensure that all parameters are identified.

Regression is adopted to model the continuous part of the model or the decision on the household vehicle mileage. In the regression, the dependent variable Y_{reg} is assumed to be a linear combination of a vector of predictors X_{reg} plus some error term:

$$Y_{reg} = X_{reg}^T \beta_{reg} + \epsilon_{reg} \quad \epsilon_{reg} \sim N(0, \sigma^2)$$

Given β_{reg}, X_{reg} and σ^2, the likelihood of observing Y_{reg} is given by the normal density function:

$$P(y_{reg}|\beta_{reg}, X_{reg}, \sigma^2) = \phi(y_{reg}|X_{reg}^T \beta_{reg}, \sigma^2) \tag{6}$$

In order to jointly capture the correlation between the discrete and continuous parts, we allow the error term of the regression to be correlated with the error terms of the utilities in the probit. Therefore, the specifications of the observable part of the utilities and of the regression remain the same, but the error terms follow an "incremental" normal distribution:

$$(\tilde{\epsilon}_{1y}, \tilde{\epsilon}_{2y}, \ldots, \tilde{\epsilon}_{ky}, \epsilon_{reg}) \sim MN(0, \Sigma_{k+1}) \tag{7}$$

In other terms:

$$\begin{bmatrix} \tilde{\epsilon}_y \\ \epsilon_{reg} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_y & \Sigma_{y,reg} \\ \Sigma_{reg,y} & \sigma^2 \end{bmatrix} \right) \tag{8}$$

The probability of observing Y and Y_{reg} can be derived as the product of the probability of observing Y_{reg} and the probability of observing Y conditional on observing Y_{reg}.

$$P(Y, Y_{reg}) = P(Y_{reg}) P(Y|Y_{reg})$$

The conditional probability of probit is:

$$P(Y|Y_{reg}) = \int_{\mathbb{R}^k} \mathbb{I}(\tilde{V}_{jy} + \tilde{\epsilon}_{jy} < 0 \quad \forall j \neq y) \phi(\tilde{\epsilon}_y) d\tilde{\epsilon}_y \tag{9}$$

where $\phi(\epsilon)$ is the density function of a multivariate distribution and

$$\tilde{\epsilon}_y \sim \mathcal{N}\left(0 + \frac{\Sigma_{disc,reg}}{\sigma^2_{reg}}(\epsilon_{reg} - 0), \Sigma_{disc} - \frac{\Sigma_{disc,reg}\Sigma_{reg,disc}}{\sigma^2_{reg}}\right)$$

Therefore the probability of the conditional probit has the same form of the multinomial probit except that $\tilde{\epsilon}_y$ follows a new mean and new variance.

The probability of the discrete part has no closed form so we rely on simulation:

$$\hat{P}(Y|Y_{reg}) = \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}(\tilde{V}_{jy} + \tilde{\epsilon}^{(b)}_{jy} < 0 \quad \forall j \neq y)$$

where $\mathbb{I}$ is an indicator of whether the statement in parentheses holds, $\tilde{\epsilon}^{(b)}_{jy}$ is a draw from a multivariate normal with mean $0 + \frac{\Sigma_{disc,reg}}{\sigma^2_{reg}}(err - 0)$ and variance $\Sigma_{disc} - \frac{\Sigma_{disc,reg}\Sigma_{reg,disc}}{\sigma^2_{reg}}$ and B is the number of simulations.

The final Simulated Log Likelihood of the model is given by the following formula:

$$SLL(\beta, \beta_{reg}, \Sigma|Y, Y_{reg}, X, J, X_{reg}) = \sum_{i=1}^{n}\log\left(\frac{B^*_i}{B} \times \phi(y_{i,reg}|X^T_{i,reg}\beta_{reg}, \sigma^2)\right)$$

where n is the total number of observations in the data, B^*_i is the number of success in the probit simulation (or the number of times $\mathbb{I}$ holds) for the ith observation. In this paper, simulations have been executed using 1000 pseudo Monte Carlo draws.

Data sources

Two data sources were used in this study: (1) the National Household Travel Surveys (NHTS) and (2) the American Community Survey (ACS).

NHTS is conducted by the Federal Highway Administration (FHWA), the United States Department of Transportation (U.S. DOT) and serves as the nation's inventory of daily travel. It collected travel data from a national sample of the civilian, non-institutionalized population of the United States. NHTS is a microdata dataset and the 2009 sample contains a total of 150,147 households, 351,275 persons, 309,163 vehicles and 1,167,321 trips.

The American Community Survey (ACS) is an ongoing statistical survey administered by the U.S. Census Bureau. It is sent to approximately 250,000 addresses monthly (or 3 million per year) and it regularly gathers information previously contained only in the long form of the decennial census. It is the largest survey other than the decennial census that the Census Bureau administers. In particular, The American Community Survey (ACS) Public Use Microdata Sample (PUMS) files show the full range of population and housing unit responses collected on individual ACS questionnaires. The PUMS files contain records for a subsample of ACS housing units and group quarters persons, with information on the characteristics of these housing units and group quarters persons plus the people in the selected housing units.

In terms of the geo-reference information, Region, Division, State, and Public Use Microdata Areas (PUMAs) are the only geographic areas identified in the ACS PUMS. Public Use Microdata Areas (PUMAs) are non-overlapping areas that partition each state into areas containing about 100,000 residents and are the most detailed geographic areas available in the ACS PUMS files (http://www.census.gov/acs/www/). In the NHTS data, Region, Division, State and the area type indicators are derived from the household's home address (confidential) and the U.S. Census boundary files.

The only built environment variable that is common to both NHTS and ACS is the density variable. According to the definition provided by the U.S. Census Bureau, urban areas are contiguous census block groups with a population density of at least 1000/sq mi with any census block groups around this core having a density of at least 500/sq mi. Urban areas are delineated without regard to political boundaries. The census has two distinct categories of urban areas. Urbanized areas have populations greater than 50,000, while urban clusters have populations of less than 50,000 but more than 2500. An urbanized area may serve as the core of a metropolitan statistical area, while an urban cluster may be the core of a micropolitan statistical area. Rural encompasses all population, housing, and territory not included within an urban area (http://www.census.gov/acs/www/).

Small area estimation

Step 1: Model estimation with NHTS data

We propose a model based SAE technique. The joint discrete–continuous modeling framework described in the methodological Section is used to estimate vehicle ownership and use household decisions. The model is estimated on the 2009 NHTS data. The entire US was divided into twelve macro-areas: four Census Regions (Northeast, Midwest, South and West) and three area types (urbanized area, urban clusters and rural) and subsequently twelve models were calibrated. Model estimation results are shown in Table 1. It should be noted that just a random sample of maximum 1500 observations was used for model estimation and for each combination of regions and area type. The discrete–continuous models is computationally

intensive and model calibration is particularly lengthy. Model specifications are the same for the twelve models estimated for consistency. The explanatory variables have been chosen based on their availability in both the NHTS and ACS databases; the final model includes the following attributes: household annual income level, household size, number of workers in the household, dummy of having child(ren), dummy of owned home, residential density, and driving cost ($/mile). Almost all of the coefficients are significant at 95% level and have the expected sign, with only few exceptions. For the discrete decision of owning a given number of cars all the estimates are positive (except for the density) and increase with the number of cars in the household; the same estimates decline for the fourth car alternative probably for the low number of observations for this alternative in our samples. Density has a negative effect on the number of vehicles owned by households in the US and the

Table 1
Estimation results of national models.

	Northeast			Midwest			South			West		
	Urban	Suburban	Rural	Urban	Suburban	Rural	Urban	Suburban	Rural	Urban	Suburban	Rural
Dependent variable: number of cars												
Constant												
1 car	−0.103	0.068	0.309	−0.003	−0.207	0.886	−0.284	−0.172	−0.044	0.313	0.452	1.014
2 cars	−21.33	−20.63	−19.78	−15.91	−8.63	−5.157	−23.60	−37.93	−14.09	−16.58	−8.03	−3.28
3 cars	−49.13	−50.86	−23.28	−26.85	−14.99	−26.47	−51.84	−56.69	−21.24	−18.60	−31.83	−11.18
4+ cars	−46.51	−52.97	−76.71	−38.73	−19.37	−20.74	−53.25	−129.8	−21.67	−36.64	−40.87	−15.19
Income												
1 car	0.086	0.084	0.08	0.074	0.166	0.031	0.103	0.113	0.164	0.065	0.105	0.089
2 cars	0.904	0.953	0.684	0.811	0.521	0.274	1.097	1.608	0.704	0.729	0.44	0.279
3 cars	1.429	1.37	0.747	0.829	0.642	0.516	1.619	1.906	0.822	0.749	1.022	0.401
4+ cars	1.03	0.863	1.125	0.801	0.559	0.434	1.668	2.737	0.658	0.885	0.997	0.37
Num. of hh members												
1 car	0.315	0.343	0.202	0.286	0.382	−0.105	0.257	0.366	−0.095	0.154	0.209	0.121
2 cars	3.376	4.135	4.528	3.289	1.933	1.387	3.88	7.211	2.829	3.31	2.1	0.897
3 cars	5.351	3.88	4.618	4.152	1.979	2.245	5.081	7.786	3.239	3.528	3.679	1.363
4+ cars	4.512	5.486	7.659	3.181	2.607	2.173	5.888	14.037	3.096	4.251	4.089	1.446
Num. of workers												
1 car	0.504	0.595	0.615	0.592	0.76	1.027	0.665	0.526	−0.17	0.094	0.328	0.789
2 cars	2.367	2.722	4.457	3.287	2.071	2.464	3.758	5.908	2.479	2.372	1.548	1.823
3 cars	5.745	8.413	5.008	4.7	4.156	6.643	7.132	8.617	3.353	2.595	4.644	3.037
4+ cars	4.025	9.495	9.631	9.813	5.11	5.982	7.622	11.447	4.148	4.748	5.481	3.339
Own home												
1 car	0.398	0.331	0.64	0.543	0.953	1.117	0.888	0.603	0.012	0.406	0.536	0.579
2 cars	6.946	5.597	7.456	5.189	3.956	3.956	7.071	12.591	5.136	5.146	2.929	2.357
3 cars	6.599	11.192	8.067	7.763	3.614	10.074	10.235	17.372	5.926	5.54	5.116	4.222
4+ cars	11.289	3.433	18.069	0.575	3.188	7.983	8.214	30.763	5.357	8.587	10.486	6.382
Res. density (1000)												
1 car	−0.063	−0.16	−0.197	−0.059	−0.089	0.38	−0.084	−0.179	0.275	−0.049	−0.239	−0.322
2 cars	−0.62	−0.721	−1.424	−1.268	−0.626	−2.718	−0.594	−2.574	−0.991	−0.589	−0.969	−0.798
3 cars	−1.171	−2.948	−1.477	−1.305	−0.843	−7.747	−1.194	−4.358	−3.758	−0.607	−2.62	−1.99
4+ cars	−1.334	−0.189	−1.974	−2.203	−1.036	−5.415	−1.382	−4.055	−4.981	−0.858	−4.991	−2.567
Dependent variable: AMT (10k)												
Constant	0.077	0.136	0.759	0.342	0.702	1.319	0.162	0.354	0.212	0.216	0.625	1.328
Income	0.062	0.072	0.055	0.06	0.099	0.051	0.071	0.087	0.071	0.069	0.07	0.054
Num. of hh members	0.252	0.255	0.22	0.166	0.221	0.154	0.271	0.23	0.239	0.387	0.301	0.183
Num. of workers	0.346	0.427	0.507	0.488	0.452	0.627	0.503	0.501	0.645	0.425	0.451	0.533
Own home	0.218	0.209	0.428	0.452	0.468	0.691	0.566	0.397	0.588	0.467	0.409	0.384
Has child(ren)	−0.004	0.165	0.207	0.168	0.179	0.328	0.128	0.406	0.144	−0.056	−0.129	0.086
Res. density (1000)	−0.041	−0.107	−0.144	−0.039	−0.017	−0.875	−0.048	−0.139	−0.102	−0.045	−0.186	−0.227
Driving cost ($ per mile)	−1.137	−1.175	−4.351	−2.732	−6.706	−7.184	−3.661	−2.996	−2.051	−5.105	−4.418	−6.048
Log-likelihood at zero	−9153	−7697	−8789	−8771	−8871	−7567	−9385	−9224	−7840	−8547	−6576	−7125
Final log-likelihood	−3114	−2476	−3488	−3396	−3202	−3838	−3457	−3580	−3927	−3517	−3205	−3836
Number of parameters	32	32	32	32	32	32	32	32	32	32	32	32
Number of observations	1500	1155	1500	1500	1333	1500	1500	1500	1500	1500	1297	1500
Adjusted R^2	0.656	0.674	0.600	0.609	0.635	0.489	0.628	0.608	0.495	0.585	0.508	0.457

absolute value of the coefficients is consistently increasing from the one car alternative to the four car alternative. Concerning the annual vehicle miles traveled, which constitutes our continuous dependent variable, household annual income level, household size, number of workers, and dummy of owned home have a positive effect, while density and cost are negative. The only variable whose sign changes across the different regions and areas is the dummy of having child(ren). We also report in Appendix A the variance–covariance matrices that result from the estimation of the fully joint discrete–continuous probit model.

STEP 2 – Model application with ACS data

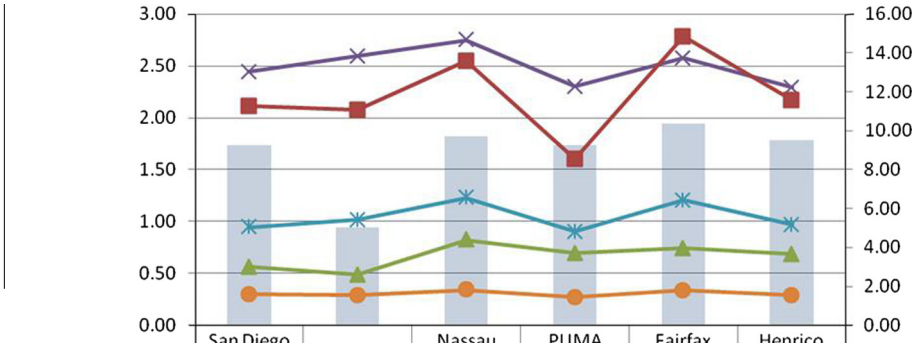
The models estimated in Table 1 for the macro-area on which the US territory has been divided are applied to households that are in the ACS to predict vehicle ownership (probability of owning 0, 1, 2, 3 or 4 cars) and use (annual mileage traveled) for the small geographical areas of interest. This is possible because we have intentionally selected in NHTS attributes that are also available in the ACS. This way to proceed will facilitate the use of the national model proposed by small planning agencies that do not have the technical capabilities to derive complex variables or to estimate their own model. The counties are selected with the objective to cover areas with different characteristics. The six counties included in our transferability study are: San Diego County in California, Queens in New York, Nassau County in New York, PUMA 1900 area (5 counties) in Texas, Fairfax County in Virginia, and Henrico County in Virginia. Fig. 3 presents general statistics derived for the six counties from the ACS PUMS files. It can be observed that the Fairfax County has the highest average household vehicle ownership rate. Queens has an average number of vehicles per household less than one, while the national average is close to two vehicles per household. Generally, the households in Fairfax County and Nassau County have bigger household size, more workers and children and much higher income. On average, about three-quarter of the households surveyed own their home, except for San Diego County and Queens where almost half of the households actually rent their place. We provide below a brief overview of the counties considered together with information on total surface, population, density and number of observations available in NHTS and ACS.

County/area descriptions

- San Diego County, CA – West, Urban

Although California has a large dataset from the statewide household travel survey, it is still worth to examine the performance of the national models on a metropolitan area with big cities; here we have selected San Diego County to perform this analysis. The basic demographic information and sample size from NHTS and ACS data are:

Total area: 4525.52 mile^2
Total population: 3,095,313 (2010 Census)
Population density: 680/mile^2
ACS: 11,653 obs.
NHTS: 3712 obs.



	San Diego County, CA	Queens, NY	Nassau County, NY	PUMA 1900, TX	Fairfax County, VA	Henrico County, VA
#veh/hh	1.74	0.94	1.82	1.74	1.95	1.78
D. owned home	0.56	0.49	0.82	0.70	0.74	0.69
hh size	2.45	2.60	2.75	2.30	2.57	2.29
#workers	0.95	1.02	1.23	0.91	1.21	0.97
D. has child(ren)	0.30	0.29	0.34	0.27	0.34	0.29
hh income	11.28	11.08	13.59	8.56	14.85	11.56

Fig. 3. Data statistics from American Community Survey.

- Queens, NY – Northeast, Urban

Queens Borough is a highly populated area in New York City, which is the most dense city in the U.S. People in this area may have different travel behavior than those residing in other regions. Meanwhile, this area has many immigrants from all around the world which may affect their travel choices as well. Again, although the New York City has good household travel surveys, it is still interesting to test the national models for this extremely dense area. The basic demographic information and sample size from NHTS and ACS data are:

Total area: 178.28 mile^2
Total population: 2,272,771 (2010 Census)
Population density: 21,116/mile^2
ACS: 6985 obs.
NHTS: 251 obs.

- Nassau County, NY – Northeast, Urban

Nassau County is located next to the east bounder of Queens in New York and many households in this county have jobs in New York City. This county is still within the New York metropolitan area and this application aims at validating the different travel styles within the same metropolitan area but different counties, thus validate the effectiveness of the national models. The basic demographic information and sample size from NHTS and ACS data are:

Total area: 453 mile^2
Total population: 1,339,532 (2010 Census)
Population density: 4669/mile^2
ACS: 4875 obs.
NHTS: 265 obs.

- PUMA 1900, TX – South, Rural

This area includes Hill County, Navarro County, Limestone County, Freestone County and Navarro County in Texas. This area is very scattered and it is located at roughly the middle point between Austin and Dallas – two big metropolitan areas in Texas. The 2009 NHTS only has less than 100 observations in this area, however ACS has around 900 observations. This is a case for which local household travel survey is not available and the national data sample has very limited observations. The basic demographic information and sample size from NHTS and ACS data are:

Hill County, TX Total area: 986 mile^2 Total population: 35,089 (2010 Census) Population density: 34/mile^2
Navarro County, TX Total area: 1086 mile^2 Total population: 47,735 (2010 Census) Population density: 18/mile^2
Limestone County, TX Total area: 933 mile^2 Total population: 23,384 (2010 Census) Population density: 23/mile^2
Freestone County, TX Total area: 892 mile^2 Total population: 19,816 (2010 Census) Population density: 21/mile^2
Navarro County, TX Total area: 779 mile^2 Total population: 17,866 (2010 Census) Population density: 57/mile^2
ACS: 894 obs.
NHTS: 93 obs.

- Fairfax, VA – South, Urban

Fairfax County is located in the Washington DC metropolitan area and west to the District of Columbia. It is one of the counties that have the highest household income in the country. Many people live in the Fairfax County commute to DC. The basic demographic information and sample size from NHTS and ACS data are:

Total area: 407 mile^2
Total population: 1,118,602 (2010 Census)
Population density: 2738.5/mile^2
ACS: 4033 obs.
NHTS: 205 obs.

- Henrico, VA – South, Urban

Henrico County is a portion of the Richmond Metropolitan area, surrounding the City of Richmond. Henrico is one of the oldest counties in the United States. The basic demographic information and sample size from NHTS and ACS data are:

Total area: 245 mile^2
Total population: 314,881 (2010 Census)

Population density: 1323/mile^2
ACS: 1274 obs.
NHTS: 379 obs.

Step 3 – Results from SAE

A prediction test (Ben-Akiva and Lerman, 1985) is adopted to compare predicted aggregate choice probabilities and VMT with the corresponding values observed in the ACS. For the discrete decision we run a sample enumeration on the

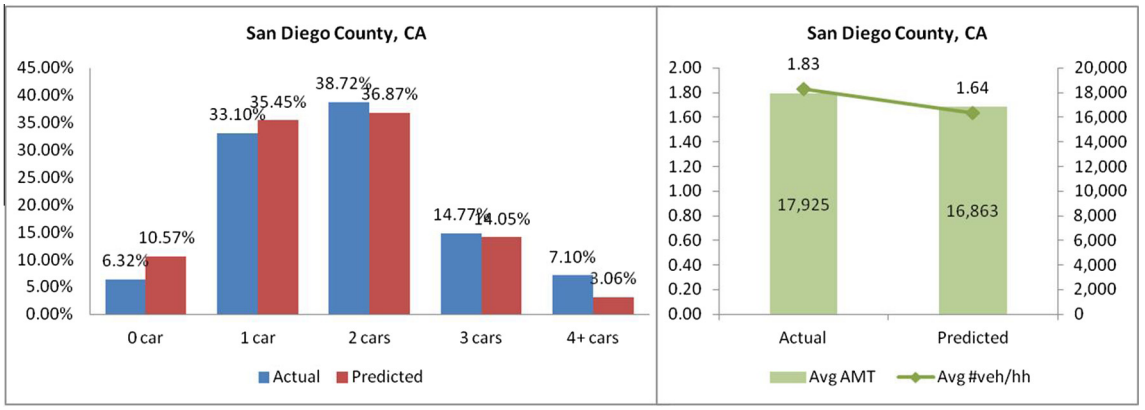


Fig. 4. Application results of San Diego County, CA.

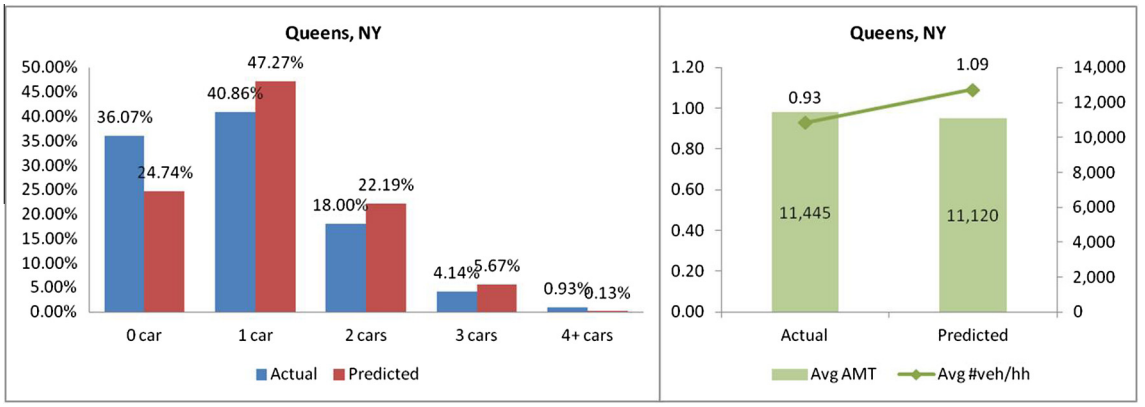


Fig. 5. Application results of Queens, NY.

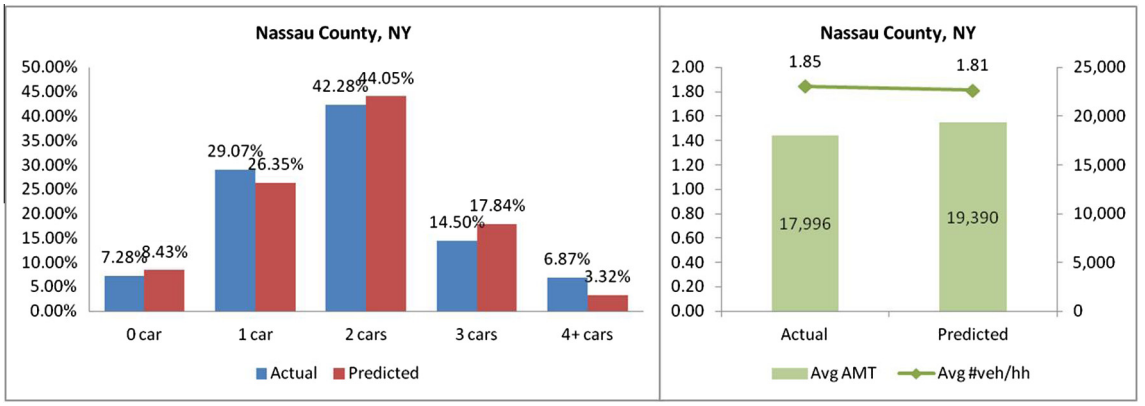


Fig. 6. Application results of Nassau County, NY.

households that are contained in the ACS, while we use the estimated regression model to calculate households vehicle miles traveled and its sample average. The comparison is done for each Census tract selected in the previous section (Figs. 4–9). Generally, the models are able to replicate quite well the percentage of household in each class of vehicle holding and the average miles traveled per year for each county/area. The model slightly underestimates the average vehicle ownership and mileage in San Diego County. For Queens, NY, the model overestimates the portion of 0-car households thus it overestimates the average number of vehicle per household. Nevertheless, the prediction of mileage is close to the actual value. Overall, the model slightly underestimates the average household vehicle ownership but overestimates the average annual

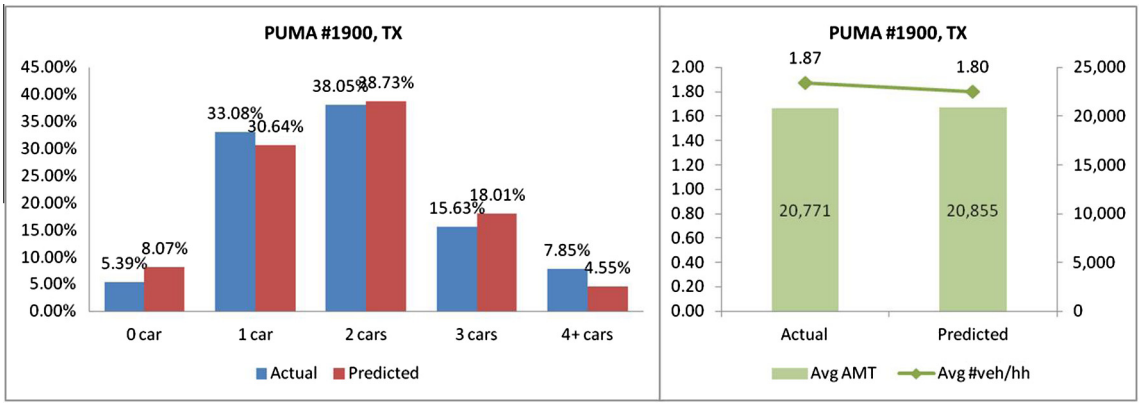


Fig. 7. Application results of PUMA area 1900, TX.

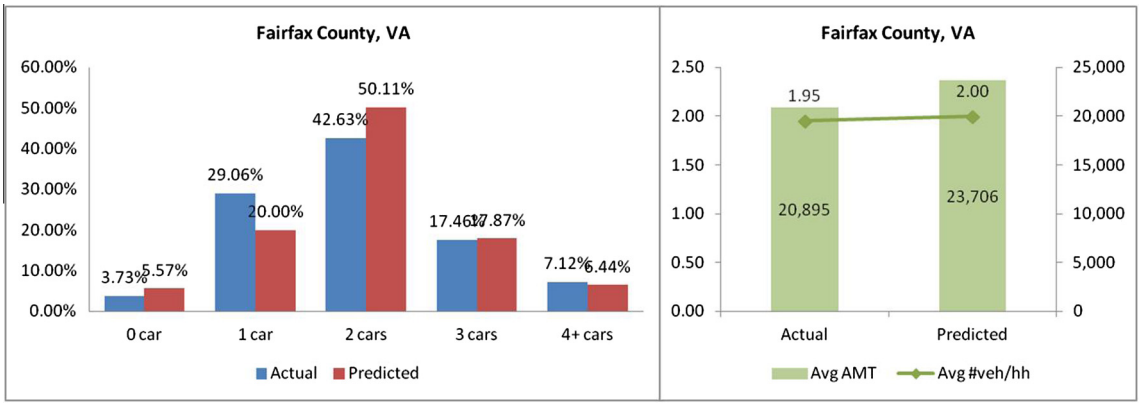


Fig. 8. Application results of Fairfax County, VA.

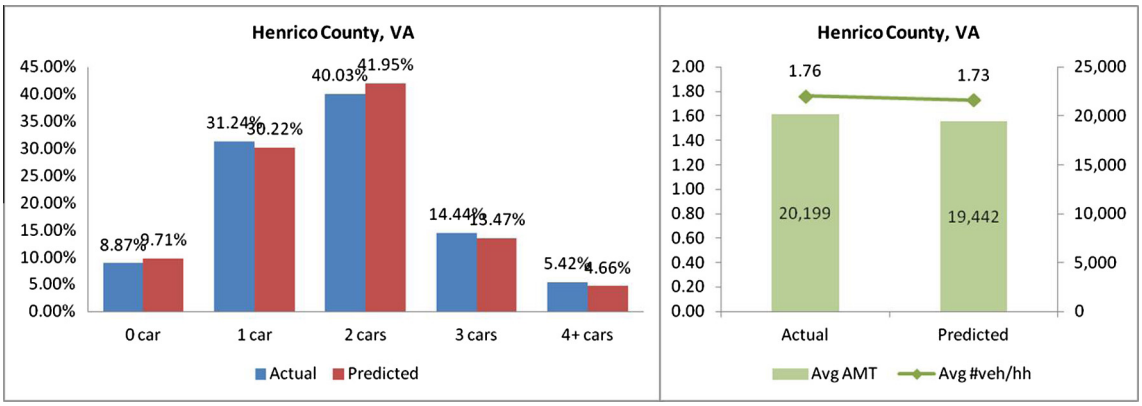


Fig. 9. Application results of Henrico County, VA.

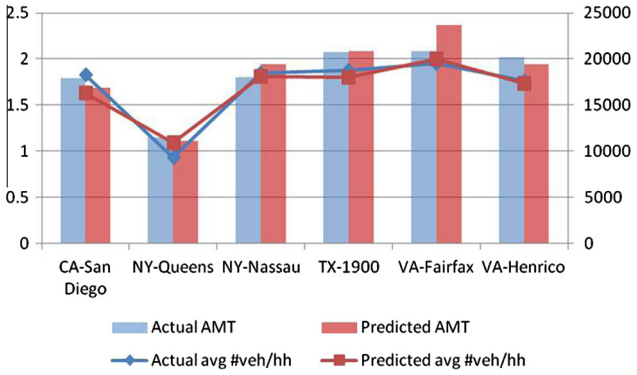


Fig. 10. Summary of applications for the six counties/areas.

mileage per household. The estimates for the PUMA 1900 area in Texas are very close to the actual values, with the exception of small shifts in the share of the alternatives. The application results for the Fairfax County shows that the model underestimates the share of 1-car households but overestimates the share of 2-car households, and it overestimates the average mileage for this county. The predictions for the Henrico County are fairly close to the real values, both for the vehicle ownership and the annual mileage. Finally, Fig. 10 summarizes the application results for the six counties/areas.

Conclusions

This paper has estimated a model system that forecasts US national demand of vehicle ownership and use. The models, that estimate jointly the number of vehicles in the households and the annual vehicle miles traveled, are based on 2009 NHTS sample households that are located in the four Census Regions (Northeast, Midwest, South, West) and three areas (urbanized area, urban clusters and rural) on which the USA has been subdivided. The analysis also integrates the NHTS data with ACS data to estimate vehicle holding and VMT for small geographic areas such as County level. Results show that the proposed method that transfers discrete–continuous models based on NHTS to smaller areas, successfully forecasts the parameters of interest. This framework constitutes a planning tool at both national level and smaller scale; it is especially valuable for those agencies that are lacking local household travel survey data or that do not have the technical capabilities to estimate demand models. Future research might be aimed at extending this modeling framework to the estimation of other key travel parameters (i.e. number of trips, frequency of trips by purpose) to small areas.

Acknowledgements

This material is based upon work supported by the National Science Foundation – United States under Grant No. 1131535. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors thank an anonymous reviewer and the editor, whose comments greatly improved the manuscript.

Appendix A. Variance–covariance matrices

$$\hat{\Sigma}_{NE,urban} = \begin{pmatrix} 2.00 & 6.25 & 9.31 & 8.98 & 1.46 \\ 6.25 & 67.98 & 129.25 & 36.71 & 4.91 \\ 9.31 & 129.25 & 332.54 & 61.85 & 7.63 \\ 8.98 & 36.71 & 61.85 & 89.06 & 6.73 \\ 1.46 & 4.91 & 7.63 & 6.73 & 1.06 \end{pmatrix}$$

$$\hat{\Sigma}_{NE,suburban} = \begin{pmatrix} 2.00 & 8.59 & 12.58 & 11.78 & 1.48 \\ 8.59 & 72.33 & 128.74 & 89.07 & 6.09 \\ 12.58 & 128.74 & 377.38 & 60.82 & 8.79 \\ 11.78 & 89.07 & 60.82 & 204.98 & 8.20 \\ 1.48 & 6.09 & 8.79 & 8.20 & 1.10 \end{pmatrix}$$

$$\hat{\Sigma}_{NE,rural} = \begin{pmatrix} 2.00 & 5.05 & 5.69 & 10.59 & 1.43 \\ 5.05 & 82.50 & 84.65 & 104.52 & 6.84 \\ 5.69 & 84.65 & 89.30 & 105.84 & 7.35 \\ 10.59 & 104.52 & 105.84 & 332.62 & 11.36 \\ 1.43 & 6.84 & 7.35 & 11.36 & 1.18 \end{pmatrix}$$

$$\hat{\Sigma}_{MW,urban} = \begin{pmatrix} 2.00 & 7.37 & 8.53 & 9.69 & 1.51 \\ 7.37 & 78.10 & 89.20 & 68.67 & 6.54 \\ 8.53 & 89.20 & 117.14 & 53.37 & 7.55 \\ 9.69 & 68.67 & 53.37 & 161.23 & 7.96 \\ 1.51 & 6.54 & 7.55 & 7.96 & 1.16 \end{pmatrix}$$

$$\hat{\Sigma}_{MW,suburban} = \begin{pmatrix} 2.00 & 6.55 & 5.77 & 6.44 & 1.06 \\ 6.55 & 22.20 & 22.37 & 24.84 & 3.57 \\ 5.77 & 22.37 & 39.18 & 38.60 & 4.62 \\ 6.44 & 24.84 & 38.60 & 43.48 & 5.01 \\ 1.06 & 3.57 & 4.62 & 5.01 & 1.22 \end{pmatrix}$$

$$\hat{\Sigma}_{MW,rural} = \begin{pmatrix} 2.00 & 3.24 & 3.43 & 4.74 & 1.48 \\ 3.24 & 13.36 & 29.38 & 30.17 & 3.33 \\ 3.43 & 29.38 & 196.83 & 131.85 & 8.23 \\ 4.74 & 30.17 & 131.85 & 101.17 & 7.51 \\ 1.48 & 3.33 & 8.23 & 7.51 & 1.31 \end{pmatrix}$$

$$\hat{\Sigma}_{S,urban} = \begin{pmatrix} 2.00 & 5.93 & 16.64 & 12.20 & 1.51 \\ 5.93 & 97.61 & 30.17 & 118.66 & 6.13 \\ 16.64 & 30.17 & 190.82 & 125.52 & 10.84 \\ 12.20 & 118.66 & 125.52 & 199.50 & 9.69 \\ 1.51 & 6.13 & 10.84 & 9.69 & 1.22 \end{pmatrix}$$

$$\hat{\Sigma}_{S,suburban} = \begin{pmatrix} 2.00 & 8.87 & 12.80 & 19.90 & 1.59 \\ 8.87 & 290.51 & 215.82 & 359.41 & 9.16 \\ 12.80 & 215.82 & 262.21 & 292.96 & 11.42 \\ 19.90 & 359.41 & 292.96 & 881.16 & 18.09 \\ 1.59 & 9.16 & 11.42 & 18.09 & 1.28 \end{pmatrix}$$

$$\hat{\Sigma}_{S,rural} = \begin{pmatrix} 2.00 & -8.82 & -1.84 & -9.54 & -0.35 \\ -8.82 & 47.61 & 14.77 & 57.01 & 4.02 \\ -1.84 & 14.77 & 45.22 & 7.22 & 5.11 \\ -9.54 & 57.01 & 7.22 & 81.25 & 5.28 \\ -0.35 & 4.02 & 5.11 & 5.28 & 1.32 \end{pmatrix}$$

$$\hat{\Sigma}_{W,urban} = \begin{pmatrix} 2.00 & -0.60 & -0.57 & 0.68 & 1.18 \\ -0.60 & 24.59 & 25.61 & 24.81 & 3.01 \\ -0.57 & 25.61 & 27.17 & 29.50 & 3.22 \\ 0.68 & 24.81 & 29.50 & 53.94 & 4.33 \\ 1.18 & 3.01 & 3.22 & 4.33 & 1.18 \end{pmatrix}$$

$$\hat{\Sigma}_{W,suburban} = \begin{pmatrix} 2.00 & 3.45 & 5.37 & 5.49 & 1.56 \\ 3.45 & 17.76 & 27.14 & 40.06 & 3.24 \\ 5.37 & 27.14 & 176.57 & 143.66 & 6.49 \\ 5.49 & 40.06 & 143.66 & 156.55 & 6.76 \\ 1.56 & 3.24 & 6.49 & 6.76 & 1.27 \end{pmatrix}$$

$$\hat{\Sigma}_{W,rural} = \begin{pmatrix} 2.00 & 4.98 & 7.11 & 8.01 & 1.29 \\ 4.98 & 13.83 & 21.37 & 24.18 & 3.07 \\ 7.11 & 21.37 & 40.02 & 41.96 & 4.48 \\ 8.01 & 24.18 & 41.96 & 48.89 & 5.03 \\ 1.29 & 3.07 & 4.48 & 5.03 & 1.29 \end{pmatrix}$$

References

Anowar, S., Eluru, N., Yasmin, S., Miranda-Moreno, L.F., 2014. Analyzing car ownership in quebec city: a comparison of traditional and latent class ordered and unordered models. Transportation 41 (5), 1013–1039.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis. The MIT Press, Cambridge, Massachusetts.

Bhat, C.R., Pulugurta, V., 1998a. A comparison of two alternative behavioral mechanisms for car ownership decisions. Transp. Res. Part B 32 (1), 61–75.

Bhat, C.R., Pulugurta, V., 1998b. A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. Transp. Res. Part B 32 (1), 61–75.

Cirillo, C., Liu, Y., 2013. A vehicle ownership model for the State of Maryland: analysis and trends from 2001 and 2009 NHTS data. J. Urban Plann. Dev. 139 (1), 1–11.

Cirillo, C., Liu, Y., Tremblay, J.M., 2016. Simulation, numerical approximation and closed forms for joint discrete continuous models with an application to household vehicle ownership and use. Transportation (in press).

Dargay, J.M., Vythoulkas, P.C., 1999. Car ownership in Rural and Urban Areas: A Pseudo Panel Analysis. ESRC Transport Studies Unit, Centre for Transport Studies, University College London.

Dargay, J.M., Gately, D., Sommer, M., 2007. Vehicle ownership and income growth, worldwide: 1960–2030. Energy J. 28 (4), 143–170.

Elbers, C., Lanjouw, J.O., Lanjouw, P., 2003. Micro-level estimation of poverty and inequality. Econometrica 71 (1), 355–364.

Hensher, D.A., Ton, T., 2002. TRESIS: a transportation, land use and environmental strategy impact simulator for urban areas. Transportation 29 (4), 439–457.

Hidiroglou, M.A., 2007. Small-area estimation: theory and practice. In: Proceedings of the Joint Statistical Meeting, Section on Survey Research Methods in Salt Lake City Utah, pp. 3445–3456.

http://www.census.gov/acs/www/.

Hu, P.S., Reuscher, T., Schmoyer, R.L., Chin, S.M., 2007. Transferring 2001 National Household Travel Survey <http://nhts.ornl.gov/tx/TransferabilityReport.pdf>.

Li, J., Walker, J.L., Srinivasan, S., Anderson, W.P., 2010. Modeling private car ownership in China: investigation of urban form impact across megacities. Transp. Res. Rec. 2193, 76–84.

Liu, Y., Tremblay, J.M., Cirillo, C., 2014. An integrated model for discrete and continuous decisions with application to vehicle ownership, type and usage choices. Transp. Res. Part A 69, 319–328.

Potoglou, D., Kanaroglou, P.S., 2008. Modelling car ownership in urban areas: a case study of Hamilton, Canada. J. Transp. Geogr. 16 (1), 42–54.

Potoglou, D., Susilo, Y.O., 2008. Comparison of vehicle-ownership models. Transp. Res. Rec. 2076, 97–105.

Rahman, A., 2008. A Review of Small Area Estimation Problems and Methodological Developments. Discussion Paper 66. NATSEM, University of Camberra.

Rao, J.N.K., 2003. Small Area Estimation. Wiley, New-York.

Vaish, A.K., Chen, S., Sathe, N.S., Folsom, R.E., Chandhok, P., Guo, K., 2010. Small area estimates of daily person-miles of travel: 2001 national household transportation survey. Transportation 37, 825–848.