Transportation Research Part A 69 (2014) 315–328

Contents lists available at ScienceDirect

Transportation Research Part A

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

An integrated model for discrete and continuous decisions with application to vehicle ownership, type and usage choices

Yangwen Liu [1], Jean-Michel Tremblay, Cinzia Cirillo *

University of Maryland, Department of Civil and Environmental Engineering, United States

ARTICLE INFO

Article history:
Received 1 September 2013
Received in revised form 8 September 2014
Accepted 9 September 2014
Available online 2 October 2014

Keywords:
Discrete–continuous model
Monte-Carlo simulation
Vehicle ownership
Vehicle miles traveled
US National Household Travel Survey

ABSTRACT

This paper proposes an integrated econometric framework for discrete and continuous choice dimensions. The model system is applied to the problem of household vehicle ownership, type and usage. A multinomial probit is used to estimate household vehicle ownership, a multinomial logit is used to estimate the vehicle type (class and vintage) choices, and a regression is used to estimate the vehicle usage decisions. Correlation between the discrete (number of vehicles) and the continuous (total annual miles traveled) parts is captured with a full variance–covariance matrix of the unobserved factors. The model system is estimated using Simulated Log-Likelihood methods on data extracted from the 2009 US National Household Travel Survey and a secondary dataset on vehicle characteristics. Model estimates are applied to evaluate changes in vehicle holding and miles driven, in response to the evolution of social societies, living environment and transportation policies.

1. Introduction

Increasing mobility demand, especially in urban areas, has resulted in growing levels of motorization, congestion and pollution. Modern societies are still highly dependent on private vehicles to satisfy demand for activities, while the fastest growing economies in the world are experiencing a rapid increase in motor vehicle ownership. It is clear that vehicle demand has to be optimally managed and regulated in order to reduce the adverse impacts of transportation. In this context, the role of analysts and researchers is to expand the basic knowledge of the problem, to develop better analytical tools and to support decision makers in their strategic choices.

The importance of modeling household vehicle fleet choices has been recognized for several decades (Hensher et al., 1992; Brendemoen, 1994; TNO-Inro, 1999), though the urgency in terms of GHG emissions and fossil fuel energy dependence is more recent (Vyas et al., 2012).

Car ownership models play an important role in transportation and land use planning and are a critical component of Transportation Modeling Systems. In the classical four-step forecasting model, the trip generation module uses the outputs from car ownership models (Golob and van Wissen, 1989; Kitamura, 2009) as its inputs. Furthermore, vehicle ownership greatly impacts mode choice (Dissanayake and Morikawa, 2010), trip frequency (Kitamura, 2009; Shay and Khattak, 2012), destination choice, trip timing, activity duration and trip chaining properties (Hatzopoulou et al., 2007; Roorda et al., 2009; Paleti et al., 2013).

* Corresponding author. Tel.: +1 (301) 405 6864; fax: +1 (301) 405 2585.
E-mail addresses: aliceliu@umd.edu (Y. Liu), jeanmi.tremblay@gmail.com (J.-M. Tremblay), ccirillo@umd.edu (C. Cirillo).
[1] Tel.: +1 (240) 898 5225.

http://dx.doi.org/10.1016/j.tra.2014.09.001
0965-8564/

Models for car ownership are of interest to both public agencies and private organizations. The US Department of Energy, the State Departments of Transportation, the auto industry, and the World Bank have supported studies on vehicle ownership and used their results for policy analysis (Train, 1986).

A number of agencies have implemented vehicle ownership in their regional transportation models. The State of California has developed the Motor Vehicle Stock, Travel and Fuel Forecast (MVSTAFF) model that uses a macroeconomic approach to modeling statewide motor vehicle holdings, vehicle miles traveled (VMT) and total fuel consumption. Other model systems that include a car ownership component are: the Maryland Statewide Transportation Model, the Coordinated Travel-Regional Activity Based Modeling Platform (CT-RAMP) for the Atlanta Region from Atlanta Regional Commission (2009), the activity based model from the Puget Sound Regional Council (2008), etc.

National governments use car ownership models to forecast tax revenues and the regulatory impact of changes in the level of taxation (Hayashi et al., 2001; Brannlund and Nordstrom, 2004; Giblin and McNabola, 2009). This type of model system examines the changes in the car market configuration, the life cycle CO_2 emission from automobile transport and the tax revenues due to different taxation policies (Hayashi et al., 2001).

Vehicle ownership models are also used by policy makers to identify factors that affect VMT, and therefore address the problems related to traffic congestion, gas consumption and air pollution (Dargay and Gately, 1997; Schipper, 2011). Models for car ownership growth in developing countries are important to estimate the implications on energy demand and price and on the global CO_2 emissions (Dargay and Gately, 1997).

This paper develops an integrated modeling framework for household decisions on vehicle ownership, class/vintage, and use. The proposed model structure allows the simultaneous analysis of both discrete and continuous dependent variables, that are potentially correlated. For instance, the number of vehicles owned by a household, their class and vintage is a typical *discrete* problem, while the total number of miles driven is represented by a *continuous* distribution. The analysis is based on a large number of policy variables and the estimates obtained are used for the evaluation of alternative scenarios.

The remaining of this paper is organized as follows. Section 2 contains a review of studies on discrete–continuous models for the vehicle ownership problem. Section 3 introduces the model formulation, and in particular (1) derives the model's properties (2) proposes a Full Information Maximum Likelihood (FIML) for discrete and continuous joint decisions and (3) describes the simulation method adopted to solve this non-closed form maximum log-likelihood estimation problem. Section 4 describes the data extracted from the 2009 National Household Travel Survey. In Section 5 and 6 the integrated model (household vehicle holding, vehicle class and vintage and total vehicle miles traveled) is estimated and applied to the Washington Metropolitan Region. Finally, Section 7 presents concluding remarks and avenues for future research.

2. Literature review: Discrete–continuous models for car ownership and use

A large number of studies have investigated vehicle ownership choices using discrete–continuous simultaneous equations. These models (Train, 1986; Mannering and Winston, 1985; de Jong, 1989) are based on the hypothesis that households choose the combination of vehicle ownership and vehicle usage that provides the highest utility. Roy's identity is applied to estimate vehicle usage and the relationship between the two modeling stages. These studies based on the indirect utility function, are consistent with the economic theory and are able to capture the interdependence between the vehicle holding and the corresponding mileage by means of observed variables.

The earliest generation of unified frameworks for car ownership and use were based on methods developed and applied in economics starting from the 70's. Heckman (1978) developed a class of econometric models for simultaneous equation systems with dummy endogenous variables; this general model includes simultaneous probit as a special case. Discrete–continuous models of consumer demand were formulated by Hanemann (1984) to jointly model the discrete choice among different brands of a commodity and the continuous choice of how many units to buy. The problem of non-linear budget sets in discrete continuous models for consumer demand was studied by Hausman (1985) and the methods proposed were applied to a labor supply model. In the same paper, the author also discusses the case of a nonlinear budget set for a household facing the decision to purchase a durable good (car), where the price per mile driven depends on the fuel efficiency of the car being used.

Dubin and McFadden (1984) consider the estimation of the Heckman selectivity problem under the assumption that the utility in the first stage model (discrete part) has a logistic rather than a normal distribution; the selection equation is of the logit rather than the probit form (Dubin and River, 1990). The model belongs to the class of single disturbance models. Because these models involve only one error term, identification of the parameters can proceed under rather weak conditions, such as symmetry of the error distribution (Chamberlain, 1986).

A micro-economic utility model was developed by de Jong in 1990. The model simultaneously determines private car ownership and private car use (measured as the annual number of kilometers), but is limited to the single car case. Both fixed and variable car costs enter the model system through the budget restriction. The model was then applied to micro-simulate increases in those costs and to test policy issues in the Netherlands. Results show that both fixed and variable car costs are effective measures for reducing traffic (growth), the former working primarily through decreasing car ownership levels, the latter having a more direct effect on car use (de Jong, 1990).

Multiple discrete–continuous extreme value (MDCEV) models, developed by (Bhat, 2005) and further applied in (Bhat and Sen, 2006; Bhat et al., 2009) are utility-based econometric models that jointly estimate the holding of multiple vehicle types

and the miles for each vehicle type. The choice and dependent variable in this model is the mileage for each vehicle type category. Utility for each household is maximized subject to a total mileage budget. Under the assumption that the error term is i.i.d. extreme value distributed, the probability function simplifies to an elegant and compact closed form, and collapses to MNL model for one car households.

Bhat and Sen (2006) presented an application of MDCEV that models simultaneous holding of multiple vehicle type (passenger car, SUV, pickup truck, minivan and van) and miles of usage of each vehicle type in a joint system, on the data drawn from the 2000 San Francisco Bay Survey. The authors analyze changes in vehicle type and usage due to the changes in demographics, employment, density and operating cost. Bhat et al. (2009) extended the 2006 study and formulated a nested model structure that includes a multiple discrete–continuous extreme value (MDCEV) component to analyze the choice of vehicle type/vintage and usage in the upper level and a multinomial logit (MNL) component to analyze the choice of vehicle make/model in the lower level. The model accommodates heteroscedasticity and/or error correlation in both the multiple discrete–continuous component and the single discrete choice component of the joint model using a mixing distribution. The joint model also incorporates random coefficients in one or both components of the joint model.

The MDCEV model recognizes multiple discreteness and is able to handle a large number of vehicle types. It captures well the interdependence between the vehicle type and the corresponding mileage and allows more complex specification forms such as heteroscedasticity and correlation. However, this model requires finer classification of vehicles as no one type of vehicle can be chosen twice for a household. This type of models is limited by the assumption of fixed total mileage budget for every household; this implies that it is not possible to predict changes in the total number of miles in response to policy changes. In conclusion, the MDCEV model is consistent with random utility, it is able to capture trade-offs among the usage of different types of vehicles and can accommodate a large number of vehicle classifications.

Fang (2008) developed the BMOPT (Bayesian Multivariate Ordered Probit and Tobit) model, which is composed of a multivariate ordered probit model for the discrete choices and a multivariate Tobit model for the continuous choice. Household decisions on the number of vehicles in one of the two categories (cars and trucks) considered are estimated by means of ordered probit model. The multivariate Tobit model is applied to estimate the household decisions on miles driven with each vehicle type. The joint model is formulated with an unrestricted covariance matrix for the discrete and continuous parts. The BMOPT model is convenient to implement, and can be applied to study policy implications. It is able to handle a large number of vehicles, and captures the interdependence (correlation) between the number of vehicles and total miles driven in each type category, with flexible specifications of error terms. There are a few limitations in the model structure. Firstly, the computation becomes intensive for a large number of vehicle categories, as the number of equations to be estimated increases proportionally with the number of vehicle types. Another concern is that the ordered mechanism may not perform as well as unordered mechanism in modeling car ownership models (Bhat and Pulugurta, 1998)). Lastly, the same variables enter both discrete and continuous sub-model; although this is due to the specification chosen by the analyst and should not be interpreted as a limitation of the BMOPT model structure. Overall, the model is well suited for predicting the changes in the number of vehicles and miles traveled for each vehicle type category modeled.

In a European context, Rouwendal and de Borger developed a model system of car ownership and car use to study the optimal tax treatment of car ownership and car use. The model assumes households have the choice between owning no car, owning one car or owning two cars. For each car there is an annual cost of ownership and, in case of car use, a variable cost per kilometer. Cars differ in terms of fixed and variable cost: for example, the first car households own is typically of higher quality than possible second cars. A crucial ingredient of the model is that multiple car ownership introduces nonlinearities in the demand for total household kilometers (Rouwendal and de Borger, 2009).

3. The modeling framework

In this paper, we propose an econometric framework for both discrete and continuous decision variables in the context of household vehicle ownership. Three main choices are taken into account: the number of vehicles, their class and vintage, the annual mileage traveled (Fig. 1). The model system explicitly accounts for a large number of alternatives in the vehicle class

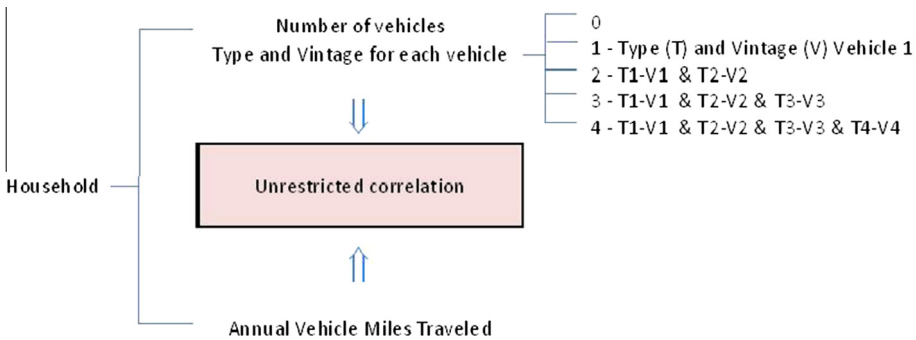

Fig. 1. Structure of the integrated model.

and vintage problem. Moreover, a flexible correlation structure between the unobserved factors of the discrete (number of vehicles in the household) and the continuous (miles driven) parts offers an integrated model system for household decisions that are naturally linked.

3.1. The discrete choice sub-model

We adopt discrete choice analysis to model both vehicle ownership and vehicle class and vintage choices. Discrete choice models forecast the outcome of a categorical dependent variable Y_{disc} using some set of predictors. In our modeling framework, we assume that the vehicle ownership model has five alternatives. The alternatives of owning 0, 1, 2, 3, 4 cars (or Y_{disc}) have a utility $(U_0, U_1, U_2, U_3, U_4)$ that consists of one observable part (systematic utility, **V**) and one non-observable part (error term ϵ). The observed utility of zero-car alternative is set to 0 for normalization purpose; for the remaining ownership alternatives, the observed utility is decomposed into two parts V_j and $V_{t_j|j}$ $(j = 0, 1, 2, 3, 4)$:

$$U_0 = \epsilon_0$$
$$U_1 = V_1 + \lambda V_{t_1|1} + \epsilon_1$$
$$U_2 = V_2 + \lambda V_{t_2|2} + \epsilon_2$$
$$U_3 = V_3 + \lambda V_{t_3|3} + \epsilon_3$$
$$U_4 = V_4 + \lambda V_{t_4|4} + \epsilon_4$$

Where V_j is the utility of the vehicle holding decision, which depends on factors that vary over j and $V_{t_j|j}$ is the utility of vehicle type choice t conditional on j. $t_j|j$ is the choice set containing all the possible combinations of car types and vintages while λ is a parameter to be estimated.

We adopt a multinomial logit model for the vehicle type-vintage sub-model; then the probability of choosing a certain type of vehicle is:

$$P_{t'_j|j} = \frac{exp(V_{t'_j|j})}{\sum_{t_j} exp(V_{t_j|j})}$$

where t'_k is the chosen alternative among total alternatives t_k. The utility that the household would obtain by its choice of vehicle type can be written as:

$$J_j = ln \sum_{t_j} exp(V_{t_j|j})$$

with J_j being the logsum of the vehicle type sub-model.

Therefore the utility of the discrete choice concerning vehicle holding can be further written as:

$$U_0 = \epsilon_0$$
$$U_1 = X_1^T \beta_1 + J_1 \lambda + \epsilon_1$$
$$U_2 = X_2^T \beta_2 + J_2 \lambda + \epsilon_2$$
$$U_3 = X_3^T \beta_3 + J_3 \lambda + \epsilon_3$$
$$U_4 = X_4^T \beta_4 + J_{4+} \lambda + \epsilon_4$$

where, $X_1, X_2, \ldots X_4$ are the attributes in the utility functions; $\beta_1, \beta_2, \ldots \beta_4$ are the parameters to be estimated; $\epsilon_0, \epsilon_1, \ldots \epsilon_4$ are the error terms. The logsum term J_k links the choice of vehicle types to the household's decision of owning a certain number of vehicles. This way to connect different levels of choice decisions is similar to the nested logit structure, but not totally equivalent. In fact errors are assumed to be normally distributed in the first choice level and Gumbel in the second level.

The decision maker is assumed to be rational and to choose the alternative with the biggest utility. In our econometric setting we adopt a *probit* model for the discrete problem and therefore the error terms follow a multivariate normal distribution with full, unrestricted, covariance matrix.

For simplicity, let's assume that:

$$Y = Y_{disc}$$
$$X = (X_1, \ldots, X_4)$$
$$J = (J_1, \ldots, J_4)$$
$$\beta = (\beta_1, \ldots, \beta_4)$$
$$\epsilon = (\epsilon_0, \epsilon_1, \ldots, \epsilon_4)$$
$$\Sigma = \text{Covariance of the error term}$$

The likelihood of *one* observation can be expressed as follow:

$$P(Y=y|X,J,\beta,\lambda,\Sigma)=\int_{\mathbb{R}^5}\mathbb{I}\left(X_y^T\beta_y+J_y\lambda+\epsilon_y>X_j^T\beta_j+J_j\lambda+\epsilon_j\quad\forall j\neq y\right)\phi(\epsilon)d\epsilon$$

where ($\mathbb{I}()$) is a Boolean indicator of whether the statement in parentheses holds and $\phi(\epsilon)$ is the density of the error term. The dimension of the integral is equal to the number of alternatives (5). The subscript y indicates the predictors and coefficients of the chosen alternative and the subscript j indicates the other alternatives.

Since only differences in utility matter, the choice probability can be equivalently expressed as $(5-1)$ - dimensional integrals over the differences between the errors. Suppose we calculate the differences against alternative y, the alternative for which we are calculating the probability. Define:

$$\begin{aligned}\tilde{\epsilon}_{jy}&=\epsilon_j-\epsilon_y\\ \widetilde{V}_{jy}&=\left(X_j^T\beta_j+J_j\lambda\right)-\left(X_y^T\beta_y+J_y\lambda\right)\\ \tilde{\epsilon}_y&=\langle\tilde{\epsilon}_{1y},\ldots,\tilde{\epsilon}_{ky}\rangle\end{aligned}$$

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

Then:

$$P(Y=y)=\int_{\mathbb{R}^4}\mathbb{I}(\widetilde{V}_{jy}+\tilde{\epsilon}_{jy}<0\quad\forall j\neq y)\phi(\tilde{\epsilon}_y)d\tilde{\epsilon}_y$$

which is a (4)-dimensional integral over all possible values of the error differences. The difference between two normals is normal and the covariance of $\tilde{\epsilon}_y$ can be easily transferred from the covariance of ϵ (see Train (2009) for more details on the procedure to be adopted for the normalization of the covariance matrix).

3.2. The continuous choice sub-model

Regression is adopted to model the continuous part of the modeling framework or the decisions on the household vehicle mileage. The model deals with the overall car mileage per household, calculated as the sum over all the car miles in the household.

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

Usually, regression is solved by using the Ordinary Least Square (OLS) estimator (Weisberg, 2005), but the same problem can be expressed in the form of a likelihood function to be maximized (McCulloch et al., 2008, p. 117). Indeed, given β_{reg}, X_{reg} and σ^2, the probability ($P()$) of observing y_{reg} is given by the normal density function ($\phi()$):

$$P(y_{reg}|\beta_{reg},X_{reg},\sigma^2)=\phi(y_{reg}|X_{reg}^T\beta_{reg},\sigma^2)\tag{1}$$

The normal density is centered at $\hat{y}=X_{reg}^T\beta_{reg}$ and has variance σ^2. The assumption of a normal error term is not necessary for OLS; however, this is necessary in this context where we estimate simultaneously both the discrete and the continuous parts by Maximum Likelihood estimation method.

3.3. The integrated discrete–continuous model

In discrete–continuous choice models, we want to model Y_{disc} and Y_{reg} jointly in order to capture the correlation between them. In our framework, we allow the error term of the regression to be correlated with the error terms of the utilities in the probit. The specifications of the observable part of the utilities and of the regression remain the same, while the error terms are assumed to follow a normal distribution with covariance matrix Σ_5:

$$(\tilde{\epsilon}_1,\tilde{\epsilon}_2,\tilde{\epsilon}_3,\tilde{\epsilon}_4,\epsilon_{reg})\sim MN(0,\Sigma_5)$$

where, the dimension of the Multivariate Normal distribution is: 5 (the number of discrete alternatives) − 1 (for normalization purposes) + 1 (error term of the regression). And $\tilde{\epsilon}_j=\epsilon_j-\epsilon_0$ ($j=1,2,3,4$). Normalization is necessary because in discrete choice models the level and the scale of utility are irrelevant. Therefore, we take the differences with respect to zero-car alternative and we set the top left element of Σ_5 equal to 2. The estimation of a full covariance matrix provides essentially no interpretable information but allows the model to represent any substitution pattern between the discrete alternatives and to capture the correlation across the discrete and continuous decision variables.

Therefore, the probability of observing Y_{disc} and Y_{reg} can be derived as the product of the probability of observing Y_{reg} and the probability of observing Y_{disc} conditional on observing Y_{reg}.

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

This is a general result about conditioning with random variables (Rice, 2007, p. 88).

3.4. Estimation with simulation

Having calculated the probability of observing Y_{reg} ($P(Y_{reg})$) from (1):

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

which can be also written as:

$$P(Y_{reg}) = \phi\left(err|\mu = 0, \sigma^2 = \sigma_{reg}^2\right)$$

where $err = Y_{reg} - \widehat{Y}_{reg}$.

The problem here is to calculate the probability of observing Y conditional on observing Y_{reg}, which is a conditional probability of probit ($P(Y_{disc}|Y_{reg})$).

In multivariate normal distribution, if ($\mathbf{A}$,$\mathbf{B}$) follow a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Then ($\boldsymbol{A}|\boldsymbol{B} = \boldsymbol{B_1}$) follows a multivariate normal distribution with mean and variance

$$\boldsymbol{\mu_A} = \boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}^{-1}}(\boldsymbol{B_1} - \boldsymbol{\mu_2})$$
$$\boldsymbol{\Sigma_A} = \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}^{-1}}\boldsymbol{\Sigma_{21}}$$

Similarly, in our problem:

$$\begin{bmatrix} \tilde{\boldsymbol{\epsilon}}_y \\ \epsilon_{reg} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{0} \\ 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{disc} & \boldsymbol{\Sigma}_{disc,reg} \\ \boldsymbol{\Sigma}_{reg,disc} & \sigma^2 \end{bmatrix}\right)$$

Thus the conditional probability of probit can be written as:

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

where $\varphi(\tilde{\epsilon}_y)$ is the density function of a multivariate distribution and

$$\tilde{\epsilon}_y \sim \mathcal{N}\left(0 + \frac{\boldsymbol{\Sigma}_{disc,reg}}{\sigma_{reg}^2}(err - 0), \boldsymbol{\Sigma}_{disc} - \frac{\boldsymbol{\Sigma}_{disc,reg}\boldsymbol{\Sigma}_{reg,disc}}{\sigma_{reg}^2}\right)$$

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:

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

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

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

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

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. Unfortunately, it was not possible to calculate the inverse of the Hessian for the computation of the standard errors which were eventually calculated using Bootstrap re-sampling techniques.

4. Data: The 2009 US National Household Travel Survey

The primary data source used in this study is the 2009 US National Household Travel Survey (NHTS). The analysis is restricted to the area of Maryland, Virginia and District of Columbia (MD-VA-DC), for which 1420 complete observations are available. Household characteristics and information on each household vehicle, including year, make, model, and estimates of yearly miles per household, are the main variables extracted from the original dataset.

The 2009 NHTS data, however, does not include information on vehicle price, fuel efficiency, seating, engine, and other vehicle characteristics by vehicle make and model, which are important attributes for the analysis of factors correlated to vehicle type decisions. The vehicle characteristic data were obtained from a secondary data source the Consumer Reports, which provides the vehicle specification data on models tested within the past 10 years, having up to four model years by performance, crash protection, fuel economy, and specifications. This database also indicates the sale price or price of each new or used car.

Table 1 lists the basic statistics relative to the household sample. For the MD-VA-DC area, the average car ownership per household is 1.87, which is lower than the national average of 2.08 cars per household. The percentage of households without a car is 7.28%, higher than the national average of 4.8%. Average household income increases for households having no cars to households having two cars, but remains stable for household with 3 or 3+ cars. The number of cars in the household is highly associated with the number of adults and number of drivers in the family. More than half of the households who do not have car do not own a house. The land use variables, such as dummy of urban area, urban size, population density and housing density, have influence on the household car ownership decisions. The household with more cars are generally located in less dense or more rural areas. In the MD-VA-DC area, the average age of the household head is around 55 years; households with zero or one car have older household heads. The average education level in this area is college or bachelor degree, while households without a car have a much lower education level. The average age of the cars in the study area is 8.6 years, the majority of the cars are between 4 and 10 years old. The average yearly mileage traveled by a household is more than 20,000 miles per year. The total mileage traveled increases, as expected, with the number of vehicles in the household.

5. Model estimation results

The integrated framework accounts for the following conditional decisions:

1. number of vehicles owned by the household;
2. class and vintage for each vehicle, and
3. total number of miles traveled.

The first two choices are discrete variables while the third one is a continuous variable.

In our modeling system, we first calculate the ' from the class/vintage sub-model and then this variable is included in the utilities of the car ownership model.

Table 1
Data statistics.

	By number of cars									
Variables	0	1	2	3	4	min	max	median	mean	s.d.
Car ownership	7.28%	26.72%	43.49%	17.03%	5.48%	0	4	2	1.87	0.96
Hhld. income level[a]	6.47	10.98	14.55	15.29	15.99	1	18	16	13.21	5.30
Num. of adults	1.31	1.40	1.99	2.21	2.76	1	5	2	1.86	0.66
Num. of workers	0.53	0.69	1.16	1.44	1.63	0	4	1	1.06	0.83
Num. of drivers	0.75	1.26	1.98	2.32	2.83	0	5	2	1.80	0.77
Owned house	0.45	0.75	0.91	0.97	0.98	0	1	1	0.85	0.36
Urban area[b]	0.94	0.84	0.75	0.60	0.56	0	1	1	0.75	0.43
Urban size[c]	4.94	4.25	3.60	2.95	2.55	1	6	5	3.70	2.29
Use of PT	0.26	0.08	0.06	0.05	0.06	0	1	0	0.08	0.27
Age of hhld head	59.13	60.51	53.12	52.16	53.00	18	95	54	55.36	14.91
Female hhld head	0.78	0.64	0.52	0.55	0.44	0	1	1	0.57	0.49
Educ. hhld head[d]	2.78	3.40	3.67	3.52	3.33	1	5	4	3.49	1.21
Housing units per sq mile	8657	4187	1525	858	632	50	30,000	750	2593	4740
Population per sq mile	13,489	7652	3648	2327	1941	50	30,000	3000	5116	6820
Percent renter-occupied	51.33	31.18	18.83	17.06	14.33	0	95	20	23.95	22.78
Workers per square mile	2870	1899	1042	640	453	25	5000	350	1303	1660
AVMT (annual vehicle miles travel.)	0	10,361	23,890	35,781	48,728	0	208,642	19,173	21,922	17,986

[a] "Hhld. income levels" *-1: <$5000; 2:$5000 - $9999; 3: $10,000 - $14,999; 4:$15,000 - $19,999; 5:$20,000 - $24,999; 6:$25,000 - $29,999; 7:$30,000 - $34,999; 8:$35,000 - $39,999; 9:$40,000 - $44,999; 10: $45,000 - $49,999; 11:$50,000 - $54,999; 12:$55,000 - $59,999; 13:$60,000 - $64,999;14:$65,000 - $69,999; 15:$70,000 - $74,999; 16:$75,000 - $79,999; 17:$80,000 - $99,999; 18: = $100,000.*

[b] "urban area" if the home address in urban area *-1: yes; 2: no.*

[c] "urban size" is the size of the urban area in which the home address is located - *1: not in an urban area; 2: 50,000–199,999; 3: 200,000–499,999; 4: 500,000–999,999; 5: 1 million or more without subway or rail; 6: 1 million or more with subway or rail.*

[d] "education" *-1: less than High School graduate; 2: High School graduate; 3: some College or Associate's degree; 4: Bachelor's degree; 5: Graduate or Professional degree.*

In the continuous part of the model the cost variable has been estimated using an instrumental variable approach. This approach is required because when the household chooses which vehicle(s) it owns, it effectively chooses the operating cost of driving the selected vehicle(s) (Train, 1986). The operating cost (endogenous variable) is regressed on the exogenous variables; those include household income, number of drivers, number of workers, owned or rental house, dummy of urban area, urban size, age of the household head and the education level of the household head. The predicted values from these regressions are obtained and used as exogenous variables to explain the vehicle miles traveled.

The discrete–continuous model is estimated on the joint decisions of number of vehicles and annual miles traveled by each household in the NHTS sample.

5.1. Vehicle class and vintage models

The types of vehicle owned by each household are categorized by classes and vintages. Vehicle classification is based on vehicle size, function and brand loyalty (domestic or imported); we report in Table 2 the classification adopted by different agencies in the US to classify vehicles on the basis of their size and function.

In this study, classes are defined on the categories proposed by NHTS (automobile, sport car, SUV, pickup truck and VAN) and on a finer classification of automobile (small, compact mid-size, large and luxury) which was based on the variable make and model also available in the NHTS vehicle file. The twelve vehicle classes in our choice set are: (1) small domestic car; (2) compact domestic car; (3) mid-size domestic car; (4) large domestic car; (5) luxury domestic car; (6) small/compact import car; (7) mid-size import car; (8) large import car; (9) sports car; (10) minivan/van; (11) pickup trucks; (12) SUVs. The 10 vintages are pre-1999 and 2000 through 2008. We report in Fig. 2 the distribution of the vehicle classes and in Fig. 3 the average vehicle age for each category of the household vehicle ownership considered.

Therefore, each household is assumed to have a choice among 12 classes and 10 vintages for each car in the household. The choice set for the class vintage models is composed of all the possible combinations of vehicle classes and vintages for each vehicle in the household. For one vehicle households the total number of alternatives in the choice set is 120, while for two vehicle households the number of possible alternatives is 120×120. Choice sets for three and four vehicle households are constructed in a similar way and by considering all the possible combinations of cars (of different class and vintage).

Table 2
Vehicle classification schemes.

Source	Vehicle classification	Basis
NHTS (FHWA, 2009)	Automobile (including wagon), van, SUV, pickup, other truck, RV, motorcycle, other	Function
NTS (BTS, 2009)	Subcompact car, compact car, intermediate car, full car, light pickup, large pickup, small van, large van, small utility, large utility	Size and function
EPA (2009)	Cars: two-seater, sedan(minicompact, subcompact, compact, mid-sized, large), station wagon (small, midsize, large); Trucks: pickup (small and standard), van (cargo and passenger), minivans, SUV, special purpose vehicle	Size and function
Consumer Reports (2009)	Convertible, small car, sedan, wagon, SUV, minivan, pickup, sporty car	Size and function

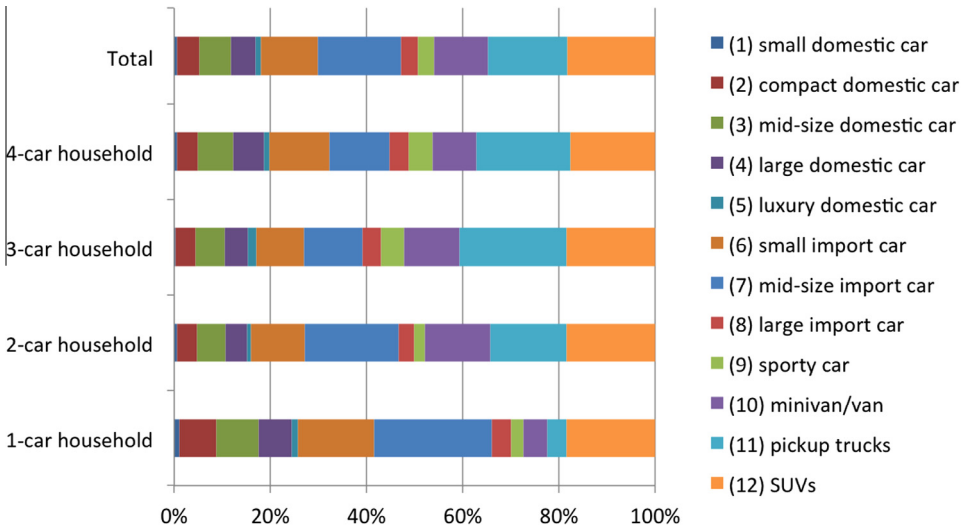

Fig. 2. Distribution of vehicle classes.

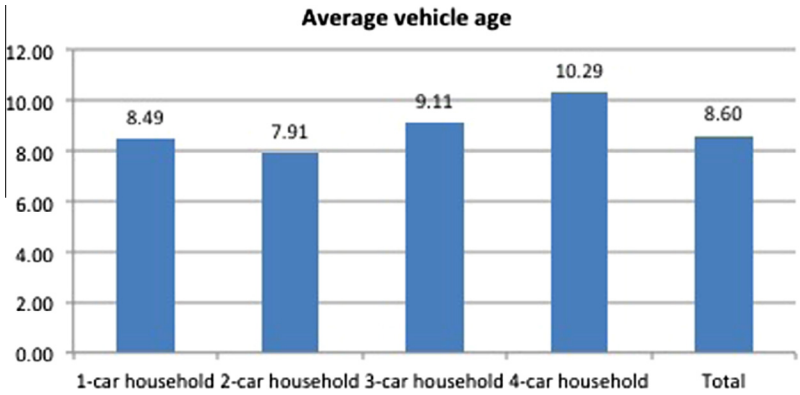

Fig. 3. Distribution of vehicle vintage.

Because of the large number of alternatives in the choice sets considered, estimation on the full set of alternatives is considered infeasible. The vehicle type sub-model is then estimated on a subset of alternatives which includes the chosen alternative and 20 alternatives randomly selected from the full choice set. Tests (Train, 1986) indicate that, once the number of alternatives exceeds a minimal threshold, the estimated parameters are not sensitive to the number of alternatives included in estimation. The definition of the independent variables for the one vehicle alternatives is straightforward; for the remaining alternatives the variables accounted for the characteristics of all the vehicles in the household. For example, for shoulder room, luggage space and price we considered the sum of the respective values for each vehicle in the household. For MPG we considered the MPG of the automobile(s) only and the difference between the max and the min MPG across all the vehicles available in the household. For the first variable, when more than one automobile was available we considered the average MPG; the second variable is an indication of the differences across vehicles in the household.

Results from the estimation of the class/vintage sub-model are reported in Table 3.

Table 3
Vehicle class and vintage choice sub-model: estimation results.

	One-car household		Two-car household		Three-car household		Four-car household	
Variable	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Sum of shoulder room	0.0329	5.6	0.04837	19.9	0.0471	18.4	0.0678	9.5
Sum of luggage space	0.2162	7.3	0.1563	5.8	-	-	0.1032	4.70
Log(n. of make/model in the class)	1.382	28.4	0.8015	40.1	0.8241	38.5	1.182	16.3
Auto's MPG	0.0026	0.3	-	-	-	-	-	-
Difference of max MPG and minMPG	-	-	0.0158	3.0	0.02465	3.1	0.02202	1.1
Dummy at least one foreign car	−0.4959	−7.6	-	-	-	-	−0.3361	−2.4
Dummy both foreign cars	-	-	1.031	14.0	-	-	-	-
Dummy one old vehicle	0.4515	8.3	0.4133	10.2	1.094	10.9	1.144	3.5
Dummy at least one SUV (Hhld 3 memb)	2.087	8.2	0.2794	3.0	-	-	−0.7771	−3.6
Dummy at least one pickup (Hhld 3 memb)	−1.332	−1.3	0.293	3.0	0.6034	5.1	0.5366	2.6
Dummy at least one van (Hhld 3 memb)	-	-	1.152	12.7	0.3519	3.0	−0.3945	−1.9
Dummy auto (Hhld 3 memb)	1.1	4.7	-	-	-	-	-	-
Dummy of both autos	-	-	−1.333	−4.7	-	-	-	-
D. of at least one sporty car	-	-	-	-	-	-	0.4704	2.1
Purchase price (Hhld income < 20 k)	−0.2037	−15.2	-	-	-	-	-	-
Purchase price (Hhld income = 20 k-40 k)	−0.0683	−8.5	-	-	-	-	-	-
Purchase price (Hhld income > 40 k)	−0.0069	−1.2	-	-	-	-	-	-
Purchase price (Hhld income < 45 k)	-	-	−0.157	−28.2	-	-	-	-
Purchase price (Hhld income = 45 k-80 k)	-	-	−0.0730	−17.8	-	-	-	-
Purchase price (Hhld income > 80 k)	-	-	−0.0457	−7.7	-	-	-	-
Purchase price (Hhld income < 55 k)	-	-	-	-	−0.1688	−22.8	-	-
Purchase price (Hhld income = 55 k-100 k)	-	-	-	-	−0.0953	−18.1	-	-
Purchase price (Hhld income > 100 k)	-	-	-	-	−0.0289	−6.1	-	-
Purchase price (Hhld income < 60 k)	-	-	-	-	-	-	−0.1841	−12.5
Purchase price (Hhld income = 60 k-100 k)	-	-	-	-	-	-	−0.1175	−11.2
Purchase price (Hhld income > 100 k)	-	-	-	-	-	-	−0.0678	−8.1
Number of observations	2995		4525		2015		545	
Initial Likelihood	−9118.34		−13,776.46		−6134.71		−1659.26	
Final Likelihood	−7859.95		−10,570.75		−3755.26		−788.66	
Rho-Squared	0.138		0.2327		0.3879		0.5247	

We observe that households prefer the vehicles that offer more space (shoulder room and luggage space). They are more likely to own car types for which more choices in terms of make and model combinations are available in the class (see the variable Log(n. of make model in the class)). The variable relative to the difference in MPGs is a proxy to test whether the households prefer owning cars that have similar engine sizes or not. The positive coefficient indicates that households with multiple cars have higher probability to own cars with different engine sizes. However, when it comes to the households with four or more cars, this factor becomes less significant. Households with only one car do not prefer foreign cars, however the preference is the opposite for the 2-car households. Households are in general holding old vehicles. One-car households are more likely to own a car if there less than three members in the family, whereas households with more than three members prefer to own SUVs. For the two-car households, the ones with three or more household members prefer to own one SUV/pickup/van rather than two autos. Similarly, the households with three cars are more likely to own a pickup or a van. However, households with four or more cars tend to own at least one pickup, but not SUVs or vans; they also have higher probability of owning a sports car. The coefficients related to vehicle purchase price are negative and significant; their magnitude is decreasing with the increase of household income. The lower income group is more sensitive to the vehicle purchase price, while higher income groups are found to be less sensitive to vehicle price (as expected). The estimates in Table 3 are then used to calculate the logsum of the class/vintage sub-model for inclusion into the discrete continuous joint model. The logsum values are calculated over the alternatives considered for model estimation.

5.2. Vehicle ownership and use integrated model

Results from the joint discrete–continuous model are reported in Table 4. Socio-demographic variables, land use variables and fuel cost enter the final specification of the joint discrete–continuous model estimated to predict car ownership and yearly vehicle miles driven by households in the Washington Metropolitan region.

Estimates from the discrete model can be interpreted as follows.

Table 4
Joint discrete–continuous model: estimation results.

Variable	Alternative	Coefficient	Std. error	*p*-value
Logsum class/vintage model	All	0.388	0.012	< 0.001
Constant	1 car	−2.863	0.237	<0.001
	2 cars	−8.700	0.098	<0.001
	3 cars	−14.404	0.188	<0.001
	4 cars	−21.385	0.201	<0.001
Income	1 car	−0.051	0.011	<0.001
	2 cars	0.056	0.006	<0.001
	3 cars	0.105	0.010	<0.001
	4 cars	0.111	0.012	<0.001
Num. of drivers	1 car	−0.010	0.007	0.182
	2 cars	3.223	0.079	<0.001
	3 cars	4.041	0.102	<0.001
	4 cars	4.432	0.092	<0.001
Gender (1: female, 0: male)	1 car	−0.129	0.551	0.815
	2 cars	−0.874	0.054	<0.001
	3 cars	−0.928	0.073	<0.001
	4 cars	−0.885	0.059	<0.001
Urban size	1 car	0.077	0.035	0.028
	2 cars	−0.120	0.074	0.103
	3 cars	−0.199	0.093	0.033
	4 cars	−0.201	0.084	0.016
Residential density (housing units per sq mile)	1 car	0.041	0.005	<0.001
	2 cars	−0.223	0.034	<0.001
	3 cars	−0.442	0.054	<0.001
	4 cars	−0.484	0.064	<0.001
Constant	Regression	1.130	0.102	<0.001
Income		0.129	0.005	<0.001
Own home		0.671	0.277	0.015
Gender (female)		−0.056	0.034	0.100
Res. density		−0.113	0.008	<0.001
Driving cost ($/mile)		−5.103	0.283	<0.001
Log-likelihood at zero	−5783.684			
Log-likelihood at convergence	−3349.812			
Number of observations	1420			

- Logsum of the vehicle type/vintage model. The logsum provides the utility that the household would obtain by its choice of vehicle type and vintage; it reflects the interdependence of household vehicle holding and vehicle type-vintage decisions. The estimated coefficient is significant and positive. It should be noted that the model proposed here is not of the nested logit type given that the first level of choice (vehicle ownership) assumes that the error term is normally distributed (probit) and the second level of choice assumes that the errors are i.i.d Gumbel (multinomial logit). Intuitively, a positive coefficient for the logsum is to be expected, given that an increase in the utility of the lower nest cannot produce a decrease in the utility of the alternative chosen in the upper level. More investigation are needed to demonstrate that the model is consistent with the utility maximization theory for all possible values of the explanatory variables. This is out of the scope of this paper.
- Household Demographics. Positive coefficients of household income indicate that households with higher income have higher tendency to own more vehicles and drive more (as expected). The coefficients of the number of drivers in the household are very significant, indicating that this factor has a strong effect on how many cars a household owns. The negative coefficients for household income and the number of drivers in the one-car alternative indicate that, the more drivers in the household, the less likely they will own only one car. For the remaining alternatives, the more drivers in the household, the higher the probability of owning more cars. Households with female household head are less likely to own one or multiple cars.
- Household location and built environment. Urban size is an indicator of the urbanization level in the area of the household location (Brownstone and Golob, 2009). The coefficients of urban size are significantly negative (except for one-car households) and have higher magnitude as the households own more cars, inferring that the households located in a more urban area have lower probability of having more cars. The households located in highly residential areas are more likely to own fewer cars.
- Residential density. High density levels are found to affect positively the utility of owning one car. Urban sprawl, characterized by low density, will induce households to own at least two cars to satisfy the mobility needs of its members.
- Vehicle use. When it comes to the continuous part of the model, we find that income and house ownership increase the household vehicle mileage. Parameters related to density, and gender (female) and fuel cost are all negative and significant.

5.3. Correlation between discrete and continuous decisions

A full correlation matrix has been estimated and results are reported in Table 5. Although the interpretation of the results obtained is difficult due to the need to drop one of the alternatives in order to make the coefficients of the probit model identifiable, it is interesting to note that the correlation terms between the discrete (the difference in utilities with respect to the 0 car alternative) and the continuous parts are all positive except for the first term.

To appreciate the effect of these correlation terms, we suppose that we observe a value of the VMT and we calculate the conditional covariance of the probit (difference in) error terms by using the following equation:

$$\overline{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

where Σ_{11} is the covariance matrix of the discrete part, Σ_{12} is the covariance across the discrete and the continuous parts, and Σ_{22} is equal to 1. The numerical terms obtained are reported in Table 6.

We can observe that the standard deviation of the errors decreases from 1 to respectively 0.99, 1.00, 0.92, and 0.93 (square root of the diagonal elements of the matrix in Table 6) for one, two, three and four cars alternatives. The improvement is low for the one and two car alternatives because the majority of the household in our sample owns one or two cars and the model is already good at predicting these choice outcomes. For the three and four cars alternatives we observe a decrease of respectively 8% and 7% in the standard deviation of the errors.

We also calculate the effect of observing VMT on the expectation of the probit residual.

$$\bar{\mu} = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$$

where μ_1 is the expectation of the unconditioned probit residuals, μ_2 is the residual of the regression and a is the VMT observed.

Suppose that the observed values for the VMT are -2 or 1 standard deviations away from the expected value; the effect of observing the continuous decision also shifts the expectation of the probit residuals as it can be seen in Table 7.

The effect of observing a VMT value less than the expected value shifts the probit residual of the one vehicle alternative by a positive quantity and the remaining residuals by a negative value. The opposite happens when we observe a VMT higher than the expected value. In order to quantify the effect of conditioning, we calculate the distribution of the residuals of the 1st and 4th alternatives. Before conditioning the probabilities of these residuals being less than zero were 50% each (by assumptions the residuals are distributed $N(0,1)$). Now if we observe that VMT is two standard deviations below the expected value, these probabilities are normally distributed with parameters $N(0.34, 0.99)$ and $N(-0.71, 0.93)$. Therefore, the probabilities that the residuals are less than zero become 37% and 78% for the 1st and 4th alternative respectively. If the VMT is one standard deviation above the expected value, these probabilities become 57% and 35%.

Table 5
Joint discrete–continuous model: correlation matrix.

$$\hat{\Sigma} = \begin{pmatrix} 1.00 & -0.63 & -0.60 & -0.79 & -0.17 \\ -0.63 & 1.00 & 0.19 & 0.51 & 0.07 \\ -0.60 & 0.19 & 1.00 & 0.94 & 0.39 \\ -0.79 & 0.51 & 0.94 & 1.00 & 0.36 \\ -0.17 & 0.07 & 0.39 & 0.36 & 1.00 \end{pmatrix}$$

Table 6
Conditional covariance matrix: discrete part.

$$\bar{\Sigma} = \begin{pmatrix} 0.97 & -0.62 & -0.53 & -0.73 \\ -0.62 & 1.00 & 0.16 & 0.49 \\ -0.53 & 0.16 & 0.85 & 0.80 \\ -0.73 & 0.49 & 0.80 & 0.87 \end{pmatrix}$$

Table 7
Shift in the expectations of the probit residuals.

VMT s.d.	*1veh.*	*2veh.*	*3veh.*	*4veh.*
−2	0.34	−0.14	−0.78	−0.71
1	−0.17	0.07	0.39	0.36

6. Model validation and application

For validation purposes, we re-estimated the model on 80% of the available observations in the dataset and we applied the model estimates on the hold out sample. The results show that the model does well in prediction. In Table 8 we report the actual choices, the choices predicted by the model and the difference between observed and predicted choices. The model slightly over-predicts vehicle ownership and mileage traveled.

The models estimated have been also applied to test policy scenarios; the variables of interest are income, density, driving cost, and vehicle price. The following four scenarios have been tested:

- Household income: 10% decrease, 5% decrease, 5% increase and 10% increase;
- Residential density: 50% decrease, 25% decrease, 25% increase and 50% increase;
- Driving cost: 50% decrease, 25% decrease, 25% increase and 50% increase;
- Vehicle price: luxury car + 10%, sports car + 20%, van + 10%, pick-up truck + 20%, suv + 10%, others no changes.

Results in Table 9 show the effects of those variables on both vehicle holding and mileage traveled. We calculate little variations in vehicle holding due to the changes in the scenarios considered. Changes in fuel cost have great effects on increasing/reducing vehicle usage, for example, vehicle usage will be reduced by around 20% when the driving cost increases by 50%. The elasticity to fuel cost is about −0.4; this value seems to be higher than elasticities calculated for the US in previous studies (see for example Litman (2013) for a very good overview of fuel elasticities). It should be noted that dataset that we used was collected in 2009 where fuel prices where particularly high and that the conditions of the US economy at that time were not particularly good. However, results obtained for density are consistent with those reported in the literature (Fang, 2008); increasing density by 50%will produce a reduction of 5.4% in annual vehicle miles traveled.

Table 8
Joint discrete–continuous model: validation results.

	Actual	Forecast	Difference
0-car household	8.44%	8.49%	0.05%
1-car household	25.31%	23.86%	−1.45%
2-car household	46.56%	46.09%	−0.48%
3-car household	15.31%	17.37%	2.06%
4-car household	4.38%	4.19%	−0.18%
Average vehicle ownership	1.82	1.85	0.03
Mileage	20,390	22,318	9.45%

Table 9
Joint discrete–continuous model: application results.

	0-car hh	1-car hh	2-car hh	3-car hh	4-car hh	Av. veh. ownership		Miles	
Actual	7.45%	22.29%	47.30%	17.79%	5.16%	1.91		22,603.00	
Income − 10%	7.42%	23.53%	47.15%	16.80%	5.11%	1.89	−1.19%	20,869.80	−7.67%
Income − 5%	7.44%	22.93%	47.23%	17.26%	5.14%	1.90	−0.62%	21,745.00	−3.80%
Income + 5%	7.45%	21.83%	47.22%	18.29%	5.21%	1.92	0.56%	23486.30	3.91%
Income + 10%	7.47%	21.33%	47.03%	18.93%	5.25%	1.93	1.17%	24343.20	7.70%
Density − 50%	7.44%	20.19%	47.81%	19.30%	5.27%	1.95	2.02%	23,831.40	5.43%
Density − 25%	7.44%	21.27%	47.60%	18.46%	5.22%	1.93	0.96%	23,203.20	2.66%
Density + 25%	7.43%	23.42%	46.82%	17.21%	5.13%	1.89	−0.91%	22,023.60	−2.56%
Density + 50%	7.40%	24.38%	46.42%	16.72%	5.08%	1.88	−1.68%	21,383.10	−5.40%
Fuel cost − 50%	7.43%	22.41%	47.11%	17.87%	5.17%	1.91	0.02%	27,056.80	19.70%
Fuel cost − 25%	7.45%	22.35%	47.22%	17.80%	5.18%	1.91	0.01%	24,835.10	9.88%
Fuel cost + 25%	7.44%	22.37%	47.24%	17.79%	5.17%	1.91	−0.02%	20,363.20	−9.91%
Fuel cost + 50%	7.43%	22.39%	47.12%	17.88%	5.18%	1.91	0.05%	18,173.00	−19.60%
Vehicle purchase price	7.36%	23.69%	48.91%	19.74%	0.30%	1.82	−4.63%	22,590.20	−0.06%

With reference to the vehicle price scenario, results indicate that the percentage of households owning 4 cars decreases significantly from 5.16% to 0.30%; the number of 0 car households remain unchanged, the percentage of 1 and 2 car households slightly increases by approximately 1% and the percentage of 3 car households increase by 2%. Overall, vehicle ownership diminishes by 4.63%. Annual vehicle miles traveled are not affected by the raising price of large or luxury vehicles.

7. Conclusions

Vehicle ownership plays an important role in the overall transportation planning process, due to its impacts on the environment, energy consumption, economic system and public health. This paper has presented a joint discrete–continuous model which aims at estimating vehicle ownership, class/vintage and use decisions. More specifically, a multinomial probit is used to estimate household vehicle ownership, a multinomial logit is used to estimate the vehicle class and vintage decisions, and a regression is used to estimate the vehicle usage decision. The unrestricted correlation pattern between the discrete and continuous parts is captured with a full variance–covariance matrix of the unobserved factors. The model is based on the properties of the multivariate normal distribution and estimated with Monte-Carlo simulation.

The 2009 National Household Travel Survey and a secondary dataset on vehicle characteristics are used to obtain estimates for the Washington DC Metropolitan area. The empirical results provide important insights into the determinants of vehicle holding, type and usage decisions of households in the study region, such as social demographics, household location, built environment, vehicle characteristics and operating cost.

The proposed econometric framework is applied to predict behavioral changes in response to the evolution of the society (income), built environment (density), and transportation policies (fuel cost). We found that the average number of vehicles owned in the Washington Metropolitan area won't be significantly affected in response to those changes. Vehicle use will be reduced by about 20% when fuel cost will increase by 50% and by less than 6% in response to a 50% increase in density. These results are consistent with previous research and attest that modern societies are still highly dependent on cars even in regions that are relatively dense.

The model system presented has a number of limitations. First the use of probit to model the discrete choice in the joint model limits the number of possible alternatives over which correlation between the discrete and the continuous parts can be accounted. Extension to a model that accounts for correlation across vehicle type and vintage choice and the number of miles driven with each car in the household is possible, but again the total number of discrete alternatives to be considered should be restricted to make the estimation of the probit model feasible. The model, in its present form, is not microeconomically consistent as the models proposed by Train (1986), de Jong (1990); however, this limitation can be relaxed by estimating a specific regression equation for households with different number of vehicles.

A number of possible research avenues are also possible. A comparison of the model proposed in this paper with ordered discrete–continuous modeling structure is under investigation. Additional policy variables, especially those related to built environment and proximity to public transportation, can be included in model specification in order to test different policy scenarios. More advanced simulation approaches based on Quasi Monte Carlo techniques will be considered to improve the estimation accuracy and reduce the computation costs. The authors are considering the possibility to apply numerical computation methods for the multivariate density function involved in the model formulation (Genz, 1992). These methods will hopefully lead to a more stable estimation of the Hessian and will allow the calculation of standard errors avoiding using Bootstrap which is computationally intensive. The model, developed in the context of a classic car ownership and use problem is general and can be applied to other decision frameworks involving discrete and continuous decision variables in transportation and related disciplines.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CMMI 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.

We thank the editor and two anonymous reviewers for their constructive comments, which helped us to improve the manuscript.

References

Bhat, C.R., 2005. A multiple discrete-continuous extreme value model: formulation and application to discretionary time-use decisions. Transp. Res. Part B 39 (8), 679–707.
Bhat, C.R., Pulugurta, V., 1998. A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. Transp. Res. Part B 32 (1), 61–75.
Bhat, C.R., Sen, S., 2006. Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. Transp. Res. Part B 40 (1), 35–53.
Bhat, C.R., Sen, S., Eluru, N., 2009. The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. Transp. Res. Part B 43 (1), 1–18.
Brannlund, R., Nordstrom, J., 2004. Carbon tax simulations using a household demand model. Eur. Econ. Rev. 48 (1), 211–233.
Brendemoen, A., 1994. Car Ownership Decisions in Norwegian Households. Statics Norway, Discussion Paper No. 116:1–25.
Brownstone, D., Golob, T.F., 2009. The impact of residential density on vehicle usage and energy consumption. J. Urban Econ. 65, 91–98.
Chamberlain, G., 1986. Asymptotic efficiency in semiparametric models with censoring. J. Econometrics 32, 189–218.
Dargay, J., Gately, D., 1997. Vehicle ownership to 2015: implications for energy use and emissions. Energy Policy 25 (14-15), 1121–1127.
de Jong, G.C., 1989. Some Joint Models of Car Ownership and Car Use; Ph.D. thesis. Faculty of Economic Science and Econometrics, University of Amsterdam.
de Jong, G.C., 1990. An indirect utility model of car ownership and car use. Eur. Econ. Rev. 34 (5), 971–985.
Dissanayake, D., Morikawa, T., 2010. Investigating household vehicle ownership, mode choice and trip sharing decisions using a combined revealed preference/stated preference nested logit model: case study in Bangkok metropolitan region. J. Transp. Geogr. 18 (3), 402–410.
Dubin, J., McFadden, D., 1984. An econometric analysis of residential electric appliance holdings and consumption. Econometrica 52 (2), 345–362.
Dubin, J., River, D., 1990. Selection bias in linear regression, logit and probit models. Sociol. Methods Res. 18 (2-3), 360–390.
Fang, H.A., 2008. A discrete–continuous model of households' vehicle choice and usage, with an application to the effects of residential density. Transp. Res. Part B 42 (9), 736–758.
Genz, A., 1992. Numerical computation of multivariate normal probabilities. J. Comput. Graphical Stat. J. Comput. Graphical Stat. 1 (2), 141–149.
Giblin, S., McNabola, A., 2009. Modelling the impacts of a carbon emission-differentiated vehicle tax system on CO_2 emissions intensity from new vehicle purchases in Ireland. Energy Policy 37 (4), 1404–1411.
Golob, T.F., van Wissen, L., 1989. A joint household travel distance generation and car ownership model. Transp. Res. Part B-Methodol. 23 (6), 471–491.
Hanemann, W.M., 1984. Discrete continuous models of consumer demand. Econometrica 52 (3), 541–561.
Hatzopoulou, M., Miller, E., Santos, B., 2007. Integrating vehicle emission modeling with activity-based travel demand modeling: case study of the Greater Toronto area, Canada. Transp. Res. Rec.: J. Transp. Res. Board 2011, 29–39.
Hausman, J., 1985. The econometrics of nonlinear budget sets. Econometrica 53 (6), 1255–1282.
Hayashi, Y., Kato, H., Teodoro, R.V.R., 2001. A model system for the assessment of the effects of car and fuel green taxes on CO_2 emission. Transp. Res. Part D: Transp. Environ. 6 (2), 123–139.
Heckman, J., 1978. Dummy endogenous variables in a simultaneous equation system. Econometrica 46 (4), 931–959.
Hensher, D.A., Barnard, P., Smith, N., Milthorpe, F., 1992. Dimensions of Automobile Demand; A Longitudinal Study of Automobile Ownership and Use. North-Holland, Amsterdam.
Kitamura, R., 2009. A dynamic model system of household car ownership, trip generation, and modal split: model development and simulation experiment. Transportation 36 (6), 711–732.
Litman, T., 2013. Understanding transport demands and elasticities. How prices and other factors affect travel behavior. Victoria Transp. Policy Inst., 1–76.
Mannering, F., Winston, C., 1985. A dynamic empirical analysis of household vehicle ownership and utilization. Rand J. Econ. 16 (2), 215–236.
McCulloch, C.E., Searle, S.R., Neuhaus, J.M., 2008. Generalized, Linear, and Mixed Models, 2nd ed. John Wiley & Sons, Hoboken, New-Jersey.
Paleti, R., Bhat, C.R., Pendyala, R.M., 2013. An integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. Transp. Res. Rec..
Rice, J.A., 2007. Mathematical Statistics and Data Analysis, 3rd ed. Duxbury, Belmont, California.
Roorda, M.J., Carrasco, J.A., Miller, E.J., 2009. An integrated model of vehicle transactions, activity scheduling and mode choice. Transp. Res. Part B-Methodol. 43 (2), 217–229.
Rouwendal, J., de Borger, B., 2009. Multiple Car Ownership, Fuel Efficiency and Substitution Between Cars. International Choice Modelling Conference, Harrogate, UK.
Schipper, L., 2011. Automobile use, fuel economy and CO_2 emissions in industrialized countries: encouraging trends through 2008? Transp. Policy 18 (2), 358–372.
Shay, E., Khattak, A.J., 2012. Household travel decision chains: residential environment, automobile ownership, trips and mode choice. Int. J. Sustain. Transp. 6 (2), 88–110.
TNO-Inro, 1999. De scenario verkenner, versie 1.2 deel 4: Het vervoervraagmodel. TNO Inro Afdeling Vervoer.
Train, K., 1986. Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand. MIT Press, Cambridge, Mass.
Train, K.E., 2009. Discrete Choice Methods with Simulation, 2nd ed. Cambridge University Press, Cambridge, England.
Vyas, G., Paleti, R., Bhat, C.R., Goulias, K.G., Pendyala, R.M., Hu, H.-H., Adler, T.J., Bahreinian, A., 2012. Joint vehicle holdings, by type and vintage, and primary driver assignment model with application for California. J. Transp. Res. Board 2302, 74–78.
Weisberg, S., 2005. Applied Linear Regression, 3rd ed. John Wiley & Sons, Hoboken, New-Jersey.