A Model System for Household Vehicle Holding, Type and Usage based on 2009 NHTS

Yangwen Liu, Jean-Michel Tremblay and Cinzia Cirillo*

University of Maryland
Department of Civil and Environmental Engineering
1173 Glenn Martin Hall
College Park, MD 20742
Phone: 301-405-6864
Fax: 301-405-2585
Email: ccirillo@umd.edu

* corresponding author

Submission Date: November 16, 2011

Word Count: 2.110 words (text) + 1 Figures + 2 Tables = 2,860 words

Abstract

This paper proposes a joint discrete-continuous model to estimate household choices on vehicle holding, type and usage. The model is estimated using the 2009 National Household Travel Survey and a secondary dataset on vehicle characteristics. The discrete components are respectively, multinomial probit for vehicle holding and multinomial logit for the vehicle type sub-models. A Multivariate Tobit allows unrestricted correlation between the discrete and the continuous parts. The estimated model is then applied to predict changes in the household decisions on vehicle holding and miles driven, in response to the evolution of social societies, living environment and transportation policies.

1 Introduction

Vehicle ownership plays an important role in transportation and land use planning. It is one of the key determinants of people's travel behavior, as it greatly impacts people's mode choice, frequency of trips, destination choice, trip timing, activity duration and trip chaining properties. Policy makers also make use of vehicle ownership models to find the best ways to reduce VMT (vehicle miles traveled), traffic congestion, gas consumption and air pollution.

A number of existing studies have investigated vehicle ownership choices with discrete-continuous models. The earliest generation of models belonging to this category were derived from conditional indirect utility function (e.g., Train, 1986; Hensher et al., 1992; de Jong, 1989a,b and 1991), which is based on micro-economic theory. Originally developed by Dubin and McFadden (1984), and Hannemann (1984), the basic concept is that the households choose the combination of vehicle ownership and vehicle usage that gives the highest utility. Roy's identity is applied to estimate vehicle usage and the relationship between the two modeling stages. Although based on single discreteness, this series of studies based on the indirect utility function are able to capture the interdependence between the vehicle holding and the corresponding mileage by means of observed variables. This elegant formulation is consistent with economic theory and simple to implement.

Multiple discrete-continuous extreme value (MDCEV) models, developed by Bhat (2005) and further applied in Bhat and Sen (2006) and 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. With the assumption that the error term distributed as iid extreme value, the probability function simplifies to an elegant and compact closed form, and collapses to MNL model for one car household. The MDCEV model recognizes multiple discreteness and is able to handle a large number of vehicle types. It well captures the interdependence between the vehicle type and the corresponding mileage and allows more complex specification forms 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. It has a fixed total mileage budget for every household so that it does not predict the change in total number of vehicles in response to the policy changes. In conclusion, the MDCEV model is consistent with random utility. It has good performance in capturing the trade-offs among the usage of different types of vehicles and can accommodate a large number of vehicle classifications.

More recently, Fang (2008) developed the BMOPT (Bayesian Multivariate Ordered Probit & 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) 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. Overall, the model is well suited for predicting the changes in the number of vehicles and miles traveled for each vehicle type category.

2 Methodology

This paper proposes a joint discrete-continuous model to estimate household choices on vehicle holding, type and usage. An unordered mechanism in the form of probit model is adopted to predict the number of vehicles per household; while a multinomial logit model is applied in the vehicle type sub-model to overcome the difficulty of estimation on a large number of alternatives. We take advantage of IIA (independency from irrelevant alternatives) property and we estimate the type-choice model on a sample of the alternatives drawn from the choice set (Train, 1986). The continuous part of the model estimates the annual miles driven by a household; Tobit model with an indirect utility function has been used here to overcome the limitation of OLS estimators. OLS provides biased estimators and fails to account for the difference between household with zero miles and the rest of the observations were the number of miles driven can be considered as a continuous random variable.

To summarize, the model system proposed estimate at household level, the total number of vehicles owned, the combination of the classes and vintages of these vehicles, and the total miles driven conditional on the number and types of vehicles (Figure 1).

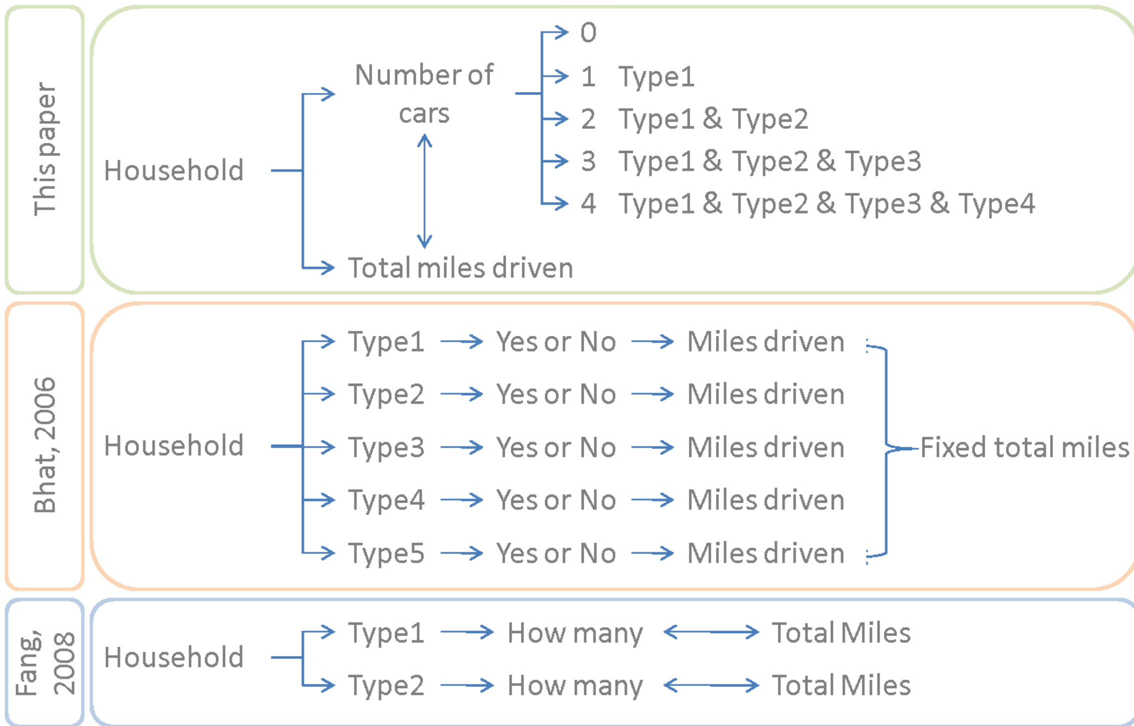

Figure 1 Model systems for joint discrete continuous decision on vehicle holding and miles driven (a comparison)

We assume that the choice set of vehicle holding sub-model includes zero, one, two, three and four or more vehicles. The types of vehicle owned by each household are categorized by classes and vintages; the vehicle classification schemes adopted are summarized in Table 1. This classification is based on the classes proposed in the 2009 National Household Travel Survey (NHTS) and in the 2009 National Transportation Statistics (NTS); it is mainly based on vehicle size, function, and brand loyalty (domestic or imported). Therefore, each household is assumed to have a choice among 12 classes and 10 vintages; 120 alternatives are in the final choice set for vehicle type and vintage sub-model. The twelve vehicle classes are: (1) small domestic car; (2) compact domestic car; (3) mid-size domestic car; (4) large domestic car; (5) luxury domestic car; (6) small import car; (7) mid-size import car; (8) large import car; (9) sporty car; (10) minivan/van; (11) pickup trucks; (12) SUVs. The 10 vintages are pre-1999 and the years 2000 through 2008. Because of the large number of alternatives, estimation of this model on the full set of alternatives is considered infeasible. The model is estimated on a subset of alternatives which includes the household's chosen alternative and 20 alternatives randomly selected from the 120 alternatives. 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. As a result, the utility function of vehicle type sub-model is written as:

$$U_{c_n|n} = \boldsymbol{\beta}_{c_n|n} \cdot \mathbf{X}_{c_n|n} + \varepsilon_{c_n} \qquad n = 1, 2, 3, 4$$

Where $V_{c_n|n}$ is a weighted sum of factors affecting the desirability to the household of owning a vehicle of class and vintage combination c_n given the household owns n vehicles. $\mathbf{X}_{\mathbf{c_n|n}}$ is a vector of characteristics of vehicles in class/vintage c_n and characteristics of household, and $\boldsymbol{\beta}_{c_n|n}$ is a vector of parameters to be estimated, ε_{c_n} is the unobserved error term which is iid extreme value distributed.

Assume the baseline utility of zero-car alternative is zero, the utility functions of vehicle holding sub-model are written as:

$$\mathrm{U}_0 = 0$$

$$\mathrm{U}_1 = \mathrm{ASC}_1 + \boldsymbol{\alpha}_1 \cdot \mathbf{X_n} + \lambda \cdot J_{c_1|1} + \varepsilon_1$$

$$\mathrm{U}_2 = \mathrm{ASC}_2 + \boldsymbol{\alpha}_2 \cdot \mathbf{X_n} + \lambda \cdot J_{c_2|2} + \varepsilon_2$$

$$\mathrm{U}_3 = \mathrm{ASC}_3 + \boldsymbol{\alpha}_3 \cdot \mathbf{X_n} + \lambda \cdot J_{c_3|3} + \varepsilon_3$$

$$\mathrm{U}_4 = \mathrm{ASC}_4 + \boldsymbol{\alpha}_4 \cdot \mathbf{X_n} + \lambda \cdot J_{c_4|4} + \varepsilon_4$$

The demand function of annual miles traveled is specified as a linear function of unit operation cost, household income and other explanatory variables, with an intercept and an error term (unobserved variable). Tobit model is applied to estimate the demand function for the vehicle usage sub-model.

$$\mathrm{M}_\mathrm{n}^* = \mathrm{Int} + \gamma \cdot p + \theta \cdot \mathrm{Y} + \psi \cdot X_\mathrm{m}^\mathrm{n} + \varepsilon_\mathrm{m}^\mathrm{n}$$

$$\mathrm{M_n} = \begin{cases} \mathrm{M}_\mathrm{n}^* & \text{if } \mathrm{U_n} > 0 \\ 0 & \text{if } \mathrm{U_n} \leq 0 \end{cases} \qquad \mathrm{n} = 1, 2, 3, 4$$

Where, M_n^* are the latent variables representing the uncensored annual miles traveled by a household. The actual vehicle usage $\mathrm{M_n}$ is equal to the latent variable M_n^* if and only if the household owns at least one car, and it is equal to zero if the household has no cars. γ, θ and ψ are parameters to be estimated.

The structure of the error terms in the submodels is specified as a multivariate normal with zero means and unrestricted covariance matrix. This specification enables the model to capture the correlation between discrete part of the model (vehicle holding sub-model and vehicle type submodel) and the continuous part of the model (vehicle usage sub-model).

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_m \end{pmatrix} \sim^{\mathrm{iid}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

Table 1 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 & function
EPA (2009)	Cars: two-seater, sedan(minicompact, subcompact, compact, mid-sized, large), station wagon (small, midsize, larg); Trucks: pickup (small & standard), van (cargo & passenger), minivans, SUV, special purpose vehicle	Size & function
Comsumer Reports (2009)	Convertible, small car, sedan, wagon, SUV, minivan, pickup, sporty car	Size & function

Notes: Vehicle function generally refers to engine size, wheel drive, and specialty.
BTS: Bureau of Transportation Statistics; EPA: Environmental Protection Agency; FHWA: Federal Highway Administration; NPTS: Nationwide Personal Transportation Survey; NTS: National Transportation Statistics.

3 Preliminary Findings

A preliminary analysis has been conducted with the aid of multinomial logit models for both vehicle holding choice and vehicle type choice with the connection in utility functions. A simpler form of vehicle usage submodel also has been tested, but without the unrestricted correlation in the error terms. The list of variables entering the model is presented in Table 2. The model has been validated by using an out-of-sample approach and has demonstrated good performance in prediction accuracy. The model has been applied to conduct sensitivity analyses with respect to a number of scenarios, and to calculate willingness to pay in terms of MPG, vehicle size and function. The joint discrete-continuous model of vehicle holding, type and usage is under investigation.

Tested scenarios analyze change in the household income, household structure, accessibility to public transportation, urbanization, aging society. The results indicate all the changes have light impacts on the average household vehicle holding. The improved public transportation system produces a greater reduction in vehicle ownership. Overall, the impacts in vehicle usage is greater than vehicle holding, meaning that people tend to drive less (or more) instead of changing vehicle holding choices.

Table 2 List of Variables

Vehicle holding submodel		Vehicle type submodel		Vehicle usage submodel	
Expected total utility from vehicle type choices	N	Vehicle purchase price (k$)	N	Unit operating cost (cents/mile)	N
Household income	C	Shoulder room	N	Household income	C
Number of adults	I	Luggage space of auto	N	Household size	I
Number of workers	I	Number of make and model in the class	I	Number of workers	I
Number of drivers	I	Foreign car	D	Number of drivers	I
Owned house	D	New vehicle	D	Number of children	I
Located in urban area	D	Old vehicle	D	Urban size	C
Urban size	C	Auto	D	Utilization of public transportation	D
Utilization of public transportation	D	SUV	D	Age of household head	I
Age of household head	I	Pickup truck	D	Gender of household head	C
Gender of household head	C	Van	D	Regional dummies *(Northeast, Midwest, South, West)*	D
Regional dummies *(Northeast, Midwest, South, West)*	D	Sporty car	D		
		MPG (miles per gallon)	N		

Notes: N = Numeric variable
C = Categorical variable
I = Numeric variable and Integer
D = Dummy variable

Reference

Bhat, C.R., and V. Pulugurta (1998), "A Comparison of Two Alternative Behavioral Mechanisms for Car Ownership Decisions", Transportation Research Part B, Vol. 32, No. 1, pp. 61-75.

Bhat, C.R., 2005. A multiple discrete-continuous extreme value model: formulation and application to discretionary time-use decisions. Transportation Research Part B 39 (8), 679–707.

Bhat, C.R., Sen, S., 2006. Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. Transportation Research 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. Transportation Research Part B 43 (1), 1-18.

Dubin, J.A., McFadden, D.L., 1984. An econometric analysis of residential electric appliance holdings and consumption. Econometrica 52 (2), 345–362.

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

Hannemann, M., 1984. The discrete/continuous model of consumer demand. Econometrica 52, 541–561.

Hensher, D.A, P.O. Barnard, N.C. Smith and F.W. Milthorpe, 1992. Dimensions of automobile demand; a longitudinal study of automobile ownership and use, North-Holland, Amsterdam.

de Jong, 1989a. 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., 1989b. Simulating car cost changes using an indirect utility model of car ownership and car use; paper presented at PTRC SAM 1989, PTRC, Brighton.

de Jong, G.C., 1991. An indirect utility model of car ownership and car use. European Economic Review, 34, 971-985.

Train, K., 1986. Qualitative choice analysis: Theory, econometrics and an application to automobile demand. The MIT Press, Cambridge, MA.