Household Vehicle Ownership by Vehicle Type: Application of a Multivariate Negative Binomial Model

by

Yong Zhao
Graduate Student Researcher
The University of Texas at Austin
yzhao@mail.utexas.edu

and

Kara Maria Kockelman
Assistant Professor of Civil Engineering
Clare Boothe Luce Professor of Civil Engineering
The University of Texas at Austin
kkockelm@mail.utexas.edu (Corresponding Author)

The University of Texas at Austin
6.90 E. Cockrell Jr. Hall
Austin, TX 78712-1076
Phone: 512-471-0210
FAX: 512-475-8744

Presented at the Transportation Research Board's 81st Annual Meeting, January 2002

ABSTRACT

Household vehicle ownership is a critical demographic characteristic influencing many aspects of travel demand. Using the 1995 NPTS data set, this research proposes a multivariate negative binomial model to investigate households' distinctions in vehicle purchases among passenger cars, SUVs, pickups, and minivans. This model structure is capable of capturing unobserved heterogeneity across the vehicle ownership levels and is equivalent to a multinomial distribution of the combinations of vehicles owned, conditioned on a negative binomial of the total vehicles owned. The results suggest that vehicle ownership decisions are firmly related to household size, income, population density (of zone of residence), and vehicle prices. For example, larger households exhibit a preference for minivans over SUVs, while SUVs are preferred to passenger cars, which are preferred to pickups. Higher-incomes households prefer SUVs most, and pickups least. In contrast, households living areas of lower population density favor pickups the most. The results of this work inform various vehicle-class-specific policy applications while providing a new statistical approach for the simultaneous vehicle-choice decision that households make.

KEY WORDS

household vehicle ownership, multivariate negative binomial model, Poisson mixture model

INTRODUCTION

Household vehicle ownership influences many aspects of travel demand. It often is used to categorize households for purposes of trip generation, and it is central to mode choices. It also affects location choices and urban form (– and is affected by these decisions as well). Vehicle ownership has a tremendous impact on personal trip-making, and, since driving generates external costs such as air pollution, crash fatalities, and congestion; vehicle ownership is a critical variable in policy analysis. It is well understood that different vehicle types impose different costs (e.g., emissions and crash distinctions *(1)*,*(2)*). Through a better understanding of household vehicle ownership decisions, planners can improve model forecasts and more effective policymaking in the areas of automobile purchase and use may occur, thus easing some of the negative impacts of automobile dependence. This paper investigates a new model of vehicle ownership, emphasizing household choices across major vehicle classes.

A large effort has been made to produce models of vehicle demand in the past twenty years, partly because of the oil crises in the late 1970's and early 1980's. Continual improvements in computation power permit more rigorous statistical estimation of more detailed and flexible models. Interest in vehicle ownership modeling appears to have slowed, perhaps because many modeling techniques have matured and improvements that can be made to existing techniques are generally marginal. Due to the relatively recent shifts in household vehicle ownership and use, towards light-duty trucks (i.e., pickups, minivans, and sport utility vehicles), and due to new energy issues and a focus on global warming policies, renewed consideration should be given to this important aspect of travel demand. Kockelman and Zhao *(3)* have investigated several dimensions of vehicle ownership and use, aiming to illuminate and identify the household and vehicle distinctions. However, their vehicle ownership model relied on a system of independent Poisson equations, across vehicle classes. The work presented in this paper moves beyond that model, to accommodate unobserved heterogeneity and intra-household ownership correlations.

The earliest research on automobile demand was performed at an aggregate level, usually from national time-series data (see, for example, *(4)*, *(5)*, and *(6)*). Almost all prior studies have examined the total number of auto purchases, ignoring consumers' choices of vehicle type. And an aggregate demand framework does not illuminate buyer distinctions. For example, Chamberlain's model of vehicle purchases *(7)* only examined the aggregate vehicle shares, by type.

or analysis of vehicle demand, discrete-choice models, such as multinomial and nested logit models, permit a disaggregate level of analysis. These models are generally more behavioral in nature and may be more policy-sensitive. Most research of this kind has used multinomial logit (MNL) models, ordered response models, or nested logit (NL) models to study individual households' choices by vehicle type and/or vehicle number (see, e.g., *(8)*, *(9)*, *(10)*, and *(11)*). Bhat and Pulugurta *(12)* compared a multinomial logit model to an ordered-response logit model when estimating a household's total number of vehicles owned. In their case, the MNL model performed somewhat better.

The application of combined discrete-continuous choice models further advanced vehicle ownership studies. (See, for example, *(13)* and *(14)*) These choice models assume that the households choose *jointly* the number of vehicles to own, the type of each vehicle (via an MNL model), and the amount each vehicle is driven. A vehicle's mileage (VMT) is clearly a key variable, acting as an output of interest and affecting vehicle operation costs and thus vehicle selection. These two types of decisions, the discrete choices of vehicle number and type and the continuous choice of VMT, are clearly interrelated and can not be considered one is made *before* the other. To accommodate this choice nature, ideally, unbiased estimation of these joint choice models would occur simultaneously, using full-information maximum likelihood methods *(15)*; but, generally, this approach is difficult so the research in this area has relied on sequential estimation. The joint choice model is usually separated into a set of sub-models and estimated sequentially; thus, the estimations lose efficiency in many cases. To reduce the complexity, their models are run separately for households owning one or two vehicles, and they neglect those households owning three or more vehicles. However, they split the major types of vehicles (i.e., passenger car, SUV, pickup, and minivan) into detailed classes (e.g., subcompact domestic passenger car).

More recently, Kockelman and Zhao *(3)* proposed an integer model of vehicle ownership across vehicle classes based on a set of simultaneous Poisson regression equations for the various numbers of different vehicle types owned, conditioned on the total. This model structure is equivalent to a multinomial distribution of the combinations of vehicles owned, with a Poisson distribution characterizing the total number of vehicles owned. In the presence of latent heterogeneity across households characterized by the same explanatory information, a more flexible model is desirable. For example, interaction of each rate parameters with a random component/error term provides a more flexible stochastic specification. If such terms were all gamma distributed and interacted multiplicatively, the result can be called a multivariate negative binomial (as used by Kockelman *(16)*).

This work relies on such a multivariate negative binomial specification to examine vehicle ownership. It also extends to a more general specification in which the interacted random components are normally distributed, and it discusses the maximum simulated likelihood (MSL) estimation method for such applications (where the likelihoods are intractable for integration but can be simulated). It employs the 1995 National Personal Travel Survey (NPTS) data in order to identify purchase distinctions across households and among vehicle types. The model structure and estimation methods are described in the following section. These are followed by a presentation of empirical results and conclusions.

MODEL FORMULATION

Based on a microeconomic theory of consumer behavior, the vehicle ownership problem may be posed as a household maximizing its utility function with respect to vehicle quantities and other consumption, subject to its income, or budget constraint. The vehicle prices are key variables.

In a particular period, the observed household vehicle ownership levels (Y_{ij}) of household *i* are a set of non-negative integer counts of different vehicle types *j* (e.g., passenger car, SUV, pickup, and minivan). Since the response values are in fact count data, the Poisson distributions are key contenders for stochastic specification of such behavior. One may expect that the mean of this distribution is represented by the optimal demand level (Y_{ij}^*), since over relatively long periods of time households can better achieve their average optimal demands. And these optima should be parameterized as functions of the explanatory variables $\vec{X}_i$, including vehicle prices, household income, and other attributes of the choice context. Virtually all vehicle ownership data sets provide information on the vehicle numbers that the households "currently" hold, rather than the numbers of vehicles that each household purchases during a window of time (e.g., five years). However, by assuming that the time window is rather long, one might still apply a rate model to present-ownership data.

Given an assumption of Poisson-distributed demands, the number of each vehicle type *j* that household *i* purchases, Y_{ij}, is drawn from a Poisson distribution with parameter λ_{ij}, as shown in Equation 1. Here, the exponential transformation ensures positive ownership levels. As Kockelman and Zhao *(3)* suggested, this set of Poisson random variables becomes simultaneous in nature if all optimal/mean demands are derived from a single indirect utility specification and thus share common parameters across their specifications.

$$\begin{aligned}
&\Pr(Y_{ij} = y_{ij}) = \frac{e^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!}, \quad y_{ij} = 0,1,2.... \\
&\textit{and} \quad \lambda_{ij} = Y_{ij}^* = \exp(\beta_j{}'\mathbf{X}_i)
\end{aligned} \tag{1}$$

where β_j is a vector of parameters to be estimated.

In the case of Kockelman and Zhao *(3)*, the price-over-income variable was theoretically confined to share the same parameter, making the estimation simultaneous. Other explanatory variables were household-specific and thus suggested preference differences across vehicle classes; there were no behavioral restrictions imposed on these other parameter values (e.g., monotonicity in consumption). The parameter values indicate the magnitudes of the household-specific explanatory variables (e.g., household size) determining the optimal demands across vehicle types, which makes it easy to interpreter the estimation results, just like any linear regression model results.

A typical Poisson regression model suffers from the fact that its variance is constrained to equal its mean. In empirical research, one often encounters situations where this restriction will be violated and overdispersion in fact exists (i.e., variance exceeds mean *(17)*). Generally, one can introduce an individual, unobserved effect into the Poisson mean to produce the following example:

$$\lambda'_{ij} = \exp(\beta_j\,'\mathbf{X}_i + \varepsilon_i) = \lambda_{ij}\exp(\varepsilon_i) = \lambda_{ij}\mu_i \tag{2}$$

where the disturbance ε_i reflects the heterogeneity that exists across households and their demand for vehicles. Then the unconditional probability distribution is:

$$p(y_{ij}) = \int \frac{e^{-\lambda_{ij}\mu_i}(\lambda_{ij}\mu_i)^{y_{ij}}}{y_{ij}!} g(\mu_i)du_i \tag{3}$$

where $g(\mu_i)$ is the density function of μ_i. Usually a gamma distribution is assumed for μ_i, which produces a negative binomial distribution, featuring overdispersion. As shown in Equation 2, the disturbance ε_i could be assumed to have zero mean if the model contains a constant term. In this case, the gamma term μ_i would have a mean of 1. With this normalization,

$$g(\mu_i) = \frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta\mu_i}\mu_i^{\theta-1} \tag{4}$$

where θ is the gamma's parameter. (Generally, a gamma distribution has two parameters, α and β, with mean $\alpha\beta$ and variance $\alpha\beta^2$; given the normalization of mean equal to one, here $\theta = \alpha = 1/\beta$).
This leads to the following probability function:

$$\begin{aligned} p(y_{ij}) &= \int_0^\infty \frac{e^{-\lambda_{ij}\mu_i}(\lambda_{ij}\mu_i)^{y_{ij}}}{y_{ij}!}\frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta\mu_i}\mu_i^{\theta-1}d\mu_i \\ &= \frac{\theta^\theta \lambda_{ij}^{\ y_{ij}}}{\Gamma(\theta)y_{ij}!}\int_0^\infty e^{-(\lambda_{ij}+\theta)\mu_i}\mu_i^{\theta+y_{ij}-1}d\mu_i \\ &= \frac{\theta^\theta \lambda_{ij}^{\ y_{ij}}\Gamma(\theta+y_{ij})}{\Gamma(\theta)y_{ij}!(\lambda_{ij}+\theta)^{\theta+y_{ij}}} \\ &= \frac{\Gamma(\theta+y_{ij})}{\Gamma(\theta)y_{ij}!}\left(\frac{\lambda_{ij}}{\lambda_{ij}+\theta}\right)^{y_{ij}}\left(\frac{\theta}{\lambda_{ij}+\theta}\right)^{\theta} \end{aligned} \tag{5}$$

which is a negative binomial distribution with mean λ_{ij} and variance $\lambda_{ij}[1+(1/\theta)\lambda_{ij}]$. The overdispersion can be measured by $1/\theta$. As illustrated via Equations 2 through 5, the negative binomial distribution assumption accommodates the effects of unobserved factors on each household's average demand levels (assuming one has begun with a Poisson assumption, for any specific household).

Using Equation 5's density function, a set of negative binomial models (one for each vehicle type) can be estimated simultaneously. This ownership model system can be considered as a multivariate negative binomial, while the marginal distribution for each vehicle type is a negative binomial. That is:

$$p(y_{i1}, y_{i2}, \ldots, y_{iJ}) = \prod_{j=1}^{J} p(y_{ij}) = \prod_{j=1}^{J}\left(\frac{\Gamma(\theta+y_{ij})}{\Gamma(\theta)y_{ij}!}\left(\frac{\lambda_{ij}}{\lambda_{ij}+\theta}\right)^{y_{ij}}\left(\frac{\theta}{\lambda_{ij}+\theta}\right)^{\theta}\right) \tag{6}$$

However, as Kockelman *(16)* showed, the use of the *same* gamma error term in all of a single household's demand functions allows for a cancellation of these terms in the probabilities of a multinomial (which is conditioned on a negative binomial for total demand). This feature is found by obtaining the joint distribution of ($y_{i1}, \ldots, y_{iJ}$) conditional on their sum.

$$p(y_{i1}, y_{i2}, ..., y_{iJ} \mid \sum_{j=1}^{J} y_{ij}) = \frac{\left(\sum_{j=1}^{J} y_{ij}\right)!}{\prod_{j=1}^{J} y_{ij}!} \prod_{j=1}^{J} p_{ij}^{y_{ij}} \tag{7}$$

where

$$p_{ij} = \frac{\lambda'_{ij}}{\sum_{j=1}^{J} \lambda'_{ij}} = \frac{\exp(\beta_j{}'\mathbf{X}_i + \varepsilon_i)}{\sum_{j=1}^{J} \exp(\beta_j{}'\mathbf{X}_i + \varepsilon_i)} = \frac{\exp(\beta_j{}'\mathbf{X}_i)\exp(\varepsilon_i)}{\sum_{j=1}^{J} [\exp(\beta_j{}'\mathbf{X}_i)\exp(\varepsilon_i)]} = \frac{\exp(\beta_j{}'\mathbf{X}_i)}{\sum_{j=1}^{J} \exp(\beta_j{}'\mathbf{X}_i)} = \frac{\lambda_{ij}}{\sum_{j=1}^{J} \lambda_{ij}} \tag{8}$$

Equations 7 and 8 produce a formulation similar to a logit model. Since demand for each vehicle type in a household takes a Poisson with mean $\lambda_{ij}\mu_i$ (conditioned on μ_i), the total demand should also take a Poisson with mean $\left(\sum_{j=1}^{J} \lambda_{ij}\right)\mu_i$, conditioned on the gamma term μ_i (with mean 1 and parameter θ).

Then, the unconditional distribution leads to the following:

$$\begin{aligned} p(y_{i1}, y_{i2}, ..., y_{iJ}) &= \frac{\left(\sum_{j=1}^{J} y_{ij}\right)!}{\prod_{j=1}^{J} y_{ij}!} \prod_{j=1}^{J} p_{ij}^{y_{ij}} \frac{\Gamma\left(\theta + \sum_{j=1}^{J} y_{ij}\right)}{\Gamma(\theta)\left(\sum_{j=1}^{J} y_{ij}\right)!} \left(\frac{\sum_{j=1}^{J} \lambda_{ij}}{\sum_{j=1}^{J} \lambda_{ij} + \theta}\right)^{\sum_{j=1}^{J} y_{ij}} \left(\frac{\theta}{\sum_{j=1}^{J} \lambda_{ij} + \theta}\right)^{\theta} \\ &= \frac{\left(\prod_{j=1}^{J} \lambda_{ij}^{y_{ij}}\right) \Gamma\left(\theta + \sum_{j=1}^{J} y_{ij}\right)}{\Gamma(\theta)\left(\prod_{j=1}^{J} y_{ij}!\right)\left(\sum_{j=1}^{J} \lambda_{ij}\right)^{\sum_{j=1}^{J} y_{ij}}} \left(\frac{\sum_{j=1}^{J} \lambda_{ij}}{\sum_{j=1}^{J} \lambda_{ij} + \theta}\right)^{\sum_{j=1}^{J} y_{ij}} \left(\frac{\theta}{\sum_{j=1}^{J} \lambda_{ij} + \theta}\right)^{\theta} \end{aligned} \tag{9}$$

The above negative binomial assumption has been used before. Rao *et al. (18)* modeled the number of children of each gender born to a pair of parents as symmetric binomials, where the total is a negative binomial. And Hausman, Leonard, and McFadden *(19)* sequentially estimated fishermen's recreational site choices as a multinomial conditioned on a fixed-effects Poisson (for the total number of trips).

Similarly, the vehicle demand model can be formulated as a multivariate negative binomial. Furthermore, under the special, same-gamma-term assumption, this multivariate negative binomial model is equivalent to a multinomial for the different vehicle types, conditioned on a negative binomial for the total number of vehicles owned.

In terms of estimating the model parameters, maximum likelihood estimation (MLE) of the various specifications proceeds in the usual manner. Clearly, the likelihood function derived from Equation 6 is relatively easier than that from Equations 9, given the fact that the two are equivalent as shown above. The likelihood function corresponding to Equation 6 is as follows:

$$L = \prod_{i=1}^{I} \prod_{j=1}^{J} \left(\frac{\Gamma(\theta + y_{ij})}{\Gamma(\theta) y_{ij}!}\right) \left(\frac{\exp(\beta_j{}'\mathbf{X}_i)}{\exp(\beta_j{}'\mathbf{X}_i) + \theta}\right)^{y_{ij}} \left(\frac{\theta}{\exp(\beta_j{}'\mathbf{X}_i + \theta}\right)^{\theta} \tag{10}$$

and the log-likelihood function is:

$$\begin{aligned} \ell = \sum_i \sum_j \{&\log[\Gamma(\theta + y_{ij})] - \log[\Gamma(\theta)] - \log(y_{ij}!) \\ &- \theta \ln[1 + \exp(\beta_j{}'\mathbf{X}_i)/\theta] - y_{ij} \log[1 + \theta/\exp(\beta_j{}'\mathbf{X}_i)]\} \end{aligned} \tag{11}$$

The log-likelihood function is maximized using standard techniques. All estimations and computations are completed using GAUSS programming language *(20)*. The analytical gradients of the log-likelihood function with respect to the parameters are coded. The standard errors of the parameters are obtained from the inverse of the Hessian matrix of the log-likelihood function. The results are presented next.

DATA AND RESULTS

The models are estimated on the sample of households that constituted the 1995 National Personal Transportation Survey (NPTS). This NPTS consists of over 42,000 households contacted between May 1995 and July 1996. The survey collected data on demographic and trip characteristics of the households. The specific NPTS data incorporated here as explanatory and response variables are shown in Table 1. After removing households with inconsistent and missing values on relevant variables, 32,596 households remained as the validation sample. The comparisons of variable distributions before and after record removal suggests that there are no significant distinctions in the full and culled samples. Thus, only complete records were used in the analysis of the various models presented here.

The NPTS households reported on over 75,200 vehicles across more than 200 makes and models. Using *Ward's Automotive Yearbook 1997(21)* and *Automotive News - 1997 Market Data Book (22),* four vehicle types were identified in the data set: passenger car, SUV, pickup, and minivan. Station wagons are considered as passenger cars and the SUV category includes all types of sport utility vehicle, from compact to full size. These are shown in Table 2. After removing missing values on relevant variables, 55,974 vehicles remain as the validation sample, associated with the 32,596 households.

In the multivariate negative binomial models developed here, the final demand specifications contain constants and household socio-economic parameters specific to each vehicle type. All specifications share a relative price parameter (i.e., the coefficient interacted with the variable "Vehicle Price/Income"), and they share the gamma parameter θ. The empirical results of this model are presented in Table 3.

The sign of the income-normalized price parameter is consistent with *a priori* expectations of the model. As the ratio of vehicle price to household income increases, households are less likely to purchase any kind of vehicle.

The relative magnitudes of household size parameters suggest that, households with larger sizes are more likely to purchase minivans and SUVs than passenger cars, while slightly less likely to buy pickups. This is consistent with expectations, since minivans and SUVs usually feature more seats and larger space than cars, and pickups have fewer seats. SUVs also are more likely, minivans slightly more likely, and pickups less likely than a passenger car to be owned as incomes (per household member) increase, *ceteris paribus*. Finally, pickups are more popular in lower-density environments, reflecting greater use of such vehicles for heavy work purposes, which are expected to be more common in relatively rural locations. All three light-duty truck (LDT) types examined here (i.e., minivan, pickup, and SUV) are more popular than passenger cars in lower-density areas; this may be related to a relative lack of parking issues for these larger vehicles in such areas.

The gamma parameter θ is relatively large. So the overdispersion measure, $1/\theta$, is rather small (0.0064). That suggests there is little unobserved heterogeneity within the vehicle ownership levels. Thus, the presented results are quite similar to those found in Kockelman and Zhao's simultaneous-Poisson model *(3)*. But the likelihood has improved, in a statistically significant way. A likelihood ratio test results in rejection of the hypothesis that there are no shared unobserved attributes among the vehicle-type decisions (the likelihood ratio test statistic in the comparison of the Poisson model and the negative binomial model is 5.76, which is larger than the chi-squared distribution with one degree of freedom at 97.5% level of significance). Thus, the addition of latent heterogeneity through inclusion of a gamma-error assumption and estimation of the parameter θ has allowed some valuable flexibility that should improve interpretation and application of this model. Yet the likelihood remains tractable.

MODEL EXTENTIONS

Note the gamma distribution assumption is for mathematical convenience and leads the Poisson mixture model to a negative binomial model, allowing straightforward estimation. Alternative distributional assumptions make the random effects Poisson model more flexible, but generally (and perhaps always) there is no closed form for the probability density function in Equation 3. For example, in assuming the random disturbance term ε_i is normally distributed with mean u and variance σ^2, μ_i will have a lognormal distribution with parameters u and σ. Normalizing *u* (the mean of ε_i) to equal zero produces the following density function:

$$g(\mu_i) = \frac{1}{\sqrt{2\pi}\sigma\mu_i} e^{-[\ln(\mu_i)]^2/(2\sigma^2)} \qquad \mu_i > 0 \tag{12}$$

In this case, the log-likelihood function of the estimation does not have a closed-form solution because of the integral in Equation 3; hence, it cannot be evaluated analytically. Thus, simulation techniques are needed to approximate the probabilities in the log-likelihood function and maximize the resulting simulated log-likelihood function. The underlying concept in such methods is to approximate the integration by computing the integrand at various values drawn from the appropriate distribution of the variable vector over which the integration is being carried out – and then taking the mean of the simulated integrands. This model was evaluated here, using such an approach.

The simulation approximations of the probabilities are the averages over the realizations from *R* random draws.

$$\tilde{P}(y_{ij}) = \frac{1}{R}\sum \tilde{P}^r(y_{ij}) \tag{13}$$

where $\tilde{P}^r(y_{ij})$ is the realization of the choice probability in the *r*th draw (*r*=1,2, …, *R*). $\tilde{P}(y_{ij})$ is an unbiased estimator of the actual probability $P(y_{ij})$, since its variance decreases as *R* increases. This estimator is smooth and twice differentiable, implying that conventional gradient-based optimization methods can apply to the simulated log-likelihood function. The simulated log-likelihood function is:

$$s\ell = \sum_i \sum_j \log\left[\tilde{P}(y_{ij})\right] \tag{14}$$

Generally, the maximum simulated log-likelihood (MSL) estimator is consistent, asymptotically efficient, and asymptotically normal (e.g., *(23)* and *(24)*). However, the logarithmic transformation of the choice probability in the likelihood function biased the MSL estimator away from the maximum likelihood estimator. As the number of simulations/repetitions increases, the bias decreases. For example, Brownstone and Train *(25)* suggested 250 repetitions would produce a rather negligible bias, in their context of a mixed logit model.

Using the lognormal specification, the relative lack of dispersion or unobserved heterogeneity in the data set ultimately meant that application of MSL estimation methods did not prove practical here: the simulations (involving 500 draws of μ_i, or equivalently 500 standard normal draws, multiplying the parameter σ and taking the exponential function, for each likelihood's estimate, over more than 32,000 observations) failed to converge. An even higher level of likelihood accuracy may be necessary for such cases. And this would be true in most any situation where one or more parameter values are so close to boundary values, yielding untenable distributions. (For example, in the vehicle-ownership case examined here, a negative dispersion factor is not permitted; a negative binomial's variance must exceed its mean.)

CONCLUSIONS

Household vehicle ownership is a key demographic characteristic that relates to and influences virtually all aspects of travel demand. This study proposed a particular multivariate negative binomial structure able to capture heterogeneity across the vehicle ownership levels. And, as shown, this structure is equivalent to a multinomial distribution of the combinations of vehicles owned conditioned on a negative binomial of the total vehicles owned.

By accommodating the correlation across different vehicle types, the multivariate negative binomial model structure used here provides a way to capture the underlying preference of vehicle types. The model structure is capable of handling panel data to study vehicle ownership decisions over time, though the data set used here provided only a cross-section.

The empirical application was based on the 1995 Nationwide Personal Transportation Survey data, and it produced evidence of household purchase differences across various vehicle types. Vehicle ownership decisions were found to be strongly influenced by household size, income, area population density, and vehicle prices. For example, SUVs and minivans were more likely to be purchased by higher income, larger households living in low-density environments. And pickups were more likely to be owned by lower-income households, living in low-density areas.

The results of this work inform various vehicle-class-specific policy applications. For instance, if a policy objective is reduction in SUV purchases (e.g., in congested and/or polluted regions), taxes on SUVs are expected to primarily affect the wealthy and are thus not generally regressive. And more stringent regulation of pickups will impact lower-income, smaller, less urban households more than other households.

Models based on stated-preference data and/or new-vehicle purchase data would be helpful for identifying more specific features of household preferences across vehicle types and more specific impacts of different policy instruments. And different statistical specifications may prove more valuable. One such option was tried here, using a lognormal random error term multiplicatively interacted with the Poisson rate parameter; however, the relative lack of dispersion in the data set stalled this effort. Another option is a two-way regression model of panel data, incorporating household-specific and period-specific random effects. In estimating such specifications, however, MSL will be necessary (as will a proper panel data set). In the interim, this work provides a new method for analyzing such data and suggests some behavioral tendencies that inform national and local transportation policy.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the University of Texas at Austin Department of Civil Engineering and the National Science Foundation CAREER Award program for their support of this research effort.

REFERENCES

1. Kockelman, K. To LDT or Not to LDT: An Assessment of the Principal Impacts of Light-Duty Trucks. In *Transportation Research Record 173*8, TRB, National Research Council, Washington, D.C., 2000, pp. 3-10.
2. Kockelman, K., and Y-J Kweon. Driver Injury Severity and Vehicle Type: An Application of Ordered Probit Models. Forthcoming in *Accident Analysis and Prevention*, 2001.
3. Kockelman, K. and Y. Zhao. Behavioral Distinctions: The Use of Light-Duty Trucks and Passenger Cars, *Journal of Transportation and Statistics* Vol. 3, No. 3, 2000, pp.47-60.
4. Chow, G. *Demand for Automobiles in the United States*. North-Holland, Amsterdam, 1957.
5. Kain, J. and M. Beesley. Forecasting Car Ownership and Use, *Urban Studies* Vol. 11, 1965.
6. Mogridge, M. The Prediction of Car Ownership. *Journal of Transport Economics and Policy*, Vol.1, 1967, pp. 52-74.
7. Chamberlain, C. A Preliminary Model of Auto Choice by Class of Car: Aggregate State Data. Discussion Paper, Transportation System Center, U. S. Department of Transportation, Cambridge, MA, 1974.
8. Manski, C.F. and L. Sherman. An Empirical Analysis of Household Choice among Motor Vehicles. *Transportation Research*, Vol. 14A, 1980, pp. 349-366.
9. Lerman, S. and C. M. Ben-Akiva. *A Behavioral Analysis of Automobile Ownership and Modes of Travel*. Report No. DOT-05-3005603, prepared by Cambridge Systematics, Inc., for the U.S. Department of Transportation and the Federal Highway Administration, 1976.

10. Mogridge, M. The Effect of the Oil Crisis on the Growth in the Ownership and Use of Cars. *Transportation*, Vol. 7, 1978, pp. 45-68.
11. Train, K. A Structured Logit Model of Auto Ownership and Mode Choice. *Review of Economic Studies*, Vol. 47, 1980, pp.357-370.
12. Bhat, C. R. and V. Pulugurta. A Comparison of Two Alternative Behavioral Choice Mechanisms for Household Auto Ownership Decisions. *Transportation Research*, Vol. 32B, No.1, 1998, pp. 61-75.
13. Mannering, F. and C. Winston. A Dynamic Empirical Analysis Of Household Vehicle Ownership And Utilization. *Rand Journal of Economics*, Vol. 16, No. 2, 1985, pp. 215-236.
14. Train, K. *Qualitative Choice Analysis: Theory, Econometrics and an Application to Automobile Demand*. The MIT Press, Cambridge, MA, 1986.
15. Mannering, F. and D. A. Hensher. Discrete/continuous Econometric Models and Their Application to Transport Analysis. Transport Reviews, Vol. 7, No. 3, 1987, pp. 227-244.
16. Kockelman, K. A Model for Time- and Budget-Constrained Activity Demand Analysis, *Transportation Research* Vol. 35B, No. 3, 2001, pp. 255-269.
17. Cameron, C., and P. Trivedi. *The Analysis of Count Data*. New York: Cambridge University Press, 1998.
18. Rao, B.R., S. Mazumdar, J. M. Waller, and C. C. Li. Correlation Between the Numbers of Two Types of Children in a Family. Biometrics, Vol. 29, 1973, pp. 271-279.
19. Hausman, J.A., G. K. Leonard, and D. McFadden, D. A Utility-Consistent, Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses due to Natural Resource Demand. *Journal of Public Economic*s, Vol. 50, 1995, pp.1-30.
20. Aptech. *GAUSS 4.0*. Aptech Systems. Maple Valley, Washington, 1999.
21. Ward's. *Ward’s Automotive Yearbook 1997*. Southfield, MI: Ward’s Communications,1997.
22. Automotive News. *Automotive News - 1997 Market Data Book*. Detroit, MI: Crain Communications Inc, 1998.
23. Lee, L. -F. On the Efficient of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models. *Econometric Theory*, Vol. 8, No. 4, 1992.
24. Hajivassiliou, V. A. and P. A. Ruud. Classical Estimation Methods for LDV Models Using Simulation. In *Handbook of Econometrics*, eds. R. F. Engle and D. L. McFadden. Vol. 4, pp. 2383-2441, North Holland, New York.
25. Brownston, D. and K. Train. Forecasting New Product Penetration with Flexible Substitution Patterns. Working paper, Department of Economics, University of California, Irvine, 1996.

LIST OF TABLES

Table 1. Description of variables used

Dependent variables			
		Mean	**SD**
Total number of vehicles	Total number of vehicles owned by household	1.72	0.93
Number of cars	Total number of passenger cars owned by household	1.16	0.81
Number of SUVs	Total number of sport utility vehicles owned by household	0.13	0.36
Number of pickups	Total number of pickup trucks owned by household	0.30	0.53
Number of minivans	Total number of minivans owned by household	0.13	0.35
Explanatory variables			
Household size	Number of household members	2.83	1.31
Population density	Population density of Census tract (persons per square mile)	3,858	5,306
Income per household member	Annual household income (1995 US$) divided by household size (where income is taken to be middle of class range)	$19,075	$13,561
Vehicle price/income	Average purchase price of new vehicle (based on 1997 sales data) divided by annual household income	0.74	0.14

Table 2. Shares of vehicle types

Vehicle Type	*No. of cases*	*Shares*
Passenger car	37885	67.7%
SUV	4120	7.4%
Pickup	9845	17.6%
Minivan	4124	7.4%
Total	55974	100%

Table 3. Results of simultaneous multivariate negative binomial model estimation

Variable	**Beta**	**SE**	**T-Stat**	**P-Value**
Vehicle Price/Income	-0.2307	0.0076	-30.26	0.000
Car:				
Constant	0.1793	0.0214	8.36	0.000
Household Size	0.0651	0.0044	14.90	0.000
Population Density	-0.1557	0.0082	-18.96	0.000
Income per HH Member	0.0153	0.0042	3.61	0.000
SUV:				
Constant	-2.7271	0.0557	-48.99	0.000
Household Size	0.2536	0.0119	21.25	0.000
Population Density	-0.4893	0.0319	-15.35	0.000
Income per HH Member	0.1761	0.0097	18.11	0.000
Pickup:				
Constant	-0.6368	0.0386	-16.48	0.000
Household Size	0.0657	0.0081	8.07	0.000
Population Density	-0.9517	0.0279	-34.09	0.000
Income per HH Member	-0.1172	0.0094	-12.53	0.000
Minivan:				
Constant	-3.1584	0.0556	-56.78	0.000
Household Size	0.4474	0.0102	43.96	0.000
Population Density	-0.4079	0.0305	-13.36	0.000
Income per HH Member	0.0284	0.0129	2.20	0.014
Gamma parameter θ	156.9	12.1093	12.96	0.000
Log Likelihood Function				
Constant only	-89028.26			
Convergence	-84396.92			
Pseudo R2	0.052			

Dependent variables: Total vehicles owned by type
Number of observations: 32,596 households