

CHARACTERIZING HOUSEHOLD VEHICLE FLEET COMPOSITION AND COUNT BY TYPE IN AN INTEGRATED MODELING FRAMEWORK

Venu M. Garikapati (*corresponding author*)

Arizona State University, School of Sustainable Engineering and the Built Environment
Room ECG252, Tempe, AZ 85287-5306.
Tel: (480) 965-3589; Fax: (480) 965-0557
Email: venu.garikapati@asu.edu

Raghuprasad Sidharthan

The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: (512) 471-4535; Fax: (512) 475-8744
E-mail: raghu@mail.utexas.edu

Ram M. Pendyala

Arizona State University, School of Sustainable Engineering and the Built Environment
Room ECG252, Tempe, AZ 85287-5306.
Tel: (480) 727-9164; Fax: (480) 965-0557
Email: ram.pendyala@asu.edu

Chandra R. Bhat

The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: (512) 471-4535; Fax: (512) 475-8744
Email: bhat@mail.utexas.edu

March 2014

Abstract

There has been considerable interest, and consequent progress, in the modeling of household vehicle fleet composition and utilization in the travel behavior research domain. The Multiple Discrete Continuous Extreme Value (MDCEV) model is a modeling approach that has been applied frequently to characterize this choice behavior. One of the key drawbacks of the MDCEV modeling methodology is that it does not provide an estimate of the count of vehicles within each vehicle type alternative represented in the MDCEV model. Moreover, the classic limitations of the multinomial logit model such as violations of the IIA property in the presence of correlated alternatives and the inability to account for random taste variations apply to the MDCEV model as well. A new methodological approach, developed to overcome these limitations, is applied in this paper to model vehicle fleet composition and count within each body type. The modeling methodology involves tying together a multiple discrete-continuous probit (MDCP) model and a multivariate count model capable of estimating vehicle counts within vehicle type categories considered by the MDCP model. The joint MDCP-multivariate count model system is estimated using a Greater Phoenix, Arizona travel survey data set. The joint model system is found to offer behaviorally intuitive results and provide superior goodness-of-fit in comparison to an independent model system that ignores the jointness between the MDCP component and the multivariate count component.

Keywords: vehicle fleet composition modeling, multiple discrete continuous probit (MDCP) model, multivariate count model, joint model estimation, vehicle type choice, activity-travel modeling

INTRODUCTION

Modeling the energy and environmental impacts of personal travel calls for the accurate representation and characterization of vehicle fleet composition and utilization choices of households in a region. As households may own a variety of vehicle types and utilize these vehicles to different degrees, the carbon footprint attributable to travel is intricately connected to the types of vehicles that households choose to own and the amount of miles that they drive the different vehicles in the household. Many metropolitan areas and policymakers are considering a host of policy, market, and technology strategies to enhance the share of alternative fuel and clean fuel vehicles as well as fuel efficient vehicles, with a view to reduce the adverse energy and environmental impacts of personal travel (1). Forecasting the potential impacts of such strategies requires the ability to accurately model vehicle type choices and utilization patterns under a wide range of scenarios. Air quality models that provide estimates of greenhouse emissions rely on information about the mix of vehicles in the fleet to compute emissions inventories. In the absence of accurate information about the vehicle fleet mix in a model region, the air quality model is prone to providing emissions estimates that are erroneous.

In light of the importance and value in modeling vehicle fleet composition and utilization, there has been considerable progress in the recent past in the modeling of these choice dimensions at the household level. Early studies in this arena did not explicitly consider the multiple discrete nature of the choice problem, i.e., households may own a variety or multitude of vehicle types, thus rendering the use of classic single discrete choice models to predict vehicle fleet composition of limited value. As a result, early studies focused on modeling household miles of travel (2), vehicle transactions including acquisition, disposal, and replacement (3), and vehicle ownership (count) and mileage (4). Hensher and Plastrier (5) developed a series of linked discrete choice models to explain household vehicle holdings and changes over time. Berkovec and Rust (6) developed nested logit models to study vehicle holdings of one-vehicle households, and thus circumvented the challenge of modeling multiple vehicle holdings in households. A key study by de Jong (7) involved the development of a disaggregate model system of vehicle type choice, duration, and usage. The system consists of separate models for vehicle type choice, vehicle holding duration, and annual mileage. This study attempted to fit this problem into a traditional single discrete choice modeling framework, resulting in the enumeration of a prohibitively large number of choice alternatives. Golob, et al (8) modeled the vehicle use of households using structural equation models. The structural equations models typically considered vehicle holdings of households as given (exogenous) and attempted to model usage (mileage). Yamamoto and Kitamura (9) developed models for actual and intended vehicle holding durations based on a panel data set collected in California; their model did not explicitly account for vehicle type choice or fleet composition. Similarly, Fang, et al (10) estimated a Bayesian Multivariate Ordered Probit and Tobit (BMOPT) model system of vehicle fuel efficiency choice and vehicle utilization measured in annual miles. The model used vehicle fuel efficiency as a proxy for vehicle type choice and did not explicitly consider the fleet mix.

In recognition of the dearth of work on vehicle fleet composition modeling and the limitations of the classic single discrete choice modeling methods in fully characterizing household fleet mix decisions, Bhat (11, 12) proposed and formulated a multiple discrete-continuous extreme value (MDCEV) modeling framework ideally suited to reflecting behavioral choice phenomena where individuals may choose multiple alternatives and utilize each of the chosen alternatives to different extents. The vehicle ownership modeling problem is an excellent example of a situation where individuals may choose multiple alternatives from a choice set and

then utilize the chosen alternatives to different degrees (as households may drive some vehicles in their fleet more or less than others). Bhat and Sen (13) formulated and estimated one of the first MDCEV models of vehicle type choice and utilization. Bhat, et al (14) updated the formulation and proposed a joint MDCEV-MNL model to fully characterize vehicle fleet composition and utilization behavior of households while including random coefficients and accommodating flexible substitution patterns across vehicles of a similar type. Eluru, et al (15) developed a joint model of household vehicle ownership, vehicle type choice, and vehicle usage and tied the entire model system with a model of residential location choice to examine how built environment attributes affected household vehicle fleet mix choices. Vyas, et al (16) further extended previous work in this domain to include assignment of an adult as a primary driver for each vehicle in the household fleet. The vehicle fleet composition and evolution simulator developed by Paleti, et al (17) utilizes a choice-occasion based approach to simulate household vehicle holdings and transactions over time.

Although the MDCEV modeling methodology constitutes a promising development in the modeling of vehicle fleet composition and utilization, it is not without its limitations. One of the key limitations of the MDCEV model is that the model does not return the exact count of vehicles that households own within each vehicle type category. Suppose a vehicle type category is defined by a combination of body type and age group as “cars 0-5 years old”. While the MDCEV model is able to indicate whether a household consumes (owns) cars 0-5 years old and the total miles that vehicle(s) in that category are driven (utilized), the model is not able to return the exact count of vehicles within the category. To overcome this problem, the vehicle type categories can be defined in such fine categories that it is virtually impossible for a household to own multiple vehicles in any of the categories. However, this may lead to the definition of a prohibitively large number of discrete choice alternatives in the MDCEV model. There is, essentially, a critical need for the ability to tie a count model to the multiple discrete-continuous framework so that counts of vehicles within each type may be accurately predicted. In addition to this key limitation, the MDCEV model has drawbacks similar to those of the traditional single discrete choice multinomial logit model including violations of the IIA property in the presence of correlated alternatives and the inability to reflect random taste variations in the behavioral choice phenomenon under investigation.

To overcome these limitations of the MDCEV model, Bhat, et al (18) recently formulated and developed a multiple discrete-continuous probit (MDCP) model that can be tied together with a multivariate count model in an integrated modeling framework. Just as the multinomial probit (MNP) model offers a methodology to overcome these limitations of the logit model, the MDCP model offers a methodology to overcome the limitations of the MDCEV model. The joint MDCP-multivariate count modeling methodology is applied in this paper to model vehicle fleet composition and utilization, and the number of vehicles (vehicle count) within each vehicle type alternative, so that the entire fleet mix of a household can be characterized. The model system is estimated on a 2008-2009 National Household Travel Survey sample drawn from the Greater Phoenix metropolitan area in Arizona.

A brief review of the modeling methodology is furnished in the next section. The data set used in the study is described in the third section. Model estimation results are furnished in the fourth section, together with goodness of fit measures that can be used to assess the efficacy of the joint MDCP-Count model. Concluding thoughts are offered in the fifth and final section.

MODELING METHODOLOGY

This section presents a brief overview of the multiple discrete-continuous probit (MDCP) – multivariate count (MC) modeling methodology employed in this paper. The complete details of the model formulation and methodology are provided in Bhat, et al (18) and hence only a brief synopsis is provided within the scope of this paper.

The use of the MDCP model in the current paper, rather than the multiple discrete-continuous extreme value (MDCEV) model (11, 12), is motivated by the need to tie the multiple discrete-continuous (MDC) model component (which caters to modeling the fleet composition dimension) with the multivariate count (MC) model (which handles the number of vehicles within each vehicle class dimension). For the MC model, a latent variable representation with normal error terms is used, and this facilitates the linkage with the MDCP model which is also based on a multivariate normal characterization of the error distribution. The model components are described further in this section.

The Multiple Discrete-Continuous Probit (MDCP) Model

The utility equation proposed by Bhat (12), where a consumer maximizes his/her utility subject to a binding budget constraint is:

$$\begin{aligned} \max U(\mathbf{x}) &= \frac{1}{\alpha_1} \psi_1 (x_1 + \gamma_1)^{\alpha_1} + \sum_{k=2}^K \frac{\gamma_k}{\alpha_k} \psi_k \left(\left(\frac{x_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right) \\ \text{st. } \sum_{k=1}^K p_k x_k &= E, \end{aligned} \quad (1)$$

where $x \geq 0$ is the consumption quantity (vector of dimension $K \times 1$ with elements x_k), and γ_k , α_k , and ψ_k are parameters associated with good k . In the linear budget constraint, E is the total expenditure (or income) of the consumer ($E > 0$), and p_k is the unit price of good k as experienced by the consumer. The utility function form in Equation (1) assumes that there is an essential outside good consumed by all behavioral units. α_k ($\alpha_k \leq 1$) and γ_k capture satiation effects and hence it is difficult to disentangle and uniquely identify the effects of both parameter vectors. Bhat (12) suggests estimating both a γ -profile and α -profile model specification (i.e., specifications in which only one of the parameter vector is free to be estimated, and the other vector is restricted) and choose the one that fits the data best. In addition to explaining satiation effects, γ_k also enables corner solutions (zero consumption) for alternatives, and hence is often preferred in empirical application contexts. ψ_k ($\psi_k > 0$) represents the stochastic baseline marginal utility; it is the marginal utility at the point of zero consumption. To complete the model structure, stochasticity is added by parameterizing the baseline utility as follows (see Bhat (12) for a detailed discussion):

$$\psi_k = \exp(\beta' z_k + \xi_k), \quad (2)$$

where z_k is a D -dimensional column vector of attributes that characterize good k , β is a corresponding vector of coefficients (of dimension $D \times 1$), and ξ_k captures the idiosyncratic (unobserved) characteristics that impact the baseline utility of good k . Bhat, et al (18) assumes

that the error terms ξ_k are multivariate normally distributed across goods k : $\xi = (\xi_1, \xi_2, \dots, \xi_K)' \sim MVN_K(\mathbf{0}_K, \Lambda)$, where $MVN_K(\mathbf{0}_K, \Lambda)$ indicates a K -variate normal distribution with a mean vector of zeros denoted by $\mathbf{0}_K$ and a covariance matrix Λ .

The Multivariate Count (MC) Model

Let y_k be the index for the count (say, of vehicles) for discrete alternative k , and let l_k be the actual count value observed for the alternative. Castro, et al (19) recast the count model for each discrete alternative using a special case of the generalized ordered-response probit (GORP) model structure as follows:

$$y_k^* = \eta_k, \quad y_k = l_k \text{ if } \psi_{k,l_k-1} < y_k^* < \psi_{k,l_k}, \quad l_k \in \{0,1,2,\dots\}, \quad (3)$$

$$\psi_{k,l_k} = f_{k,l_k}(s) = \Phi^{-1} \left[e^{-\lambda_k} \sum_{r=0}^{l_k} \left(\frac{\lambda_k^r}{r!} \right) \right], \text{ where } \lambda_k = e^{\zeta_k' s_k}.$$

In the above equation, y_k^* is a latent continuous stochastic propensity variable associated with alternative k that maps into the observed count l_k through the ψ_k vector, which is itself a vertically stacked column vector of thresholds $(\psi_{k,0}, \psi_{k,1}, \psi_{k,2}, \dots)'$. This variable, which is equated to η_k in the GORP formulation above, is a standard normal random error term. ζ_k is a vector of parameters (of dimension $\tilde{C} \times 1$) corresponding to the conformable vector of observables s_k (including a constant).

The η_k terms may be correlated across different alternatives because of unobserved factors. Formally, define $\eta = (\eta_1, \eta_2, \eta_3, \dots, \eta_K)'$. Then η is assumed to be multivariate standard normally distributed: $\eta \sim MVN_K(0_K, \Gamma)$, where Γ is a correlation matrix.

Joint Model System and Estimation Approach

An important feature of the proposed joint model system is that y_k (the count corresponding to discrete k) is observed only if there is some positive consumption of the alternative k as determined in the MDC model. That is, y_k is observed only if $x_k^* > 0$, and $y_k > 0$ in this case (y_k is not observed if $x_k^* = 0$). Thus, the proposed model resembles the hurdle model used in the count literature, albeit with the flexibility that the error components of the MDC model (ξ) and the MC model (η) can be correlated. As a result, the estimation approach involves the joint estimation of the MDC and MC model components. For details on the derivation of the likelihood expression, and the estimation procedures, please see Bhat, et al (18). The estimation process involves the evaluation of a multivariate normal cumulative distribution (MVNCD) function, which can be computationally challenging as the number of alternatives increases if typical simulation based approaches are used to approximate the function. As an efficient alternative to this, Bhat and Sidharthan (20) proposed a fast analytic approximation method called the maximum approximate composite marginal likelihood (MACML) approach to evaluate multidimensional integrals. We adopt the same approach in the current empirical context.

DATA DESCRIPTION

The data set used in this study is derived from the 2008-2009 National Household Travel Survey (NHTS), which is a survey of the nation's travel behavior conducted by the US Department of Transportation on a periodic basis. In the 2008-2009 version of the survey, individual jurisdictions were provided the option to purchase additional samples for their region to aid in model development and travel behavior analysis at the local and regional level. The Maricopa Association of Governments (MAG), the planning agency for the Greater Phoenix metropolitan area, purchased more than 4,400 such add-on sample households, thus obtaining a large sample household travel survey data set that could be used for model development and estimation purposes. Each respondent household was geolocated within a traffic analysis zone (TAZ). Using secondary TAZ and network skim data provided by MAG, the study team augmented the data set with an extensive set of built environment and accessibility variables. The built environment variables characterized the density and development patterns within the residential location TAZ of each household. The accessibility variables served as measures of the amount of employment in the region that could be accessed from the household's residential TAZ within certain travel time bands by auto (10 and 30 minute bands) and transit modes (30 and 60 minute bands). As built environment and accessibility variables are likely to be important predictors of vehicle fleet composition and utilization, it was considered important to augment the travel survey data set with such secondary variables.

Vehicle type choice was characterized by five distinct body type alternatives, namely, car, van, sport utility vehicle (SUV), pick-up truck, and motorbike. Further disaggregation of vehicle type alternatives can be done. For example, it is possible to consider a disaggregation of vehicle body types by age category and fuel type category. While such a disaggregation of vehicle type classification is appealing, the number of households with non-zero consumption in the various categories could easily become too thin to support model estimation of such a joint model system. Moreover, with the inclusion of a multivariate count model within the modeling framework, it is not necessary to try and create a disaggregate categorization where households consume (choose to own) only one vehicle. An annual mileage value (continuous dimension) is associated with each vehicle in the estimation data file.

In addition to the five vehicle choices, an outside good that is consumed by all households is introduced in the choice set to account for zero-vehicle households. This outside good is the non-motorized vehicle mileage. All households have to walk (and/or bicycle) for at least some non-zero distance over the course of an entire year. For households that report walk and bicycle trips in the survey, the reported non-motorized distance is scaled up to compute an annual non-motorized vehicle mileage. For households that report absolutely no walk and bicycle trips in the survey, a value for this consumption is estimated as 0.5 miles/person/day x 365 days/year x household size. This approximation is found to be reasonable and model parameter estimates are robust to alternative mileage computation schemes for the outside good (16). The result of the exercise is the creation of a data set where every household has six alternatives, one of which is consumed by all.

The data set was subjected to an extensive quality check and cleaning process to ensure that the data would be able to support the model estimation effort for this paper. The final cleaned data set includes 4,262 households owning 7,785 vehicles. The socio-economic and demographic characteristics of the sample are quite reasonable with no anomalies that might affect model estimation efforts. In the interest of brevity, a detailed tabulation of descriptive characteristics of the sample is not provided. On average, households owned 1.95 vehicles per

household. Average household size is 2.43 persons per household. There are 1.9 adults per household, and just about one worker per household (on average). There are 1.83 drivers and 0.53 children per household. A vast majority (95.8 percent) live in single family dwelling units and just about 85 percent of the respondent sample owns the home in which they live. An examination of the income distribution shows that 47 percent of the households have incomes that fall within the band of \$25,000 to \$75,000 per year. About one-fifth of the households have incomes greater than \$100,000.

Table 1 presents the vehicle ownership profile in the survey sample. The average age of pick-up trucks is larger than other vehicles in the fleet. Sport utility vehicles tend to be newer relative to the other vehicle types. The table also shows the age distribution and mileage distribution for the vehicles in each body class. Cars tend to be driven fewer miles, while SUVs and vans tend to be driven more miles. This is consistent with the notion that vans and SUVs tend to be vehicles owned and driven by family households located in the suburbs. Such households tend to drive longer distances, both because of household size and obligations and because of the potential distances that need to be traversed to access destinations. An examination of the age distribution suggests that SUVs tend to be newer vehicles, while cars and pick-up trucks tend to be older vehicles in the fleet.

MODEL ESTIMATION RESULTS

This section provides a summary of the model estimation results. The model estimation effort involved a systematic attempt at including explanatory variables such that the model offered behaviorally intuitive and statistically significant interpretations. Some variables were retained in the model specification even if they were statistically insignificant for considerations of behavioral sensitivity and intuitiveness.

MDCP Component

Estimation results for the MDCP component are furnished in Table 2. A γ -profile of MDCP model was estimated with one outside good. A baseline utility equation is estimated for each vehicle type. The values of the coefficient estimates indicate whether a certain characteristic or variable positively or negatively contributes towards ownership (consumption) of that vehicle type. Cars tend to be owned by smaller households evidenced by the negative coefficients on child presence and household size. It is to be expected that larger households, and households with children, have a higher baseline preference to own SUVs and vans; this is indeed supported by the model estimation results as single person households show a negative propensity to own larger vehicle types. Households at all income levels show a proclivity towards owning SUVs (presumably at different vintage levels), with the highest positive coefficient exhibited by the high income household category. Retired households with no children and those renting their single family housing unit have a lower preference to own SUVs. Home ownership, on the other hand, is positively associated with SUV ownership.

It is found that high income households shy away from owning vans, a finding that is consistent with expectations. It is found that multi-worker households are less likely to own vans, a finding that merits further investigation. Again, it is likely that these households are higher income households who prefer to own luxury vehicles. Retired households, single person households, very large households, and households in single family dwelling units all have a low preference to own pick-up trucks. As pick-up trucks tend to be more specialized and likely to be used as utility and work-related vehicles to haul cargo, it is not surprising that there is a general

disinclination to own pick-up trucks across the board. Households that own their residence and households with two workers show a positive inclination to own trucks.

Single person households and large households are less likely to own motorbikes (similar to pick-up trucks). It is likely that motorbikes and pick-up trucks are owned by two or three person households (not single person and not too large). Households in all income categories show an inclination to own motorbikes with those in the middle range exhibiting larger coefficients. Retired households, as expected, are less likely to own motorbikes. Households in rural area, households that own their single family dwelling unit, and two worker households are more likely to own motorbikes (again similar to pick-up trucks). As both pick-up trucks and motorbikes tend to be rather specialized vehicles, they appear to exhibit common traits.

The translation parameters are also furnished in the table. Van has the highest translation parameter, suggesting that it tends to be driven most. Vans tend to be multipurpose family vehicles, and are used for long distance family vacation trips. Thus this finding is consistent with expectations. SUV has the next highest translation parameter, once again consistent with expectations. These vehicles are more likely to be driven longer distances. Cars and motorbikes show lower translation parameters, presumably because these vehicles are driven shorter distances

Multivariate Count Component

Estimation results for the multivariate count model component are furnished in Table 3. The parameters in Table 3 refer to the elements of the ζ_k vector ($k=1,2,3,4$) embedded in the threshold functions. The constant coefficient in the ζ_k vector does not have any substantive interpretation. For the other variables, a positive coefficient in the ζ_k vector for a specific vehicle type k shifts all the thresholds toward the left of the count propensity scale for that vehicle type, which has the effect of reducing the probability of one vehicle of type k if the household decides to own vehicles (which is determined by the MDC component). That is, the household has a higher probability of owning multiple vehicles of type k , should it hold any vehicles at all of that type. On the other hand, a negative coefficient shifts all the thresholds toward the right of the count propensity scale, which has the effect of increasing the probability of one vehicle of type k (or decreasing the probability of multiple vehicles of type k), conditional on owning a vehicle of type k .

It is found that households with children are less likely to own multiple cars or multiple pick-up trucks. This is consistent with expectations as such households are likely to own a mix of vehicle types (such as a van and a car). Single person households are less likely to own more than one vehicle of any type. This is behaviorally intuitive as single person households would not generally own more than one vehicle. Households with three or more workers are more likely to own multiple cars, presumably because these households need multiple cars to meet their commuting needs. These household may also have a higher income, making it possible for them to own multiple cars.

An examination of the income dummy variables shows that lowest income households are less likely to own multiple cars, pick-up trucks, or SUVs, while high income households are more likely to own multiple vehicles – particularly in the car and SUV categories. These findings are consistent with expectations. Households in rural areas are more likely to own multiple pick-up trucks and less likely to own multiple cars or SUVs. As these households tend to own a pick-up truck, they are likely to own just one (if any) of the other vehicle types. Retired households are less likely to own multiple SUVs, a finding that is consistent with expectations.

These households would not have the need for multiple large vehicles; likewise, these households are less likely to own multiple pick-up trucks. On the other hand, retired households are more likely to own multiple cars. Households that own their single family dwelling unit are more likely to own multiple SUVs. Households of small size (one or two person) are less likely to own multiple pick-up trucks; as these tend to be specialized vehicles, it is unlikely that small households would need to own multiple vehicles of this category. The negative coefficients on these variables are indicative of this.

MODEL GOODNESS-OF-FIT AND ASSESSMENT

The model system is found to offer a good fit with the log-likelihood of the final model at convergence equal to -20989.96. The log-likelihood of the base model with only constants in the baseline utility, translation parameters, and constants in the count models is -22191.08. The log-likelihood ratio for the estimated model is 2402.23, which is significantly larger than the critical χ^2 value with 98 degrees of freedom at any level of significance.

In addition to examining model goodness-of-fit statistics, an assessment of the efficacy of estimating a joint model system (MDCP-MC) was performed. A complete calibration and validation of the model system is beyond the scope of this effort, but a simple assessment can be made by comparing the fit and indications offered by the joint model against those offered by an independent MDCP-MC model system where error correlations across the discrete-continuous and count components of the model system are ignored. The latter is akin to estimating two model components separately and then applying them in forecast mode in a sequential fashion – first, apply the MDCP model to predict the vehicle fleet mix by body type, and second, given the predictions of this model component, apply the count model for each body type consumed by a household to estimate the number of vehicles owned in each class. An independent MDCP-MC model system was estimated by setting all error correlations equal to zero. The coefficient estimates and goodness of fit statistics were compared between the two models. It was found that the independent model system offered coefficient estimates that were considerably different from those provided by the joint model and the goodness of fit was significantly worse. This comparison offered the first indication that the joint modeling approach is critical to modeling vehicle fleet composition, utilization, and count in a holistic framework.

An examination of the error covariance matrix (not presented in the interest of brevity) shows that there are a number of significant error correlations across the alternatives in the MDCP model component and the MC (count) model component. In general, it is found that within body type correlations are positive while cross body type correlations are negative. For example, the error correlation for the car alternative across the two model components is positive. This suggests that unobserved factors that contribute to car consumption in the MDCP component also contribute to owning more cars in the MC component. Such positive correlations are seen for all vehicle body types. This is consistent with expectations; it is very likely that unobserved attributes that contribute to greater mileage of a certain vehicle type will also contribute to a higher vehicle count for this class. A household whose members appreciate and desire comfortable and roomy vehicles are likely to choose and drive larger vehicles (such as vans and SUVs), and the same unobserved factors (desire for comfortable and roomy vehicles) will also contribute to such households owning multiple large vehicles. Across vehicle categories, error correlations are generally found to be negative, suggesting that there is an inherent inverse effect across body types. In the above example, the unobserved factors (desire for comfortable and roomy vehicles) are the very same factors that will negatively impact the

choice of smaller vehicles such as cars or vehicles with harsher rides such as pick-up trucks. Thus, the correlation between cars and vans (or SUVs and pickup trucks) is negative, both within the model component and across model components. The large number of significant error correlations leads to two noteworthy considerations. First, the joint model is capable of accounting for error correlations that may exist across choice dimensions. Ignoring such error correlations, when in fact they exist, will lead to inconsistent parameter estimates unsuitable for forecasting applications. Second, it points to the presence of unobserved factors that affect behavior and yet remain accounted in the model specifications. Qualitative research methods should be employed to identify these factors, and survey designs should be enhanced to measure these variables so that they may be included as observed covariates in the model specifications.

In addition to an examination of the error correlations, a comparison of the joint and independent model systems was performed by computing the log-likelihood value on a per household basis for a number of subsamples in the survey data set. If the log-likelihood in one model is higher than that in the other model, then the model with the higher log-likelihood may be considered superior from a statistical perspective. If the improvement in log-likelihood per household is seen across all (or nearly all) subsamples, then it indicates that such a model is likely better able to predict vehicle ownership, fleet composition, count, and utilization patterns for all socio-economic and demographic market segments. This comparison is presented in Table 4.

The comparison in Table 4 suggests that the joint model is consistently performing better than the independent model across all socio-economic and demographic market segments. The log-likelihood value per household is consistently higher (and therefore better) in the joint model relative to the independent model. There is only one subsample for which this does not hold true – household size=1. For single person household subsample, it is found that the independent model system is very marginally better. For every other market segment depicted in the table, the joint model offers a stronger fit as evidenced by the higher likelihood value.

CONCLUSIONS

The motivation for this paper stems from the growing interest in modeling household vehicle fleet composition and utilization behavior so that richer predictions of vehicle fleet mix and miles of travel by vehicle type can inform energy and environmental analysis. Recent work in this domain has focused on the development and application of techniques that recognize the multiple discrete-continuous nature of the vehicle fleet composition and utilization modeling problem. Recent work involving the use of the multiple discrete continuous extreme value (MDCEV) model has provided a promising approach to model vehicle fleet composition and utilization behavior. However, the MDCEV model is not able to offer predictions of the count of vehicles within each vehicle class, thus necessitating the statistically inefficient and behaviorally counter-intuitive stitching of a separate count model system (to the MDCEV model) capable of predicting vehicle counts. Such an approach ignores the presence of possible common unobserved factors affecting both the consumption of alternative vehicle types and the number of vehicles owned within each vehicle type. In order to overcome this limitation and account for such presence of common unobserved factors, this paper employs a joint model that incorporates a multiple discrete-continuous probit (MDCP) model component and a multivariate count (MC) model that takes the form of the generalized ordered probit model structure. The use of the probits in the two model components allows the use the multivariate normal distribution to characterize the error covariance structure accommodating correlations between the MDCP

component (that models vehicle type choice and mileage) and the MC component (that models the number of vehicles or vehicle count within each chosen vehicle class).

The model is estimated on a household travel survey data set of 4,262 households drawn from the Greater Phoenix region of Arizona in the United States. The model system is found to offer plausible parameter estimates with a host of socio-economic, demographic, and built environment variables affecting both the MDCP model of vehicle type choice and mileage, and the MC model of vehicle counts. The model is found to fit the data well, and a comparison of the goodness of fits between the joint model presented in this paper and an independent model that ignores error correlations across the choice dimensions (not presented in this paper) shows that the joint model consistently outperforms the independent model system. The comparison involved an examination of the per-household log-likelihood value between the two model systems; the model with the larger log-likelihood value offers the better fit to the data. The joint model is found to offer a better fit for all socio-economic and demographic market segments of interest. In addition, it was found that there were a number of significant error correlations across the two choice dimensions in the joint model. The presence of significant error correlations implies that there are common unobserved factors that affect both the MDC dimension (vehicle type choice and mileage) and the count of vehicles. For example, a person who is fun-seeking and gregarious in nature may like to own and drive sports cars. The unobserved attitudinal trait (being fun-seeking and gregarious) is likely to influence both the mileage (this person will likely drive more miles, thus representing a higher level of vehicle consumption/utilization), and the count of cars (as this individual might purchase additional sports cars that are fun to drive). There are likely to be a number of such attitudinal and contextual factors that are unobserved and yet influence both the multiple discrete continuous and multivariate count components of the model system.

The modeling of vehicle fleet composition, utilization, and counts by vehicle type is critical to performing energy and environmental impact analysis for a variety of policy, market, and technology scenarios. The introduction of vehicle fleet composition and utilization model systems is particularly made possible by the implementation of microsimulation-based activity-based travel demand model systems in practice. By accurately modeling vehicle fleet composition and usage patterns, planning agencies will be able to address energy sustainability and environmental concerns and implement policy actions that promote a more sustainable and energy friendly fleet mix and vehicle utilization pattern in the region. The MDCP-MC model system presented in this paper can be used to fill this modeling need. Future work in this domain should focus on including additional explanatory variables to make the model sensitive to policy, pricing, and market/technology changes. The data set used in this study did not support the inclusion of such variables. Household travel surveys should be designed to collect such data so that model systems capable of responding to a wide variety of scenarios can be estimated and deployed in practice. Future research efforts also should be aimed at reporting results of model validation and sensitivity analysis to demonstrate the ability of the model system to replicate base year conditions and respond in behaviorally intuitive ways when subjected to changes in input variables.

REFERENCES

1. Poudenx, P. The effect of transportation policies on energy consumption and greenhouse gas emission from urban passenger transportation. *Transportation Research Part A*, Vol. 42, No. 6, 2008, pp. 901–909.
2. Choo, S., and P. L. Mokhtarian. What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A*, Vol. 38, No. 3, 2004, pp. 201–222.
3. Mohammadian, A., and E. J. Miller. Empirical investigation of household vehicle type choice decisions. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1854, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 99–106.
4. 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.
5. Hensher, D.A., and V. L. Plastrier. Towards a dynamic discrete-choice model of household automobile fleet size and composition. *Transportation Research Part B*, Vol. 19, No. 6, 1985, pp. 481–495.
6. Berkovec, J., and J. Rust. A nested logit model of automobile holdings for one vehicle households. *Transportation Research Part B*, Vol. 19, No. 4, 1985, pp. 275–285.
7. de Jong, G.C. A disaggregate model system of vehicle holding duration, type choice and use. *Transportation Research Part B*, Vol. 30, No. 4, 1996, pp. 263–276.
8. Golob, T.F., D. S. Bunch, and D. Brownstone. A vehicle use forecasting model based on revealed and stated vehicle type choice and utilization data. *Journal of Transport Economics and Policy*, Vol. 31, No. 1, 1997, pp. 69–92.
9. Yamamoto, T., and R. Kitamura. An analysis of household vehicle holding durations considering intended holding durations. *Transportation Research Part A*, Vol. 34, No. 5, 2000, pp. 339–351.
10. Fang, H. A. A discrete-continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B*, Vol. 42, No.9, 2008, pp. 736–758.
11. Bhat, C. R. A multiple discrete-continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B*, Vol. 39, No. 8, 2005, pp. 679–707.
12. Bhat, C. R. The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B*, Vol. 42, No. 3, 2008, pp. 274–303.
13. Bhat, C. R., and S. Sen. Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. *Transportation Research Part B*, Vol. 40, No. 1, 2006, pp. 35–53.
14. Bhat, C. R., S. Sen, and N. Eluru. The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B*, Vol. 43, No. 1, 2009, pp. 1–18.
15. Eluru, N., C. R. Bhat, R. M. Pendyala, and K. C. Konduri. A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. *Transportation*, Vol. 37, No. 4, 2010, pp. 603–626.

16. Vyas, G., R. Paleti, C. R. Bhat, K. G. Goulias, R. M. Pendyala, H. Hu, T. J. Adler, and A. Bahreinian. Joint vehicle holdings, by type and vintage, and primary driver assignment model with application for California. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2302, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 74–83.
17. Paleti, R., N. Eluru, C. R. Bhat, R. M. Pendyala, T. J. Adler, and K. G. Goulias. Design of comprehensive microsimulator of household vehicle fleet composition, utilization, and evolution. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2254, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 44–57.
18. Bhat, C. R., S. K. Dubey, R. Sidharthan, and P. Bhat. A multivariate hurdle count data model with an endogenous multiple discrete-continuous selection system. Technical Report, 2013, Department of Civil, Architectural, and Environmental Engineering, University of Texas at Austin.
19. Castro M., R. Paleti, and C. R. Bhat. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B*, Vol. 46, No. 1, 2012, pp. 253–272.
20. Bhat, C. R., and R. Sidharthan. A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B*, Vol. 45, No. 7, 2011, pp. 940–953.

List of Tables

- TABLE 1 Vehicle Fleet Mix and Mileage Characteristics
- TABLE 2 Model Estimation Results (MDC Component)
- TABLE 3 Model Estimation Results (Count Component)
- TABLE 4 Comparison of Measures of Fit - Per Household Log-likelihood by Subsample

TABLE 1 Vehicle Fleet Mix and Mileage Characteristics

	Vehicle Body Type				
	CAR	VAN	SUV	PICK-UP	MOTORBIKE
Average Age	8.55	7.46	6.52	9.52	9.21
Average Mileage	10204.41	11317.66	11296.57	10722.98	3838.92
Number of Vehicles	3,997	635	1,537	1,376	240
Vehicle Type vs Annual mileage					
<i>Annual Mileage</i>					
0 - 4,999	27.5%	18.4%	21.1%	24.9%	71.3%
5,000 - 9,999	30.6%	31.3%	28.6%	29.4%	15.8%
10,000 - 14,999	21.4%	26.9%	26.2%	22.8%	7.9%
15,000 - 19,999	11.3%	13.9%	12.8%	12.5%	2.9%
≥ 20,000	9.1%	9.4%	11.3%	10.5%	2.1%
Total	100.0%	100.0%	100.0%	100.0%	100.0%
Vehicle Type vs Age					
<i>Age</i>					
0 - 5 Years	42.0%	40.9%	52.4%	34.2%	43.3%
6 - 11 Years	35.2%	44.3%	35.3%	39.5%	35.4%
≥ 12 Years	22.8%	14.8%	12.2%	26.3%	21.3%
Total	100.0%	100.0%	99.9%	100.0%	100.0%

TABLE 2 Model Estimation Results (MDC Component)

Baseline Utility							
	Explanatory Variables	Coef	t-statistic		Explanatory Variables	Coef	t-statistic
Car	Constant	1.20	9.29	PickUp	Constant	1.24	12.10
	Child presence	-0.13	-2.82		Child presence	-0.17	-3.04
	Household size	-0.19	-12.67		Household size	-0.14	-7.14
	High income household (\$75,000 - \$99,999)	0.06	1.50		Household size = 1	-0.24	-4.33
	Retired household (one/two person) with no children	-0.14	-3.81		Highest income household (\geq \$100,000)	-0.13	-3.38
	Single family housing unit (owned)	0.29	7.03		Retired household (one/two person) with no children	-0.30	-6.96
	Household residing in TAZ with low density	0.60	4.88		Single family housing unit	-0.42	-4.05
	Household residing in TAZ with medium density	0.63	5.18		Single family housing unit (owned)	0.58	9.12
	Household residing in TAZ with high density	0.61	4.97		Two worker household	0.13	3.52
SUV	Constant	0.93	8.94	Motorbike	Constant	-0.04	-0.21
	Household size = 1	-0.16	-13.71		Household size	-0.27	-8.93
	Lowest income household ($<$ \$25,000)	0.12	2.11		Household size = 1	-0.24	-2.15
	Low income household (\$25,000 - \$49,999)	0.36	6.26		Low income household (\$25,000 - \$49,999)	0.24	1.83
	Medium income household (\$50,000 - \$74,999)	0.29	4.59		Medium income household (\$50,000 - \$74,999)	0.42	3.04
	High income household (\$75,000 - \$99,999)	0.43	7.17		High income household (\$75,000 - \$99,999)	0.39	2.78
	Retired household (one/two person) with no children	-0.22	-5.59		Highest income household (\geq \$100,000)	0.24	1.68
	Single family housing unit	-0.21	-2.05		Retired household (one/two person) with no children	-0.38	-4.83
	Single family housing unit (owned)	0.51	8.20		Household in rural area	0.24	3.80
	Three worker household	-0.10	-1.12		Single family housing unit (owned)	0.49	4.28
Proportion of households in the lowest income quintile	-0.39	-3.38	Two worker household	0.11	1.62		
Van	Constant	0.31	2.53	Translation Parameters			
	Household size = 1	-0.11	-1.74	Non-motorized vehicle	0.00	0.00	
	Highest income household (\geq \$100,000)	-0.16	-2.97	Car	22.75	74.80	
	Single family housing unit	-0.26	-1.84	Van	73.11	27.89	
	Single family housing unit (owned)	0.39	5.33	SUV	36.54	53.41	
	Two worker household	-0.12	-2.45	PickUp	29.14	52.43	
	Three worker household	-0.34	-3.58	Motor	10.90	26.66	
	Household residing in TAZ with medium density	0.08	2.05				

TABLE 3 Model Estimation Results (Count Component)

Explanatory Variables			Estimates	t-statistic	Explanatory Variables			Estimates	t-statistic
Car	Constant		-0.12	-0.66	Van	Constant	-2.49	-12.94	
	Child presence		-0.40	-5.88		Three or more worker Household	0.74	1.45	
	Household size = 1		-1.26	-12.37	Constant	0.23	1.25		
	Household size = 2		-0.29	-4.75	Child presence	-0.51	-3.99		
	Zero worker household		-0.23	-3.38	Household size = 1	-1.24	-5.06		
	Three or more worker household		0.43	6.08	Household size = 2	-0.58	-4.64		
	Lowest income household (< \$25,000)		-0.14	-1.65	Zero worker Household	-0.19	-1.15		
	High income household (\$75,000 - \$99,999)		0.18	2.85	One worker household	-0.29	-3.01		
	Highest income household (>= \$100,000)		0.27	5.11	Lowest income household (< \$25,000)	-0.45	-2.36		
	Proportion of single family housing units in the TAZ		0.28	2.31	Retired household (one/two person) with no children	-0.41	-2.85		
	Retired household (one/two person) with no children		0.07	1.14	Household in rural area	0.14	1.52		
	Household in rural area		-0.07	-1.29	Single family housing unit	-1.46	-4.52		
	Single family housing unit		-0.51	-3.00	Single family housing unit (owned)	0.71	2.53		
	Single family housing unit (owned)		0.36	3.13					
SUV	Constant		-2.25	-6.47	Goodness of Fit (Joint Model)				
	Household size = 1		-2.02	-3.73	Log-likelihood of final model at convergence	-20989.97			
	Three or more worker household		0.47	3.82	Degrees of freedom of final model	112.00			
	Lowest income household (< \$25,000)		-0.44	-1.95	Log-likelihood of base model at convergence	-22191.08			
	Low income household (\$25,000 - \$49,999)		-0.66	-3.81	Degrees of freedom of base model	14.00			
	Highest income household (>= \$100,000)		0.51	5.74	Log-likelihood ratio	2402.23			
	Proportion of single family housing units in the TAZ		0.72	2.91	Critical Chi-squared value (98 df)	146.99			
	Retired household (one/two person) with no children		-0.34	-3.16					
	Household in rural area		-0.13	-1.47					
	Single family housing unit (owned)		0.56	2.20					

TABLE 4 Comparison of Measures of Fit - Per Household Log-likelihood by Subsample

Sample details	Number of households	Joint Model	Independent Model
<i>Full Sample</i>	4262	-4.9249	-5.0585
<i>Household Size</i>			
Household size = 1	992	-1.8513	-1.8384
Household size = 2	1830	-4.9422	-5.0938
Household size greater than 2	1440	-7.0204	-7.2320
<i>Household income</i>			
Lowest income household (< \$25,000)	759	-2.6603	-2.6880
Low income household (\$25,000 - \$49,999)	1210	-3.9486	-4.0106
Medium income household (\$50,000 - \$74,999)	800	-5.3207	-5.4750
High income household (\$75,000 - \$99,999)	640	-6.4789	-6.6743
Highest income household (>= \$100,000)	853	-6.7878	-7.0514
<i>Number of workers in household</i>			
Zero worker Household	1496	-5.8713	-6.0516
One worker household	1597	-4.9777	-5.0945
Two worker household	1011	-6.7623	-7.0083
Three or more worker household	158	-9.2027	-9.6031
<i>Household TAZ density</i>			
Lowest density	16	-8.5189	-8.6719
Household in TAZ with low density	620	-5.4100	-5.5464
Household in TAZ with medium density	2158	-5.1054	-5.2616
Household in TAZ with high density	1468	-4.4155	-4.5146
<i>Single family housing unit</i>			
No	179	-3.4535	-3.5164
Yes	4083	-4.9894	-5.1261
<i>Single family housing unit (owned)</i>			
No	642	-3.0915	-3.1396
Yes	3620	-5.2501	-5.3988
<i>Retired household (one/two person) with no children</i>			
No	2504	-5.7956	-5.9667
Yes	1758	-3.6848	-3.7650
<i>Household in rural area</i>			
No	3570	-4.6936	-4.8194
Yes	692	-6.1182	-6.2921