Alternative Approaches for Estimating Annual Mileage Budgets for a Multiple Discrete-Continuous Choice Model of Household Vehicle Ownership and Utilization: An Empirical Assessment

Bertho Augustin
Atkins North America
4030 West Boy Scout Blvd., Suite 700, Tampa, Fl 33607
Tel: (813)281-4576, Fax (813)974-2957, Email:betho.augustin@atkinsglobal.com

Abdul R. Pinjari*
Department of Civil & Environmental Engineering
University of South Florida, ENC 2503
4202 E. Fowler Ave., Tampa, FL 33620
Tel: (813) 974- 9671, Fax: (813) 974-2957, Email: apinjari@usf.edu

Naveen Eluru
Department of Civil, Environmental and Construction Engineering
University of Central Florida
12800 Pegasus Drive, Room 301D, Orlando, FL 32816
Tel: (407) 823-4815, Fax: (407) 823-3315, Email: naveen.eluru@ucf.edu

Ram M. Pendyala
School of Civil and Environmental Engineering
Georgia Institute of Technology
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: (404) 385-3754, Fax: (404) 894-2278; Email: ram.pendyala@ce.gatech.edu

* Corresponding author

Submitted for Presentation and Publication
94rd Annual Meeting of the Transportation Research Board
Committee: ADB40 Travel Demand Forecasting

Submitted: August 1, 2014
Revised submission: Nov 15, 2014

Word count: 6864(text) + 2tables x 250 + 1figures x 250 = 7614 equivalent words

ABSTRACT

This paper presents an empirical comparison of the following approaches to estimate annual mileage budgets for multiple discrete-continuous extreme value (MDCEV) models of household vehicle ownership and utilization: (1) The log-linear regression approach to model observed total annual household vehicle miles traveled (AH-VMT), (2) The stochastic frontier regression approach to model latent annual vehicle mileage frontier (AH-VMF), and (3) Other approaches used in the literature to assume annual household vehicle mileage budgets. For the stochastic regression approach, both MDCEV and multiple discrete-continuous heteroscedastic extreme value (MDCHEV) models were estimated and examined. When model predictions were compared with observed distributions of vehicle ownership and utilization in a validation data sample, the log-linear regression approach performed better than other approaches. However, policy simulations demonstrate that the log-linear regression approach does not allow for AH-VMT to increase or decrease due to changes in vehicle-specific attributes such as changes in fuel economy. The stochastic frontier approach overcomes this limitation. Policy simulation results with the stochastic frontier approach suggest that increasing fuel economy of a category of vehicles increases the ownership and usage of those vehicles. But this doesn't necessarily translate into an equal decrease in usage of other household vehicles confirming previous findings in literature that improvements in fuel economy tend to induce additional travel. In view of policy responsiveness and prediction accuracy, we recommend using the stochastic frontier regression (for estimating mileage budgets) in conjunction with the MDCHEV model for discrete-continuous choice analysis of household vehicle ownership and utilization.

1 INTRODUCTION

Analysis of household automobile ownership and utilization continues to be an important topic for transportation planners and researchers. Automobiles are the dominant mode of passenger travel in the United States (US) and many other countries. 95% of households in the US owned at least one automobile in 2009 and 87% of daily trips were made by automobiles (*1*). It is not surprising that the literature abounds with studies on this topic.

A variety of modeling approaches have been used for examining automobile ownership and utilization (see (*2*) for a review). Until a decade ago, standard discrete choice techniques (e.g., (*3-5*)) had been the mainstay of modeling vehicle ownership and/or vehicle-type choice decisions. These models, however, do not consider vehicle usage (mileage) endogenously in conjunction with vehicle ownership. Joint, discrete-continuous vehicle type choice and usage models have been formulated to address this issue (*6-8*).

More recently, there has been a growing interest in analyzing households' vehicle fleet composition (i.e., the types and number of vehicles owned by households) and utilization (i.e., the mileage accrued on each vehicle owned). This is motivated from an increasing interest in promoting policies aimed at encouraging the ownership and use of more energy-efficient and less polluting automobiles and for reducing the vehicle miles traveled. Evaluation of such policy actions requires modeling approaches that can provide credible forecasts of household vehicle fleet composition and usage under a variety of demographic, land-use, and policy scenarios.

An important aspect of household vehicle fleet composition is "multiple discreteness", where households own multiple types of vehicles depending on their preferences and travel needs (*9-11*). Recent literature has seen significant strides in developing model structures that explicitly recognize multiple discreteness in household vehicle holdings as well as model vehicle holdings and utilization in a joint fashion. Specifically, two distinct streams of modeling advances have been made: (a) random utility maximization-based multiple discrete-continuous choice models, particularly the multiple discrete-continuous extreme value (MDCEV) model proposed by Bhat (*9-11*), and (b) statistically-based discrete-continuous choice models that tie the discrete and continuous choice model equations for multiple vehicle categories into a joint statistical system based on error term correlations (*12-15*).

The MDCEV formulation has now been used in a number of studies on modeling household vehicle fleet holdings and utilization (*10, 11, 16-18*). The elegance of the MDCEV formulation, ease of estimation, and recent advances on applying the model for forecasting (*19*) makes it an attractive approach. Some transportation planning agencies have started implementing the formulation in their travel demand model systems for forecasting residential vehicle fleet mix and usage in their regions. Despite all these advances, a particular issue has been that most MDCEV formulations of vehicle holdings and utilization assume an exogenous (or fixed) total household mileage budget. The MDCEV model is used to allocate such exogenously available mileage budget among different types of vehicles to determine whether each type of vehicle is owned by the household and the extent to which each vehicle is utilized. Given the budget is exogenously determined, the MDCEV formulation does not allow the total household mileage to increase or decrease in response to changes in vehicle-specific attributes and relevant policies (e.g., increase in fuel economy of a particular vehicle type). Any such policies, with a fixed mileage budget, lead to only a reallocation of the mileage budget among different vehicle type categories.

The second stream of studies mentioned earlier on formulating statistically-based multiple discrete-continuous models (*12-15*) are not saddled with the above disadvantage. However, they are typically less theoretically-based and largely require computationally intensive simulation techniques to estimate and implement for simultaneous analysis of vehicle fleet holdings and usage while considering error correlations among all model components.

This budget issue is also addressed in the MDCEV formulations to a limited extent by including a non-motorized alternative along with the motorized vehicle alternatives in the formulation (*11*). The non-motorized alternative allows for the total mileage on motorized household vehicles to increase or decrease as a result of vehicle-specific attribute changes. This formulation, however, implies that a decrease/increase in total motorized vehicle mileage implies an equal amount of increase/decrease in non-motorized vehicle mileage, which may not necessarily be realistic.

More recently, Augustin et al. (*20*) proposed a stochastic frontier regression approach for estimating budgets for the MDCEV model in the context of analyzing individuals' daily out-of-home time-use choices. They conceive the presence of a latent frontier (or a maximum possible extent) of the resource being consumed (e.g., time, money, mileage). The frontier, in turn, is assumed to be the budget governing resource allocation among different choice alternatives. By design, the frontier is defined as greater than the observed total consumption, because the frontier is the maximum possible extent of the resource the consumer is willing to invest on the choice under consideration. Therefore, an *outside* choice alternative is introduced into the MDCEV model to represent the difference between the frontier value and the actual expenditure on all *inside* choice alternatives of interest. In other words, the *outside* alternative represents the portion of the frontier that is not expended for consumption. As such, when alternative-specific attributes change, the *outside* alternative acts as a "reservoir" to allow for the total consumption among the other choice alternatives to either increase or decrease. This concept potentially can be useful for estimating the budgets for MDCEV models of household vehicle ownership and utilization as well.

In view of the above discussion, the objective of this paper is to empirically compare alternative approaches to estimating budgets for MDCEV models of household vehicle ownership and utilization. Specifically, the following approaches are compared:

(a) The traditional log-linear regression approach to model observed total annual household vehicle miles traveled (AH-VMT),
(b) The stochastic frontier regression approach to model a latent annual household vehicle mileage frontier (AH-VMF),

(c) Introduction of a non-motorized alternative in the MDCEV model, as in (*11*), to allow for the AH-VMT to change in response to changes in vehicle-specific attributes (in this case the AH-VMT plus the household non-motorized mileage becomes the budget), and
(d) Assumption of an arbitrarily determined, uniform mileage budget for all households in the data

With the annual household mileage budgets estimated or assumed from each of the above approaches, we estimate MDCEV models of household vehicle holdings and utilization using household travel survey data from Florida. Each of these MDCEV models is applied on a validation dataset to assess the prediction accuracy (of MDCEV models) for different ways of estimating annual household vehicle mileage budgets. Furthermore, the influence of a policy scenario is simulated where the fuel economy is improved for selected categories of vehicles to understand how the different MDCEV models (with mileage budgets from different approaches) respond.

With mileage budgets from the stochastic frontier approach (i.e., AH-VMFs), in addition to examining the results of the MDCEV model, we assess if using the multiple discrete heteroscedastic extreme value (MDCHEV) model helps improve the predictions of household vehicle ownership and utilization patterns. This is because, by design, AH-VMFs are greater than AH-VMTs. As discussed later (in Section 3), the estimated AH-VMFs in the current empirical context are much larger in magnitude when compared to observed AH-VMTs. With such large budget values, it is likely that the MDCEV model might not appropriately allocate the mileage budget (AH-VMF) among different choice alternatives; particularly for the allocation of mileage budget between the *outside* alternative and *inside* alternatives. This issue potentially can be addressed by allowing for the variance of the random utility component of the *outside* alternative to be different from that of the *inside* choice alternatives. Therefore, we employ the MDCHEV model to allow for heteroscedasticity between the random utility specifications of the *outside* and *inside* alternatives. [1]

The remainder of the paper is organized as follows. Section 2 presents the modeling methodology. Section 3 presents the empirical analysis, including the data used, model estimation results, prediction assessments, and policy simulations. Section 4 concludes the paper.

2 METHODOLOGY

2.1 Stochastic Frontier Model for Annual Household Vehicle Mileage Frontier (AH-VMF)

In the stochastic frontier approach used in this paper, the annual mileage budget available to (or perceived by) a household is assumed to be a latent AH-VMF. While survey data provide measurements of AH-VMT, they do not provide measurements of AH-VMF. Stochastic frontier regression is employed to model such an unobserved limit households perceive.

[1] The MDCHEV model can be used to allow for heteroscedasticity across the different inside alternatives as well. However, we chose not to do so. This is because the intent of allowing heteroscedasticity in this study is specifically for allowing higher variance in the outside alternative utility term for addressing prediction issues arising from large budget values obtained from the stochastic frontier approach. For the same reason, we did not explore MDCHEV in conjunction with the other approaches used to estimate household mileage budgets.

Following Banerjee et al. (*21*), consider the notation below:
T_i = the observed AH-VMT for household *i*, assumed to be log-normally distributed;
τ_i = the unobserved AH-VMF for household *i*, assumed to be log-normally distributed;
v_i = a normally distributed random term specific to household *i*, with variance σ_v^2;
u_i = a non-negative random term assumed to follow half-normal distribution, with variance σ_u^2;
$\boldsymbol{X}_i$ = a vector of observable household characteristics; and $\boldsymbol{\beta}$ = coefficient vector of $\boldsymbol{X}_i$.

The unobserved AH-VMF (τ_i) of a household is assumed a function of demographics, location attributes, and fuel prices as:

$$\ln(\tau_i) = \boldsymbol{\beta}'\mathbf{X}_i + \nu_i \tag{1}$$

The unobserved AH-VMF can be related to the observed AH-VMT (T_i) as:

$$\ln(T_i) = \ln(\tau_i) - u_i \tag{2}$$

Note that since u_i is non-negative, the latent AH-VMF is by design greater than observed AH-VMT. Combining Equations (1) and (2) results in the following stochastic frontier regression equation:

$$\ln(T_i) = \boldsymbol{\beta}'\mathbf{X}_i + \nu_i - u_i \tag{3}$$

Once the model parameters are estimated (see *(22)* on estimating stochastic frontier models), using Equation (1), one can compute expected value of AH-VMF for household *i* as:

$$E\left[\hat{\tau}_i\right] = E\left[\exp\left(\hat{\boldsymbol{\beta}}'\mathbf{X}_i + \nu_i\right)\right] = \exp\left(\hat{\boldsymbol{\beta}}'\mathbf{X}_i + \hat{\sigma}_v^2/2\right) \tag{4}$$

The expected AH-VMF may be used as the mileage budget in the second-stage MDCEV model of vehicle type/vintage holding and usage.

2.2 MDCEV Model Structure for Household Vehicle Type/Vintage Holdings and Usage

A household is assumed to make its vehicle holdings and utilization choices (i.e., which vehicle types/vintages to own and how many annual miles to accrue on each vehicle type/vintage) for maximizing the following utility function (*9*):

$$U_i(\mathbf{t}_i) = \sum_{k=1}^{K} \gamma_{ik}\psi_{ik} \ln\left\{\left(t_{ik} / \gamma_{ik}\right) + 1\right\} + \left\{\psi_{io} \ln t_{io}\right\}, \tag{5}$$

subject to a maximum amount of annual miles the household is willing to travel (i.e., a household vehicle mileage budget constraint).

In Equation (5), $U_i(\mathbf{t}_i)$ is the total utility derived by a household *i* from its vehicle holdings and annual mileage choices. t_{ik} is the annual mileage on vehicle type/vintage category *k*, $\forall k = 1,2,...,K$. The term $\gamma_{ik}\psi_{ik} \ln\left\{\left(t_{ik} / \gamma_{ik}\right) + 1\right\}$ represents the utility accrued by driving t_{ik} miles on vehicle type/vintage category *k*, $\forall k = 1,2,...,K$. The term $\left\{\psi_{io} \ln t_{io}\right\}$ is used in the utility function to include t_{io}, an *outside* alternative representing the difference between the mileage budget and the sum of annual miles travelled on all household vehicles $\sum_{k=1 \text{ to } K} t_{ik}$. This can be viewed as the unexpended portion of the mileage budget.

The specification of the annual household vehicle mileage constraint depends on the approach used for the total available mileage budget. As discussed earlier, we tested three different approaches. The first approach is the stochastic frontier approach, where the expected value of AH-VMF is used as the budget; i.e., the constraint then becomes $\sum_{k=1 \text{ to } K} t_{ik} + t_{i0} = E\left[\hat{\tau}_i\right]$. As discussed earlier, while changes in vehicle-specific attributes do not allow for the mileage frontier ($E\left[\hat{\tau}_i\right]$) to change, the AH-VMT (= $\sum_{k=1 \text{ to } K} t_{ik}$) can potentially change because t_{io} serves as a "reservoir" to hold mileage for decreasing or increasing AH-VMT.

The second approach is to use AH-VMT, which is observed in the data for model estimation purposes and can be estimated via a log-linear regression model for prediction purposes. In this case, the budget constraint would be $\sum_{k=1 \text{ to } K} t_{ik} = T_i$, where T_i is the AH-VMT for household *i* ($E\left[\hat{T}_i\right]$ is used for prediction purposes). Note that in this specification the t_{io} term is specified as zero because the sum of annual miles on all household vehicles or AH-VMT ($\sum_{k=1 \text{ to } K} t_{ik}$) is itself assumed as the budget.

The third and fourth approaches specify or assume a budget amount greater than the observed AH-VMTs in the sample. Therefore, in both these approaches, similar to the stochastic frontier approach, the t_{io} term is positive.

In the utility function in Equation (5), ψ_{ik}, labelled the baseline marginal utility of household *i* for alternative *k*, is the marginal utility of mileage allocation to vehicle type/vintage *k* at the point of zero mileage allocation. Between two choice alternatives, the alternative with greater baseline marginal utility is more likely to be chosen. In addition, ψ_{ik} influences the amount of miles allocated to alternative *k*, since a greater ψ_{ik} value implies a greater marginal utility of mileage allocation. γ_{ik} allows corner solutions (i.e., the possibility of not choosing an alternative) and differential satiation effects (diminishing marginal utility with increasing consumption) for different vehicle types/vintages. When all else is same, an alternative with a greater value of γ_{ik} will have a slower rate of satiation and therefore a greater amount of mileage allocation (see *(9)* for more details).

The influence of observed and unobserved household characteristics and built environment measures are accommodated as $\psi_{i0} = \exp(\xi_{i0})$, $\psi_{ik} = \exp(\boldsymbol{\theta}'\mathbf{z}_{ik} + \xi_{ik})$, and $\gamma_{ik} = \exp(\boldsymbol{\delta}'\mathbf{w}_{ik})$; where, $\mathbf{z}_{ik}$ and $\mathbf{w}_{ik}$ are vectors of observed demographic and activity-travel environment measures influencing the choice of, and mileage allocation to, vehicle type/vintage

k, $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ are corresponding parameter vectors, and ξ_{ik} ($k=0,1,2,\ldots,K$) is the random error term in the sub-utility of choice alternative k. Assuming that the random error terms follow the independent and identically distributed (iid) standard Gumbel distribution leads to the standard MDCEV model *(9)*. On the other hand, allowing heteroscedasticity in the random terms across choice alternatives leads to the MDCHEV model (*25*).

It was observed in the data that, although many households owned vehicles from multiple vehicle type/vintage categories, a vast majority did not own multiple vehicles within any single vehicle type/vintage category. Therefore, along with the MDCEV (or MDCHEV) structure for modeling vehicle type/vintage choice (to recognize multiple discreteness), a simple multinomial logit (MNL) structure was used for vehicle make/model choice within each vehicle type/vintage category *(10)*. Specifically, the baseline utility (ψ_{ik}) specification of each vehicle type/vintage combination includes a log-sum variable from the corresponding MNL model of vehicle make/model choice. The log-sum variables carry information on vehicle-specific attributes specified in the MNL models to the MDC model utility functions (*11*).

3 EMPIRICAL ANALYSIS

3.1 Data

The primary data used for this analysis comes from the Florida add-on of 2009 US National Household Travel Survey (NTHS), which included detailed information on household vehicle fleet composition and usage for over 15,000 households. Secondary data sources used to collect vehicle-specific attributes include CarqueryApi.com (*23*) and Motortrend.com (*24*). All vehicles in the data were categorized into nine vehicle types and three vintage (*i.e.,* vehicle age) categories to form a total of 27 vehicle type and vintage alternatives. The vehicle type categories are: (1) Compact (2) Subcompact (3) Large Sedan (4) Mid-size Sedan (5) Two-seater (6) Van (7) SUV (8) Pickup Truck and (9) Motorcycle. The three vintage categories are: (1) 0 to 5 years (2) 6 to 11 years and (3) 12 years or older. After data cleaning and quality checks, the final sample comprises 10,294 household-records of households owning at least one vehicle. 8,500 of these households were randomly selected for model estimation and the remaining 1,794 households were kept aside for validation.

Table 1 shows the descriptive statistics of household vehicle type/vintage holdings and utilization. The second and third columns present the number of households owning a vehicle in each vehicle type/vintage category and the average annual household mileage for each vehicle type/vintage, respectively. It can be observed that households in Florida show a higher ownership of SUVs and mid-sized sedans in the 0-5 year and 6-11 year old categories than other vehicle type/vintage categories. The average annual mileage figures show a higher utilization rate for vans, pickup trucks and SUVs in the 0-5 year vintage category.

The last column shows the number of vehicle make/model alternatives owned by different households in the sample in each vehicle type/vintage category. As mentioned earlier, MNL structure was used to model the choice of vehicle make/model within each vehicle type/vintage category. The table does not show any vehicle make/model categories for motorcycles; because we did not model motorcycle choice in such a detail.

The demographic characteristics of the households in the estimation sample were found to be reasonably representative of the demographic makeup in Florida. However, descriptive statistics of the sample's demographic characteristics are not presented here to conserve space (but available from the authors).

3.2 Empirical Models for Estimating Annual Household Vehicle Mileage Budgets

Recall from Section 1 that we employed four different approaches for estimating annual household vehicle mileage budgets:

(a) Use of a stochastic frontier regression model for latent AH-VMF,
(b) Use of a log-linear regression model for observed total AH-VMT,
(c) Introduction of a non-motorized alternative in the MDCEV model, and
(d) Assumption of a uniform mileage budget for all households in the data.

The parameter estimates of the stochastic frontier model for AH-VMFs are not presented here to conserve space, but select empirical findings are discussed. Households with male householder and households with a younger householder were found to have a higher VMF than their counter parts (i.e., households with female householder and households with an older householder). As expected, AH-VMFs increased with household income level. Number of licensed drivers in the household, number of employed adults, and presence of children in the household are positively associated with AH-VMF, presumably because an additional member of each of these types is likely to increase household travel needs. Households located in urban areas tend to have lower VMFs compared to households located in rural areas. Similarly, households located in higher employment density and higher residential density neighborhoods have lower VMFs, possibly due to greater accessibility to employment and other activity opportunities within a closer proximity in higher density neighborhoods. An increase in fuel cost ($/gallon), as expected, tends to decrease households' VMFs.

The log-linear regression approach provided similar substantive interpretations (of the impacts of household sociodemographics and land use characteristics on AH-VMT) to those from the stochastic frontier model of AH-VMF discussed above. Therefore these results are not discussed exclusively here.

In the third approach, where we introduce a non-motorized alternative in the MDCEV model, we set the annual household mileage budget as the sum of annual non-motorized miles traveled (NMT) and total observed annual household vehicle miles traveled (AH-VMT). The annual NMT was calculated for each household assuming a walking distance of 0.5 miles per day for all household members (> 4 years old) for 100 days a year. For the fourth approach, we assumed a uniform annual household mileage budget of 119505 miles for every household, which is equal to the maximum observed annual household mileage travel (AH-VMT) in the dataset (119,405 miles) plus 100 miles.

3.3 Empirical Models for Vehicle Type/Vintage Holdings and Utilization

We estimated four different MDCEV models of vehicle type/vintage holdings and usage, one for each of the above discussed approaches for estimating annual household vehicle mileage budgets. In addition, we estimated an MDCHEV model, specifically for the annual household vehicle mileage budget obtained from the stochastic frontier approach.

The parameter estimates from all the different MDC models estimated in this study were found to be intuitive and consistent (in interpretation) with previous studies. The substantive interpretations of the influence of different explanatory variables are found to be similar across all different MDC models. For brevity, the model parameter estimates are not reported in the form of tables but only the important empirical findings are discussed here. Among socio-demographic characteristics, higher income households have lower baseline preference for older vehicle types and a higher baseline preference for new SUVs. As expected, households with more children are more likely to own and use vans. For householder characteristics, the results suggest that households with male householders are more likely to own and use pickup trucks, motorcycles, and old vans. Older households have higher preference for mid-age large sedans and vans. Among ethnicity variables, blacks are less likely to prefer trucks compared to other ethnic groups. Hispanics are more likely to prefer large sedans whereas Asians are less likely to prefer pickup trucks but more likely to prefer old compact vehicles.

Households located in rural areas have a higher preference for pickup trucks compared to households located in urban areas. Households located in low residential density neighborhoods prefer vans, SUVs and pickup trucks compared to households in high density neighborhoods. Also, households located in high employment density neighborhoods have lower preference for pickup trucks.

In each of the MDC models estimated, the baseline utility specification of each vehicle type/vintage combination includes a log-sum variable from the corresponding MNL model of vehicle make/model choice. The log-sum variables carry information of vehicle-specific attributes – purchase prices, operating costs (using gasoline price and fuel economy of the specific vehicle make/model for the given vehicle type and vintage), vehicle dimensions such as payload capacity, engine performance, and fuel type (premium vs. regular) – from the MNL model into the MDCEV model utility functions. The MNL model results suggest that, for any vehicle type/vintage, households prefer to own vehicle makes/models that are less expensive to purchase and operate, albeit the sensitivity to purchase prices and operating costs decreases with household income level. A greater preference was found for vehicle makes/models with superior engine performance (ratio of horsepower to weight), for all-wheel-drive vehicles, and for regular fuel vehicles. For pickup trucks, a higher preference was found for makes/models with high payload capacity.

3.4 Comparison of Predictive Accuracy Assessments Using Validation Data

This section presents a comparison of predictive accuracy assessments for the different MDCEV models estimated using different approaches for estimating annual household vehicle mileage budgets. As mentioned earlier, we had kept aside a random sample of 1,794 households for validation. All MDC model predictions were undertaken using the forecasting algorithm proposed by Pinjari and Bhat (*19*), using 100 sets of random draws to cover the error distributions for each of these households.

The predicted ownership (i.e., discrete choice) for each vehicle type/vintage category was computed as the proportion of instances the category was predicted with a positive mileage across all 100 sets of random draws for all households. These aggregate predictions from different MDC models (with annual household mileages estimated from different approaches) were compared with the percentages of households owning each vehicle type/vintage category. While not shown in figures or tables to conserve space, all the approaches resulted in similar results except when the budget was assumed to be 119,505 miles for all households. The last approach resulted in relatively poor predictions.

The predicted aggregate mileage for a vehicle type/vintage category was computed as average of the mileage predicted across all random draws for all households with a positive mileage prediction. To compare the different approaches used to estimate mileage budgets, we plotted distributions of the observed mileage and the predicted mileage for each vehicle type/vintage using different approaches for the mileage budgets. To conserve space, we present these distributions for only a few vehicle types in the new vintage (0-5yrs age) category. The distributions are presented in the form of box-plots in Figure 1, with nine sub-figures (one sub-figure for each vehicle type). In all these sub-figures, there are two different results for the stochastic frontier approach, one for the MDCEV model and the other for the MDCHEV model. For the MDCHEV model, baseline utility function for the outside good (t_{io}) was specified to have a different variance than the utility functions for all other goods; i.e., vehicle type/vintage categories (t_{ik}). The MDCHEV model was explored because the AH-VMFs estimated from the stochastic frontier models were much larger in magnitude when compared to the observed AH-VMTs (recall that by design AH-VMF > AH-VMT). With such large values of annual mileage, the MDCEV model might not be able to appropriately allocate the mileage budget between the outside good (t_{io}) and the different vehicle type/vintage categories (t_{ik}). The MDCHEV model helps in rectifying this issue *(25)*.

Figure 1 suggests that, when compared to the observed vehicle mileage distribution, predictions from all four MDCEV models and those from the MDCHEV model exhibit higher variance. Also, all model predictions exhibit a discernible likelihood of over prediction in mileage as evidenced by larger values of the 95th percentile values when compared to that of the observed 95th percentile value. Among the different MDCEV models, in terms of predicting annual mileage on household vehicles, the MDCEV model with uniform budget assumption (of 119,505 miles) exhibits poor performance, with a significant extent of over-prediction of annual mileage for all vehicle types. On the other hand, the MDCEV model using budgets (i.e., AH-VMT) from the log-linear regression approach performs relatively better than the MDCEV models with budgets from all other approaches. The MDCEV model with budgets (AH-VMF) from the stochastic frontier regression approach, when compared to the MDCEV model with budgets from the log-linear regression approach, exhibits a relatively higher over-prediction of annual mileage for all vehicle types. However, when the MDCHEV model (instead of the MDCEV model) was used with stochastic frontier budgets (AH-VMF), the predicted annual mileage distributions improve discernibly and become close to those of the MDCEV model used in conjunction with log-linear budgets. This is because the MDCHEV model allowed a higher variance of the error term on the outside good (t_{io}) in comparison with those of the vehicle type/vintage categories, which in turn helped in better allocation of AH-VMF between t_{io} and all vehicle type/vintage categories in the model.

In summary, the results indicate that the MDCEV model with budgets from the log-linear regression model resulted in better predictions than all other approaches used to estimate budgets. The MDCHEV model with mileage budgets from the stochastic frontier regression model provided predictions that were close to that of the MDCEV model with log-linear approach.

3.5 Simulations of the Effect of Fuel Economy Changes on Vehicle Type/Vintage Holdings and Usage

Here, we compare the policy predictions of the different MDCEV (and MDCHEV) models estimated in this study (with mileage budgets from the different approaches discussed earlier) by examining the effect of increasing fuel economy (miles/gallon) on vehicle holdings and mileage allocation patterns of the 1,794 households set aside for validation. Specifically, we increased the fuel economy for new (0-5 years) compact, subcompact, large and mid-size vehicles by 25%. This change is reflected in the operating cost variable in the MNL models of vehicle make/model choice for each vehicle type/vintage category. The log-sum variables constructed using the MNL model parameters were used to carry this change to the MDCEV models.

Note that since the fuel economy variable does not appear in the stochastic frontier or log-linear regression models, the estimated mileage budgets do not differ between the base-case (*i.e.*, before-policy) and the policy-case (*i.e.*, after policy) for these two approaches. The other approaches considered also assume the same mileage budgets between the base-case and the policy-case.

For the different approaches to estimate mileage budgets, we employed the corresponding MDCEV models to predict vehicle holdings and usage for the base-case and the policy-case. Subsequently, the policy effect was quantified as two different measures of differences between the policy-case and base-case, as shown in Table 2: (1) The "% Change in Holdings" column shows the percentage change in the holdings (or ownership) of the corresponding vehicle type/vintage, and (2) The "Change in Mileage" column indicates the average change in annual vehicle mileage for households in which a change occurred in the usage (or mileage) for the corresponding vehicle type/vintage category.

We now make several observations from the table, beginning with the similarities in results from all different approaches. First, across all different approaches, an increase in fuel economy of new (0-5yrs age) compact, subcompact, large and mid-sized vehicles leads to an increase in the holding (or ownership) of vehicles in those categories. The results also indicate a decrease in the holding of almost all other vehicle type/vintage categories. Overall, this is an intuitive result since an increase in fuel economy reduces operating cost and, *ceteris paribus*, households prefer vehicles that are less costly to operate (consistent with MNL model results). Second, in the context of vehicle usage (i.e., annual mileage), results from all different approaches suggest that fuel economy improvements led to increase in usage of all vehicle type/vintage categories for which the fuel economy was improved. Also, the results indicate a decrease in the average mileage for all other vehicle types/vintage categories. When such decreases in annual mileages are examined closely within each vehicle type, it can be observed that there is a higher decrease in the usage of older vehicle types that that of newer vehicle types. This is an intuitive result since older vehicles tend to have lower fuel economy compared to newer vehicles, which makes older vehicle types more expensive to operate.

Notwithstanding the above similarities, there are some important differences in policy predictions from all the different approaches examined in this study. Specifically, when examining where the additional mileage for new compact, subcompact, large and mid-size vehicles comes from, results from the log-linear regression approach differ fundamentally from all other approaches. In this approach, the annual mileage budget is simply reallocated among the different vehicle types/vintages. That is, increases in annual mileage of certain vehicle type/vintage categories must come from a decrease in the annual mileage of other vehicle types/vintage categories. This result is counter intuitive and in contrast to previous empirical evidence in the literature that improvements in fuel economy tend to induce additional travel (*26*). On the other hand, the stochastic frontier approach and the other approaches provide a "buffer" in the form of an unspent mileage alternative (t_{io}) from where the additional mileage can be drawn. As a result, for all approaches other than the MDCEV model that uses annual mileage budgets from the log-linear regression approach, the increased usage of new compact, subcompact, large and mid-sizes vehicles doesn't necessarily translate into an equal decrease in usage of other household vehicles. Instead, the overall household annual VMT across all vehicles increases, suggesting that improvements in fuel economy tend to induce additional travel (this can be observed from the last row of the table for all approaches except the log-linear regression approach). This finding is intuitive and consistent with other studies in the literature (*26*).

The natural next question is which approach provides a more reasonable estimate of the induced travel than other approaches? Assuming a uniform annual mileage budget of 119505 miles shows an average induced travel of 554 miles per annum per household. Given the poor prediction performance of this approach (discussed in the earlier section) the estimate of 554 miles per annum per household is perhaps less reliable than the estimates from other approaches. The approach of adding a non-motorized mileage alternative to the MDCEV model shows an unrealistically small induced travel of 10 miles per annum per household (in response to 25% improvement in fuel economy). The stochastic frontier approach, on the other hand, with both MDCEV and MDCHEV models, appears to result in more reasonable estimates of induced travel – 258 miles per annum per household from the MDCEV model and 230 miles per annum per household from the MDCHEV model. Of course, it is difficult to assertively assess the reliability of these estimates without comparing and contrasting the estimates with findings from the literature. Further work is necessary for a deeper examination of these estimates and a more extensive testing of the different approaches used to estimate annual household vehicle mileage budgets.

4 Conclusions

This paper presents an empirical comparison of the following approaches to estimate annual mileage budgets for multiple discrete-continuous extreme value (MDCEV) models of household vehicle ownership and utilization, using household survey data from Florida:

(a) The traditional log-linear regression approach to model observed total annual household vehicle miles traveled (AH-VMT),

(b) The stochastic frontier regression approach to model latent (or unobserved) annual vehicle mileage frontier (AH-VMF),

(c) Introduction of a non-motorized choice alternative in the MDCEV model, assuming that the total household mileage is equal to the total annual mileage (AH-VMT) plus the total non-motorized mileage (NMT), and

(d) Assumption of an arbitrarily determined, uniform mileage budget for all households in the data.

For the stochastic regression approach, both MDCEV and MDCHEV models were estimated and examined.

In terms of prediction performance in a validation sample, assuming an arbitrarily determined uniform annual vehicle mileage budget for all households resulted in the most distorted predictions vis-à-vis observed distributions in the validation sample. Therefore, we recommend not using this approach to approximate annual household vehicle mileage budgets for MDCEV models of vehicle ownership and usage.

On the other hand, the MDCEV model using budgets (i.e., AH-VMT) from the log-linear regression approach performed better than all other approaches. The MDCEV model with budgets (AH-VMF) from the stochastic frontier regression approach, when compared to the MDCEV model with budgets from the log-linear regression approach, exhibits a relatively higher over-prediction of annual mileage for all vehicle types. However, when the MDCHEV model (instead of the MDCEV model) was used with stochastic frontier budgets (AH-VMF), the predicted annual mileage distributions improve discernibly and become close to those of the MDCEV model used in conjunction with log-linear budgets.

Policy predictions of the different MDCEV (and MDCHEV) models estimated in this study were compared by examining the effect of increasing fuel economy (miles/gallon) on vehicle ownership and usage. The policy predictions demonstrate an important drawback of the log-linear approach for estimating annual mileage budgets for MDCEV models of household vehicle ownership and utilization. Specifically, this approach does not allow for the total AH-VMT to increase or decrease due to changes in vehicle-specific attributes such as changes in fuel economy of specific vehicle type/vintage categories. In this approach, the total AH-VMT is simply reallocated among the different vehicle type/vintage categories. MDCEV models with budget estimates form the other three approaches – stochastic frontier regression, introduction of a non-motorized choice alternative, and the assumption of a uniform annual mileage budget – overcome this problem. This is because all these approaches provide a "buffer" for the AH-VMT to increase or decrease as needed. As a result, consistent with other studies in the literature, improvements in fuel economy induce an increase in total AH-VMT, as opposed to mere reallocation of the current AH-VMT across different household vehicles. Among the three approaches examined in this study that allow for the AH-VMT to increase or decrease, the stochastic frontier approach provides the most reasonable results in terms of the magnitude of induced travel.

Taking into consideration all the above results, in view of policy responsiveness and prediction accuracy considerations, we recommend using the stochastic frontier approach for estimating annual household vehicle mileage budgets for multiple discrete-continuous models of household vehicle ownership and utilization. Furthermore, with the stochastic frontier approach to estimating annual household vehicle mileage budgets, we recommend using the MDCHEV model over the MDCEV model for better prediction accuracy.

The empirical work in this paper can be extended by a more rigorous assessment of the predicted influences of fuel economy improvements vis-à-vis the existing literature on induced travel and rebound effects (*26*). Methodologically, the mileage budgets from the stochastic frontier regression approach and that of the log-linear regression approaches were derived by taking an expected value of the corresponding regression equations. Instead, the entire distributions of the budget equations can be utilized to estimate the MDCEV models, by integrating the budget equation and the MDCEV specification into a joint modeling framework.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. DUE 0965743. Comments from anonymous reviewers helped improve the discussion of results in this paper.

References

1. U.S. Department of Transportation, Federal Highway Administration, *2009 National Household Travel Survey.* URL: http://nhts.ornl.gov.

2. Anowar S., N. Eluru, and L. Miranda-Moreno. Alternative Modeling Approaches Used for Examining Automobile Ownership: A Comprehensive Review. *Transport Reviews*, 34 (4), 2014, pp. 441-473.

3. Lave, C.A. and K. Train. A disaggregate model of auto-type choice. *Transportation Research Part A,* 13 (1), 1979, pp. 1–9.

4. Berkovec, J. and J. Rust. A nested logit model of automobile holdings for one vehicle households. *Transportation Research Part B* 19 (4), 1985, pp. 275–285.

5. Mohmmadian, A. and E.J. Miller. An Empirical Investigation of Household Vehicle Type Decisions. *Transportation Research Record,* 1854, 1985, pp. 99-106.

6. Mannering, F. and C. Winston. A dynamic empirical analysis of household vehicle ownership and utilization. *Rand Journal of Economics* 16 (2), 1985, pp. 215–236.

7. de Jong, G. C.. An indirect utility model of car ownership and car use. *European Economic Review*, 34(5), 1990, pp. 971–985.

8. Spissu, E., A.R. Pinjari, R.M. Pendyala, and C.R. Bhat. A Copula-Based Joint Multinomial Discrete-Continuous Model of Vehicle Type Choice and Miles of Travel. *Transportation*, 36(4), 2009, pp. 403-422.

9. 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,* 42(3), 2008, pp. 274-303.

10. 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*, 40(1), 2006, pp. 35-53.

11. 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,* 43(1), 2009, pp. 1-18.

12. 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,* 42(9), 2008, pp. 736-758.

13. Rouwendal, J. and de Borger, B. Multiple car ownership, fuel efficiency and substitution between cars. *International Choice Modelling Conference*, Harrogate, UK, July 2011.

14. Liu,Y., J. Tremblay, C. Cirillo. An Integrated model for Discrete and Continuous Decisions with Application to Vehicle Ownership, Type, and Usage Choices. *Presented at the 93rd Transportation Research Board Annual Meeting, Washington D.C., January 2014.*

15. Paleti, R., N. Eluru, C.R. Bhat, R.M. Pendyala, T.J. Adler, and K.G. Goulias. The Design of a Comprehensive Microsimulator of Household Vehicle Fleet Composition, Utilization, and Evolution. *Transportation Research Record,* 2254, 2011, pp. 44-57.

16. Ahn, J., G. Jeong, and Y. Kim. A forecast of household ownership and use of alternative fuel vehicles: a multiple discrete-continuous choice approach. *Energy Economics*, 30(5), 2008, pp. 2091-2104.

17. Jaggi, B., C. Weis, and K.W. Axhausen. Stated Response and Multiple Discrete-Continuous Choice Models: Analysis of Residuals. *Journal of Choice Modelling*, 6, 2011, pp. 44-59.

18. You, D., V. M. Garikapati, R.M. Pendyala, C.R. Bhat, S. Dubey, K. Jeon and V. Livshits. Development of a Vehicle Fleet Composition Model System for Implementation in an Activity-Based travel Model. *Forthcoming, Transportation Research Record,* 2014.

19. Pinjari, A.R., and C.R. Bhat. Computationally Efficient Forecasting Procedures for Kuhn-Tucker Consumer Demand Model Systems: Application to Residential Energy Consumption Analysis. Technical paper, University of South Florida, 2011.

20. Augustin, *B.*, A.R. Pinjari, A. Faghih Imani, V. Sivaraman, N. Eluru, and R.M. Pendyala. Stochastic Frontier Estimation of Budgets for Kuhn-Tucker Demand Systems: Application to Activity Time-Use Analysis. Technical paper, University of South Florida, 2014.

21. Banerjee, A., X. Ye, and R.M. Pendyala. Understanding Travel Time Expenditures Around the World: Exploring the Notion of a Travel Time Frontier. *Transportation*, 34(1), 2007, pp. 51-65.

22. Aigner, D., C.A.K. Lovell, and P. Schmidt. Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics*, 6(1), 1977, pp. 21-37.

23. CarqueryAPI. The Vehicle Data API and Database, Full Model/Trim data. 2014. Website: http://www.carqueryapi.com

24. MotorTrend. Motor Trend, 2014. Website: http://www.motortrend.com

25. Sikder, S., and A.R. Pinjari. The Benefits of Allowing Heteroscedastic Stochastic Distributions in Multiple Discrete-Continuous Choice Models. Forthcoming, *Journal of Choice Modelling,* Vol 9, 2014, pp. 39-56.

26. Small, K., and K. Van Dender. If cars were more efficient, would we use less fuel? *Access*, Vol 31, 2007, pp. 8-13.

LIST OF FIGURES

LIST OF TABLES

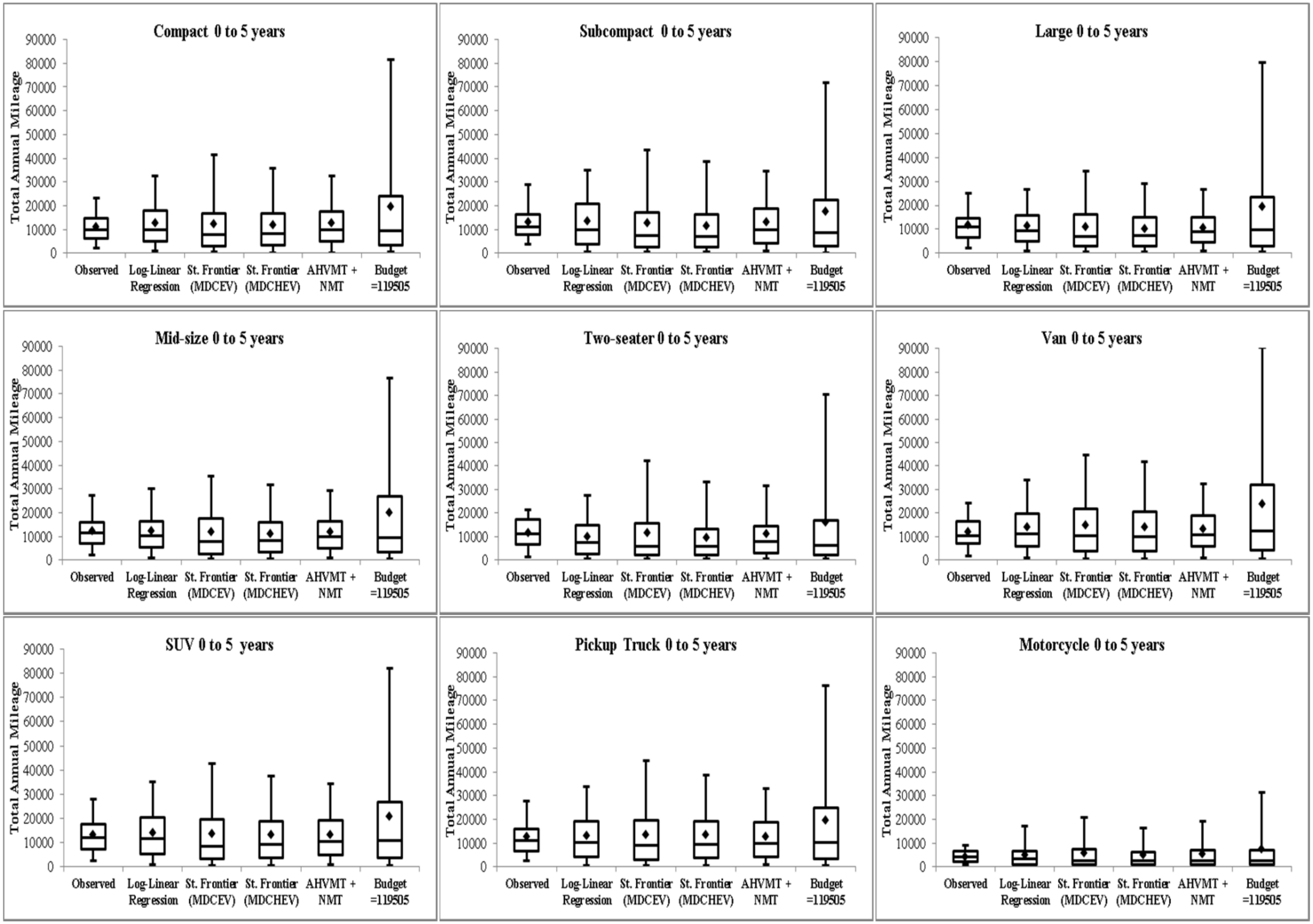

FIGURE 1 Observed and Predicted Distributions of Total Annual Mileage by Vehicle Type/Vintage

TABLE 1 Descriptive Statistics of Vehicle Type/Vintage Holdings and Usage in the Estimation Sample

Vehicle Type/Vintage	Total number (%) of households owning	Average Annual Mileage	Number of vehicle make/model alternatives for MNL model
Compact 0 to 5 years	887 (10.4%)	11363	36
Compact 6 to 11 years	802 (9.4%)	10471	45
Compact 12 years or older	391 (4.6%)	8254	29
Subcompact 0 to 5 years	301 (3.5%)	11104	23
Subcompact 6 to 11 years	246 (2.9%)	9998	21
Subcompact 12 years or older	251 (3.0%)	8276	27
Large 0 to 5 years	624 (7.3%)	10754	25
Large 6 to 11 years	566 (6.7%)	9573	19
Large 12 years or older	336 (4.0%)	8282	20
Mid-size 0 to 5 years	1299 (15.3%)	11079	32
Mid-size 6 to 11 years	1223 (14.4%)	10183	35
Mid-size 12 years or older	417 (4.9%)	7921	35
Two-seater 0 to 5 years	101 (1.2%)	8625	21
Two-seater 6 to 11 years	97 (1.1%)	8345	14
Two-seater 12 years or older	93 (1.1%)	8193	13
Van 0 to 5 years	522 (6.1%)	13184	20
Van 6 to 11 years	522 (6.1%)	11222	22
Van 12 years or older	195 (2.3%)	8898	20
SUV 0 to 5 years	1512 (17.8%)	12851	52
SUV 6 to 11 years	1067 (12.6%)	11920	41
SUV 12 years or older	279 (3.3%)	9428	24
Pickup Truck 0 to 5 years	852 (10.0%)	13046	17
Pickup Truck 6 to 11 years	818 (9.6%)	11598	16
Pickup Truck 12 years or older	540 (6.4%)	8948	14
Motorcycle 0 to 5 years	153 (1.8%)	4305	NA
Motorcycle 6 to 11 years	126 (1.5%)	3461	NA
Motorcycle 12 years or older	99 (1.2%)	2194	NA
Total Observed Annual Mileage	NA	18010	NA

TABLE 2 Impact of Increasing Fuel Economy for New (0-5 years) Compact, Subcompact, Large, and Mid-sized Vehicles

	Log-linear Regression		**Stochastic Frontier (MDCEV)**		**Stochastic Frontier (MDCHEV)**		**Budget = AH-VMT + NMT**		**Budget = 119505 miles**	
Vehicle Type and Vintage	**% Change in Holdings**	**Change in Mileage***	**% Change in Holdings**	**Change in Mileage***	**% Change in Holdings**	**Change in Mileage***	**% Change in Holdings**	**Change in Mileage***	**% Change in Holdings**	**Change in Mileage***
Unspent Mileage (t_o)	-	-	-	-258	-	-230	-	-10	-	-554
Compact 0 to 5 years	**1.03%**	**404**	**1.28%**	**431**	**1.17%**	**428**	**1.04%**	**267**	**1.16%**	**669**
Compact 6 to 11 years	-0.36%	-292	-0.12%	-153	-0.21%	-189	-0.26%	-308	-0.07%	-100
Compact 12 years or older	-0.70%	-345	-0.33%	-179	-0.39%	-220	-0.49%	-339	-0.06%	-113
Subcompact 0 to 5 years	**0.09%**	**193**	**0.95%**	**243**	**0.93%**	**332**	**0.63%**	**202**	**0.59%**	**314**
Subcompact 6 to 11 years	-0.43%	-345	-0.25%	-174	-0.59%	-236	-0.47%	-401	-0.21%	-114
Subcompact 12 years or older	-0.44%	-340	-0.30%	-164	-0.42%	-199	-0.55%	-312	-0.18%	-108
Large 0 to 5 years	**0.81%**	**352**	**1.02%**	**322**	**1.43%**	**351**	**0.96%**	**225**	**1.20%**	**538**
Large 6 to 11 years	-0.48%	-404	-0.26%	-164	-0.29%	-214	-0.40%	-344	-0.14%	-95
Large 12 years or older	-0.71%	-550	-0.40%	-231	-0.68%	-300	-0.50%	-475	-0.29%	-145
Mid-size 0 to 5 years	**0.93%**	**348**	**1.12%**	**325**	**1.04%**	**321**	**0.76%**	**209**	**0.95%**	**546**
Mid-size 6 to 11 years	-0.35%	-270	-0.17%	-144	-0.18%	-160	-0.30%	-274	-0.06%	-86
Mid-size 12 years or older	-0.43%	-404	-0.31%	-175	-0.45%	-205	-0.37%	-365	-0.20%	-109
Two-seater 0 to 5 years	0.00%	-161	-0.18%	-126	-0.50%	-138	-0.23%	-185	-0.39%	-78
Two-seater 6 to 11 years	-0.25%	-267	-0.22%	-164	-0.38%	-223	-0.78%	-257	0.00%	-92
Two-seater 12 years or older	-0.61%	-216	-0.58%	-121	-0.65%	-195	-0.46%	-225	0.00%	-83
Van 0 to 5 years	-0.53%	-370	-0.17%	-149	-0.37%	-185	-0.37%	-361	-0.09%	-97
Van 6 to 11 years	-0.61%	-367	-0.12%	-151	-0.38%	-187	-0.47%	-322	-0.21%	-102
Van 12 years or older	-0.61%	-445	-0.35%	-202	-0.76%	-271	-0.69%	-469	-0.05%	-116
SUV 0 to 5 years	-0.20%	-214	-0.10%	-107	-0.11%	-122	-0.24%	-191	-0.04%	-68
SUV 6 to 11 years	-0.26%	-257	-0.16%	-138	-0.23%	-158	-0.28%	-252	-0.09%	-93
SUV 12 years or older	-0.74%	-326	-0.22%	-171	-0.49%	-196	-0.65%	-349	-0.14%	-91
Pickup Truck 0 to 5 years	-0.35%	-278	-0.19%	-159	-0.38%	-187	-0.32%	-291	-0.10%	-102
Pickup Truck 6 to 11 years	-0.33%	-310	-0.22%	-170	-0.09%	-214	-0.37%	-314	-0.04%	-107
Pickup Truck 12 years or older	-0.58%	-319	-0.29%	-205	-0.49%	-232	-0.64%	-318	-0.23%	-123
Motorcycle 0 to 5 years	-0.74%	-170	-0.51%	-75	-0.81%	-119	-0.48%	-144	-0.18%	-51
Motorcycle 6 to 11 years	-0.63%	-134	-0.08%	-82	-1.05%	-107	-0.83%	-132	0.00%	-63
Motorcycle 12 years or older	-0.29%	-89	-0.65%	-55	-0.40%	-79	-0.55%	-95	-0.47%	-34
Change in AH-VMT	**0 miles**		**258 miles**		**230 miles**		**10 miles**		**554 miles**	

*When a change in annual mileage occurred for this vehicle type/vintage category