January 2014

Comparison of Different Approaches to Estimating Budgets for Kuhn-Tucker Demand Systems: Applications for Individuals' Time-Use Analysis and Households' Vehicle Ownership and Utilization Analysis

Bertho Augustin
University of South Florida, bertho@mail.usf.edu

www.manaraa.com

Comparison of Different Approaches to Estimating Budgets for Kuhn-

Tucker Demand Systems: Applications for Individuals' Time-Use Analysis

and Households' Vehicle Ownership and Utilization Analysis

by

Bertho Augustin

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Engineering Science
Department of Civil and Environmental Engineering
College of Engineering
University of South Florida

Major Professor: Abdul Rawoof Pinjari, Ph.D.
Qing Lu, Ph.D.
Yu Zhang, Ph.D.

Date of Approval:
July 3, 2014

Keywords: Stochastic frontier model, MDCEV model, activity
participation, multinomial logit model, vehicle type/vintage

DEDICATION

This thesis is dedicated to my father, Berjean Augustin and my mother, Marie Claire Augustin for their encouragement, love, affection and support throughout my life. I would also like to dedicate this thesis to my brothers and sisters: Berline Augustin, Bertrand Augustin, Jean Bernard Augustin and Bergeline Augustin for their friendship, love and support. Finally, I would also like to dedicate this thesis for my fiancée, Christie Mauretour for her love, her understanding and for being there for me when I needed it the most.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

ii

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis compares different approaches to estimating budgets for Kuhn-Tucker (KT) demand systems, more specifically for the multiple discrete-continuous extreme value (MDCEV) model. The approaches tested include: (1) The log-linear regression approach (2) The stochastic frontier regression approach, and (3) arbitrarily assumed budgets that are not necessarily modeled as a function of decision maker characteristics and choice-environment characteristics.

The log-linear regression approach has been used in the literature to model the observed total expenditure as way of estimating budgets for the MDCEV models. This approach allows the total expenditure to depend on the characteristics of the choice-maker and the choice environment. However, this approach does not offer an easy way to allow the total expenditure to change due to changes in choice alternative-specific attributes, but only allows a reallocation of the observed total expenditure among the different choice alternatives. To address this issue, we propose the stochastic frontier regression approach. The approach is useful when the underlying budgets driving a choice situation are unobserved, but only the expenditures on the choice alternatives of interest are observed. The approach is based on the notion that consumers operate under latent budgets that can be conceived (and modeled using stochastic frontier regression) as the maximum possible expenditure they are willing to incur.

To compare the efficacy of the above-mentioned approaches, we performed two empirical assessments: (1) The analysis of out-of-home activity participation and time-use (with a budget on the total time available for out-of-home activities) for a sample of non-working

adults in Florida, and (2) The analysis of household vehicle type/vintage holdings and usage (with a budget on the total annual mileage) for a sample of households in Florida. A comparison of the MDCEV model predictions (based on budgets from the above mentioned approaches) demonstrates that the log-linear regression approach and the stochastic frontier approach performed better than arbitrarily assumed budgets approaches. This is because both approaches consider heterogeneity in budgets due to socio-demographics and other explanatory factors rather than arbitrarily imposing uniform budgets on all consumers. Between the log-linear regression and the stochastic frontier regression approaches, the log-linear regression approach resulted in better predictions (vis-à-vis the observed distributions of the discrete-continuous choices) from the MDCEV model. However, policy simulations suggest that the stochastic frontier approach allows the total expenditures to either increase or decrease as a result of changes in alternative-specific attributes. While the log-linear regression approach allows the total expenditures to change as a result of changes in relevant socio-demographic and choice-environment characteristics, it does not allow the total expenditures to change as a result of changes in alternative-specific attributes.

1.1 Background

Numerous consumer choices are characterized by "multiple discreteness" where consumers can potentially choose multiple alternatives from a set of discrete alternatives available to them. Along with such discrete-choice decisions of which alternative(s) to choose, consumers typically make continuous-quantity decisions on how much of each chosen alternative to consume. Such multiple discrete-continuous (MDC) choices are being increasingly recognized and analyzed in a variety of social sciences, including transportation, economics, and marketing.

A variety of approaches have been used to model MDC choices. Among these, an increasingly popular approach is based on the classical microeconomic consumer theory of utility maximization. Specifically, consumers are assumed to optimize a direct utility function $U(\mathbf{t})$ over a set of non-negative consumption quantities $\mathbf{t} = (t_1, ..., t_k, ..., t_K)$ subject to a budget constraint, as below:

$$Max\ U(\mathbf{t}) \text{ such that } \sum_{k=1}^{K} p_k t_k = y \text{ and } t_k \geq 0\ \forall\ k = 1, 2, ..., K \tag{1}$$

In the above Equation, $U(\mathbf{t})$ is a quasi-concave, increasing, and continuously differentiable utility function of the consumption quantities, $p_k\ (k = 1, 2, ..., K)$ are unit prices for all goods, and *y* is a budget for total expenditure. A particularly attractive approach for deriving the demand functions

---

[1] Part of this thesis has been submitted for publication and conference proceeding please refer to Augustin et al (2014) and Pinjari et al (2014).

from the utility maximization problem in Equation (1), due to Hanemann (1978) and Wales and Woodland (1983), is based on the application of Karush-Kuhn-Tucker (KT) conditions of optimality with respect to the consumption quantities. When the utility function is assumed to be randomly distributed over the population, the KT conditions become randomly distributed and form the basis for deriving the probability expressions for consumption patterns. Due to the central role played by the KT conditions, this approach is called the KT demand systems approach (or KT approach, in short).

Over the past decade, the KT approach has received significant attention for the analysis of MDC choices in a variety of fields, including environmental economics (von Haefen and Phaneuf, 2005), marketing (Kim et al., 2002), and transportation. In the transportation field, the multiple discrete-continuous extreme value (MDCEV) model formulated by Bhat (2005, 2008) has led to an increased use of the KT approach for analyzing a variety of choices, including individuals' activity participation and time-use (Habib and Miller, 2008; Chikaraishi et al., 2010), household vehicle ownership and usage (Ahn et al., 2008; Jaggi et al., 2011), recreational/leisure travel choices (von Haefen and Phaneuf, 2005; Van Nostrand et al., 2013), energy consumption choices, and builders' land-development choices (Farooq et al., 2013; Kaza et al., 2010). Thanks to these advances, KT-based MDC models are being increasingly used in empirical research and have begun to be employed in operational travel forecasting models (Bhat et al., 2013[a]). On the methodological front, recent literature in this area has started to enhance the basic formulation in Equation (1) along three specific directions: (a) toward more flexible, non-additively separable utility functions that accommodate rich substitution and complementarity patterns in consumption (Bhat et al., 2013[b]), (b) toward more flexible stochastic specifications

2

for the random utility functions (Pinjari, 2011), and (c) toward greater flexibility in the specification of the constraints faced by the consumer (Castro et al., 2012).

1.2 Gaps in Research

Despite the methodological advances and many empirical applications, one particular issue related to the budget constraint has yet to be resolved. Specifically, almost all KT model formulations in the literature, including the MDCEV model, assume that the available budget for total expenditure, i.e. y in Equation (1), is fixed for each decision-maker (or for each choice occasion, if repeated choice data is available). Given the fixed budget, any changes in the choice alternative attributes, or the choice environment can only lead to a reallocation of the budget among different choice alternatives. The formulation itself does not allow either an increase or a decrease in the total available budget. Consider, for example, the context of households' vehicle holdings and utilization. In most applications of the KT approach for this context (Bhat et al., 2009, Ahn et al., 2008), a total annual mileage budget is assumed to be available for each household. This mileage budget is obtained exogenously for use in the KT model, which simply allocates the given total mileage among different vehicle types. Therefore, any changes in vehicle attributes (e.g., prices and fuel economy) and gasoline prices can only lead to a reallocation of the given mileage budget among the different vehicle types without allowance for either an increase or a decrease in the total mileage. Similarly, in the context of individuals' out-of-home activity participation and time-use, most applications of the KT approach consider an exogenously available total time budget that is allocated among different activity type alternatives. The KT model itself does not allow either an increase or decrease in the total time expended in the activities of interest due any changes in the alternative-specific characteristics.

3

It is worth noting that the fixed budget assumption is not a theoretical/conceptual flaw of the consumer's utility maximization formulation per se. Classical microeconomics typically considered the consumption of broad consumption categories (such as food, housing, and clothing). In such situations, all consumption categories potentially can be considered in the model while considering natural constraints such as total income for the budget. Similarly, several time-use analysis applications can use natural constraints individuals face as their time budgets (e.g., 24 hours in a day). However, many choice situations of interest involve the analysis of a specific broad category of consumption, with elemental consumption alternatives within that broad category, as opposed to all possible consumption categories that can possibly exhaust naturally available time and/or money budgets. For example, in a marketing context involving consumer purchases of a food product (say, yogurt), one can observe the different brands chosen by a consumer along with the consumption amount of each brand, but cannot observe the maximum amount of expenditure the consumer is willing to allocate to the product. It is unreasonable to assume that the consumer would consider his/her entire income as the budget for the choice occasion.

The above issue has been addressed in two different ways in the literature, as discussed briefly here (see Chintagunta and Nair, 2010; and von Haefen, 2010). The first option is to consider a two-stage budgeting process by invoking the assumptions of separability of preferences across a limited number of broad consumption categories and homothetic preferences within each broad category. The first stage involves allocation between the broad consumption categories while the second stage involves allocation among the elemental alternatives within the broad category of interest. The elemental alternatives in the broad consumption category of interest are called *inside goods*. The second option is to consider a

4

Hicksian composite commodity (or multiple Hicksian commodities, one for each broad consumption category) that bundles all consumption alternatives that are not of interest to the analyst into a single *outside good* (or multiple outside goods, one for each broad consumption category). The assumption made here is Hicksian separability, where the prices of all elementary alternatives within the outside good vary proportionally and do not influence the choice and expenditure allocation among the inside goods (see Deaton and Muellbauer, 1980). The analyst then models the expenditure allocation among all inside goods along with the outside good.

Many empirical studies use variants of the above two approaches either informally or formally with well-articulated assumptions. For instance, one can informally mimic the two-stage budgeting process by modeling the total expenditure on a specific set of choice alternatives of interest to the analyst in the first stage. The natural instinct may be to use linear (or log-linear) regression to model the total expenditure in the first stage. Subsequently, the second stage allocates the total expenditure among the different choice alternatives of interest. This approach is straightforward and also allows the total expenditure (in the first-stage regression) to depend on the characteristics of the choice-maker and the choice environment. The problem, however, is that the first-stage regression cannot incorporate the characteristics of choice alternatives in a straight forward fashion. Therefore, changes in the attributes of choice alternatives, such as price change of a single alternative, will only lead to reallocation of the total expenditure among choice alternatives without allowing for the possibility that the overall expenditure itself could increase or decrease. This is considered as a drawback in using the MDCEV approach for modeling vehicle holdings and usage (Fang, 2008) and for many other applications. Besides, from an intuitive standpoint, the observed expenditures may not necessarily represent the budget

5

for consumption. It is more likely that a greater amount of underlying budget governs the expenditure patterns, which the consumers may or may not expend completely.

1.3 Objectives

The purpose of this thesis is to compare different approaches to estimate budgets for the multiple discrete-continuous extreme value (MDCEV) models. One of the approaches is log-linear regression. Specifically, log-linear regression is used to model the total expenditure on all choice alternatives available to the decision-maker. The total expenditure estimated using the log-linear regression approach is subsequently used as budget for the MDCEV model. We use log-linear regression as opposed to a standard linear regression in order to avoid situations of predicting negative budgets. This approach is straightforward and also allows the total expenditure (in the first-stage regression) to depend on the characteristics of the choice-maker and the choice environment. However, this approach does not offer an easy way to allow the budgets to vary with alternative-specific characteristics.

To address the above issue, we propose the use of stochastic frontier regression approach to estimate budgets for the MDCEV models. Stochastic frontier regression models have been widely used in firm production economics (Aigner et al, 1977; Kumbhakar and Lovell, 2000) for identifying the maximum possible production capacity (i.e., production frontier) as a function of various inputs. While the actual production levels and the inputs to the production can be observed, a latent production frontier is assumed to exist. Such a production frontier is the maximum possible production that can be achieved given the inputs. Conversely, one can conceive of a cost frontier that is the minimum possible cost at which a good can be produced.

In travel behavior research, the stochastic frontier approach has been used to analyze: (1) the time-space prism constraints that people face (Kitamura et al., 2000), and (2) the

6

maximum amount of time that people are willing to allocate to travel in a day (Banerjee et al., 2007). In the former case, while the departure times and arrival times at fixed activities (such as work) are observed in the survey data, the latest possible arrival time or the earliest possible departure time are unobserved and therefore modeled as stochastic frontiers. In the latter case, while the daily total travel time can be measured, an unobserved Travel Time Frontier (TTF) is assumed to exist that represents the maximum possible travel time an individual is willing to undertake in a day.

Analogous to the above examples, in many consumer choice situations, especially in time-use situations, one can conceive of latent time and/or money frontiers that govern choice making. In the case of household ownership and usage, one can also perceive a maximum total annual mileage. Such frontiers can be viewed as the limit, or maximum amount of expenditure the individuals/households are willing to incur, or the expenditure budget available for consumption. We invoke this notion to use stochastic frontier models for estimating the budgets for consumption. Following the two-stage budgeting approach discussed earlier, the estimated budgets can be used for subsequent analysis of choices and allocations to different choice alternatives of interest. The same assumptions discussed earlier, such as weak separability of preferences, are needed here. However, an advantage of using the stochastic frontier approach over the traditional regression models (to estimate budgets) is that the frontier, by definition, is greater than the observed total expenditure. Therefore, the budget estimated using the stochastic frontier approach provides a "buffer" for the actual total expenditure to increase or decrease. This can be easily accommodated in the second stage consumption analysis (using KT models) by designating an *outside good* that represents the difference between the frontier and the actual expenditure on all the *inside goods* (i.e., choice

7

alternatives of interest to the analyst). Given the frontier as the budget, if the attributes of the choice alternatives change, the second stage consumption analysis allows for the total expenditure on the inside alternatives to change (either increase or decrease). Specifically, within the limit set by the frontier, the outside good can either supply the additional resources needed for additional consumption of inside goods or store the unspent resources. The theoretical basis of the notion of stochastic frontiers combined with the advantage just discussed makes the approach particularly attractive for estimating the latent budgets necessary for Kuhn-Tucker demand analysis.

Finally, we use various assumptions on the estimation of budgets for the MDCEV models. These assumptions on the budgets are not necessarily estimated as a function of socio-demographic characteristics or built environment. Instead, we specify an arbitrary budget amount greater than the observed expenditure. Therefore, similar to the stochastic frontier approach, the analyst can specify an outside good in the MDCEV models to represent the difference between the arbitrary budget and the total available expenditure. The outside good, in turn, allows for the total expenditure to increase or decrease due to changes in alternative-specific attributes. In the context of time-use, the analyst can use the natural available budget which is 24 hours or just an assumed time budget.

To compare the efficacy of the different approaches to estimate budgets for the MDCEV models, we performed two empirical assessments: (1) The analysis of out-of-home activity participation and time-use (time as budget) for non-working adults in the State of Florida, and (2) The analysis of household vehicle type/vintage holdings and usage (annual mileage as budget) in Florida. We present both empirical assessments to compare the efficacy

of all the above approaches both in terms of prediction accuracy and the reasonableness of the changes in the total expenditure due to changes in alternative-specific variables.

1.4 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter 2 provides an overview of the stochastic frontier modeling methodology and the MDCEV model. Chapter 3 presents the application of the proposed approach for an empirical analysis of daily out-of-home activity participation and time-use patterns in a survey sample of non-working adults in Florida. Chapter 4 provides another case study for an empirical analysis of household vehicle holdings and usage in a sample of households in Florida. Finally, Chapter 5 discusses the conclusions of the thesis along with avenues for future research.

CHAPTER 2 MODELING METHODOLOGY

2.1 Stochastic Frontier Model

The stochastic frontier modeling methodology is employed to model the underlying budget driving a choice situation that is unobserved. While the actual production levels and the inputs to the production can be observed, a latent production frontier is assumed to exist. Such a production frontier is the maximum possible production that can be achieved given the inputs.

Following Banerjee et al. (2007), consider the notation where T_i is the observed total expenditure for decision-maker i, τ_i is the unobserved frontier (i.e., the maximum possible total expenditure) for decision-maker i, v_i is a normally distributed random component specific to decision-maker I and u_i is a non-negative random component assumed to follow a half-normal distribution. Also, X_i is a vector of observable decision-maker characteristics, β is a vector of coefficients of X_i and $\varepsilon_i = (v_i - u_i)$.

Let τ_i be a log-normally distributed unobserved frontier of a decision-maker i, while T_i is a log-normally distributed observed expenditure of the decision. Both these variables are assumed to be log-normally distributed to recognize the positive skew in the distribution of observed expenditure and to ensure positive predictions. τ_i of a decision-maker is assumed to be a function of the decision-maker demographic, attitudinal, and built environment characteristics, as:

$$\ln(\tau_i) = \beta' X_i + v_i \tag{2}$$

The unobserved frontier can be related to the observed expenditure T_i as:

$$\ln(T_i) = \ln(\tau_i) - u_i \tag{3}$$

Note that since u_i is non-negative, the observed expenditure is by design less than the unobserved frontier.

Combining Equations (2) and (3) results in the following regression Equation:

$$\ln(T_i) = \beta' X_i + v_i - u_i = \beta' X_i + \varepsilon_i \tag{4}$$

In the above equation, the expression $\beta' X_i + v_i$ may be considered as representative of the location of the unobserved frontier for $\ln(T_i)$ with a random component v_i. Consistent with the formulation of the stochastic frontier model (Aigner et al, 1977), a half-normal distribution (with variance σ_u^2) is assumed for u_i and a normal distribution (with mean 0 and variance σ_v^2) is assumed for v_i. These two error components are assumed to be independent of one another to derive the distribution of ε_i as:

$$h(\varepsilon_i) = \frac{2}{\sqrt{2\pi}\sigma}\{1 - \Phi(\varepsilon_i\lambda/\sigma)\} \exp\left[-\frac{\varepsilon_i^2}{2\sigma^2}\right]; -\infty < \varepsilon_i < \infty \tag{5}$$

where, $\sigma^2 = \text{var}(v_i + u_i) = \sigma_v^2 + \sigma_u^2$, and $\lambda = \sigma_u/\sigma_v$. The ratio, λ, is an indicator of the relative variability of the sources of error in the model, namely v_i, which represents the variability among decision-makers, and u_i, which represents the portion of the frontier that remains unexpended (Aigner et al, 1977). The log likelihood function for the sample of observations is given by:

$$LL = \sum_{i=1}^{n} \ln\left[h(\varepsilon_i)\right] \tag{6}$$

Maximum likelihood estimation of the above function yields consistent estimates of the unknown parameters, β, σ_u and σ_v.

11

From Equation (2), one can write the unobserved frontier as: $\tau_i = \exp(\beta'X_i + v_i)$. Using this expression and the parameter estimates β and σ_v, once can compute the expected value of frontier for decision-maker i as:

$$E[\tau_i] = E\left[\exp(\beta'X_i + v_i) = \exp\left(\beta'X_i + \sigma_v^2\Big/2\right)\right] \tag{7}$$

The expected frontier may be used as the budget in the second-stage analysis.

2.2 MDCEV Model Structure

The models estimated in this study are based on Bhat's (2008) linear expenditure system (LES) utility form for the MDCEV model:

$$U_i(\mathbf{t}) = \sum_{k=1}^{K} \gamma_{ik}\psi_{ik}\ln\left(\frac{t_{ik}}{\gamma_{ik}} + 1\right) \tag{8}$$

In the above function, $U_i(\mathbf{t})$ is the total utility derived by a decision maker i from the decision maker consumption. Decision-makers are assumed to choose consumption patterns (i.e., which product to consume and the amount to consume) to maximize $U(\mathbf{t})$ subject to a linear budget constraint on the available budget. The specification of this constraint depends on the approach used for the total available budget. As discussed earlier, we tested three different approaches, as discussed next.

The first approach is the stochastic frontier approach, where the *frontier* (τ_i) is used as the budget; i.e., the linear constraint then becomes $\sum_{k=1 \text{ to } K} t_{ik} = \tau_i$. We use the expected value of frontier as an estimate for τ_i, resulting in $\sum_{k=1 \text{ to } K} t_{ik} = E[\tau_i]$ as the actual budget constraint used.

The second approach is to simply use the total expenditure (T_i), which is observed in the data for model estimation purposes and can be estimated via a log-linear regression model for prediction

12

purposes. In this case, the budget constraint would be $\sum_{k=1 \text{ to } K} t_{ik} = T_i$, where T_i is the total

expenditure. The third approach is to specify an arbitrarily assumed budget amount (greater than

the observed expenditures in the sample) on the right side of the budget constraint.

In the above formulation, when the stochastic frontier approach is used to determine the

budget, the first choice alternative ($k = 1$) in the utility function is designated as the *outside good*

that represents the difference between the expected frontier and the observed expenditure (i.e.,

$t_1 = \tau_i - T_i$), while the other alternatives ($k = 2, 3,…, K$) are the *inside goods* representing

different alternatives. Similarly, when an arbitrarily assumed budget (greater than the observed

expenditure) is used, the *outside good* represents the difference between the assumed budget and

the observed expenditure. On the other hand, when the observed expenditure (T_i) is itself used as

the budget, there is no *outside good* in the formulation.

In the utility function, ψ_{ik}, labelled the baseline marginal utility of decision-maker i for

alternative k, is the marginal utility of consumption with respect to alternative k at the point of

zero consumption. Between two choice alternatives, the alternative with greater baseline

marginal utility is more likely to be chosen. In addition, ψ_{ik} influences the consumption

quantities to alternative k, since a greater ψ_{ik} value implies a greater marginal utility of

consumption. 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 alternatives. Specifically, 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 consumption quantities.

13

The influence of observed and unobserved decision-maker characteristics and built environment measures are accommodated as $\psi_1 = \exp(\xi_1), \psi_k = \exp(\theta' z_k + \xi_k)$ and $\gamma_k = \exp(\delta' w_k)$ where, z_k and w_k are vectors of observed socio-demographic and built environment measures influencing the choice of and consumption quantity to alternative k, θ and δ are corresponding parameter vectors, and ξ_k ($k=1,2,...,K$) is the random error term in the sub-utility of alternative k. Assuming that the random error terms ε_k ($k=1,2,...,K$) follow the independent and identically distributed (iid) standard Gumbel distribution leads to a simple probability expression (see Bhat, 2005) that can be used in the familiar maximum likelihood routine to estimate the unknown parameters in θ and δ.

For more details on the formulation, properties, and estimation of the MDCEV model, the reader is referred to the papers by Bhat (2005) and Bhat (2008).

CHAPTER 3 ANALYSIS OF INDIVIDUAL'S ACTIVITY PARTICIPATION AND TIME-USE PATTERNS

3.1 Introduction

This chapter presents an empirical analysis of individuals' activity participation and time-use choices for assessing the efficacy of the different approaches to estimate (or assume) budgets for the MDCEV model. In the context of individuals' out-of-home activity participation and time-use, most applications of the KT approach consider an exogenously available total time budget that is allocated among different activity type alternatives. As discussed earlier, the KT approach itself does not allow either an increase or decrease in the total time expended in the activities of interest due to changes in the alternative-specific characteristics. In this chapter, we use the different approaches mentioned earlier to estimate time budgets of the MDCEV models.

The first approach used is the log-linear regression approach which models the total observed expenditure to estimate time budgets. Log-linear regression is used as opposed to linear regression to avoid situations where negative time budgets might be predicted. The concept of out-of home activity time expenditure (OH-ATE) is used to represent amount of time that people are spending in out-of home activities. Then, the estimated total OH-ATE is used in the MDCEV model prediction. Next, we propose the use of stochastic frontier approach to estimate time budgets for the MDCEV models. In the stochastic frontier approach, we use the notion of an out-of-home activity time frontier (OH-ATF) that represents the maximum

15

amount of time that an individual is willing to allocate to out-of-home (OH) activities in a day. Stochastic frontier regression is performed on the observed total out-of-home activity time *expenditure* to estimate the unobserved out-of-home activity time *frontier* (OH-ATF). The estimated *frontier* is viewed as a subjective limit or maximum possible time individuals can allocate to out-of-home activities and used to inform time budgets for a subsequent MDCEV model of activity time-use. Finally, we use various assumptions on the time budget, without necessarily estimating it as a function of individuals' demographic characteristics. These assumed time budgets include:

1. An arbitrarily assumed time budget of 875 minutes for every individual, which is equal to the total maximum observed OH-ATE in the sample plus 1 minute,

2. An arbitrarily assumed time budget of 918 minutes for every individual, which is equal to 24 hrs minus an average of 8.7 hours of sleep time for non-workers (obtained from the 2009 American Time-use Survey),

3. An arbitrarily assumed time budget of 1000 minutes for every individual,

4. 24 hrs (1440 minutes) as the total time budget for every individual in the sample, and

5. 24 hrs minus observed in-home activity duration.

The approaches listed above (1 to 5) specify an arbitrary budget amount greater than the observed OH-ATEs[2]. Therefore, similar to the stochastic frontier approach, the analyst can specify an outside good in the time-use model to represent the difference between the arbitrary budget and the total OH-ATE. The outside good, in turn, allows for the total OH-ATE to increase or decrease due to changes in alternative-specific attributes. The different approaches

---

[2] Among the approaches listed from 1 through 5, all approaches except *e* assume an equal amount of budget across all individuals, while 5 allows the budget to be different across individuals depending on the differences in their in-home activities. While the approach *e* (i.e., utilizing 24 hrs. minus in-home duration as the budget) does allow for different budgets across different individuals, it does not recognize the variation as a result of systematic demographic heterogeneity.

are compared based on the predictive accuracy (of the corresponding MDCEV models) and the reasonableness of the changes in time-use patterns due to changes in alternative-specific variables.

The rest of the chapter is organized as follows. Section 3.2 describes the Florida sample of the National Household Travel Survey (NHTS) data used for the empirical analysis. Section 3.3 presents the empirical results, and Section 3.4 concludes the chapter.

3.2 Data

3.2.1 Data Sources

The primary data source used for the analysis is the 2009 National Household Travel Survey (NHTS) for the state of Florida. The survey collected detailed information on all out-of-home travel undertaken by the respondents. The information includes trip purpose, mode of travel, and travel start and end time, and dwell time (time spent) at the trip destination. Several secondary data sources were used to derive activity-travel environment measures of the neighborhoods in which the sampled households are located. The secondary sources are: (1) 2009 property appraiser data for all 67 counties in Florida, (2) 2007 infoUSA business directory, (3) 2010 NAVTEQ data, and (4) GIS layers of: (a) all parcels in Florida from the property appraiser data, (b) employment from the 2007 infoUSA business directory, and (c) intersections from the NAVTEQ data.

3.2.2 Sample Formation

In order to prepare data for the analysis of the activity participation and time-use, several steps were undertaken:

1. In the person file, only the adult non-workers (aged 18 years or over) who were surveyed on a weekday that was not a holiday were selected.

17

2. Using the activity file, all out-of-home activities in the NHTS data were aggregated into eight broad activity categories: (1) Shopping, (2) Other maintenance (buying goods/services), (3) Social/Recreational (visiting friends/relatives, go out/hang out, visit historical sites, museums and parks), (4) Active recreation (exercise and playing sports), (5) Medical, (6) Eat out (going out for meal) (7) Pick up/drop, and (8) Other activities.

3. The amount of time spent in each of these activity categories was calculated by using the "dwell time" variable in the NHTS data. The time spent in in-home activities was computed as total time in a day (24 hours) minus the time allocated to the above mentioned out-home activities, sleep (8.7 hours , 2010 American Time Use Survey) , and travel activities.

4. To develop the activity-travel environment measures from secondary data sources, various GIS layers (from property appraiser, infoUSA and NAVTEQ data) were overlaid onto circular buffers centered on the NHTS household locations. The buffer sizes used for this purpose are: ¼ mile, ½ mile and 1 mile. Accessibility variables such as recreational accessibility (such as gymnasiums, parks), retail accessibility (such as department stores, financial institutions), and other accessibility were also created for a 5 mile buffer size centered on the household locations.

5. After preparing the data from the activity file and the person file, the activity-travel environment measures and the accessibility variables were added based on the household file. The records with missing or inconsistent data were removed from the final data set.

18

3.2.3 Data Description

Table 3.1 provides descriptive information on the estimation sample used in this analysis. The sample comprises 6218 individuals who participated in at least one out-of-home activity on the survey-day. Only the interesting characteristics of the sample are discussed here. A large portion of the sample comprises elderly; partly due to a large share of elderly in Florida's population and also due to a skew in the response rates of different age groups to the survey.

The dominant share of elderly in the sample explains a greater share of females, a higher than typical proportion of smaller size households, larger share of households without children and those with no workers, and predominantly urban residential locations. A large share of the sample is Caucasian, able to drive, and owns at least one vehicle in the household. Several other demographic variables reported in the table are relevant to the models estimated in this paper.

The last part of the table presents the OH activity participation and time-use statistics observed in the sample. On average, individuals in the sample spent around two-and-half hours on OH activities. Majority of them participated in shopping activities, followed by personal business, social/recreation, eat out, medical, active recreation, pickup/drop-off, and other activities. Note that the percentages of participation in different activities add up to more than 100, because a majority of individuals participate in multiple activities. On average, individuals in the sample participated in 2.6 OH activities; 32% participated in two activities and 36% participated in at least 3 activities. This calls for the use of the multiple-discrete choice modeling approach for modeling time-use. In terms of time allocation, those who participate in social recreation do so for an average of 2 hours. The average time allocation to shopping, personal business, active recreation, eat out, or medical activities ranges from 45 minutes to an hour, while that for pickup/drop-off and other activities is around 15 minutes.

19

While not reported in the tables, some useful patterns observed in the data and relevant to the modeling results presented later are: (a) greater proportion of females participate in shopping and social/recreation activities and for larger durations, (b) older people participate more in medical activities while younger people participate more in social/recreational activities, (c) those with a driver's license are likely to do more out of home activities, especially pickup/drop-off, (d) those with children undertake more pickup/drop-off activities, and (e) higher income individuals participate more in social and active recreation and eat out activities. In summary, the sample shows reasonable time allocation patterns that are typical of the non-working population in Florida.

3.3 Empirical Results

3.3.1 Stochastic Frontier Model of OH Activity Time Frontier (OH-ATF)

Table 3.2 presents the results of the stochastic frontier model for OH-ATFs. Interestingly, female non-workers are found to have larger OH ATFs than male non-workers in Florida. Upon closer examination, this result can be traced to larger participation of females in shopping and social/recreation activities that tend to be of larger duration. As expected, the frontier is larger for people of younger age groups and for those who have driver licenses. Blacks seem to have larger frontiers than Whites and others; see Banerjee et al. (2007) for a similar finding. Internet use is positively associated with OH-ATF. People from single person households, high income households, and zero-worker households tend to have larger OH-ATFs; presumably because of the greater need for social interaction for single-person households, greater amount of money among higher income households to buy home maintenance services and free-up time for OH activity (as well as greater affordability to consume OH activities), and lower time-constraints of zero-worker households. People living in urban locations have larger OH-ATFs than those in

20

rural locations, perhaps due to a greater presence of OH activity opportunities in urban locations. Mondays are associated with smaller perceived frontiers for OH non-worker activity, possibly due to pronounced OH activity pursued over the weekend just before Monday and also due to the effect of Monday being the first work day of the week. Several other demographic variables were explored but turned non-influential in the final model. These include education status, vehicle ownership, presence of children, and own/rent house. This may be because the income effects in the model act as surrogate for many of these variables.

The stochastic frontier models can be used to estimate the expected OH-ATF for each individual in the survey sample to generate a distribution of expected ATFs. The average value of the expected ATF in the estimation sample is around 400 minutes (6 and half hours), whereas the average total OH time expenditure is 152 minutes (about 2 and half hour), suggesting that people are utilizing close to 40% of their perceived time budgets for OH activity. Of course, the percentage utilization varies significantly with greater utilization for those with larger observed OH activity expenditures and smaller utilization for those with smaller observed expenditures.

3.3.2 Out-of-home Activity Time-use Model Results

We estimated seven different MDCEV models of time-use with different assumptions discussed earlier on time budgets. Overall, the parameters estimates from all the models were found to be intuitive and consistent (in interpretation) with each other and previous studies. This section presents (in Table 3.3) and discusses only the results of the model in which the expected OH-ATFs (estimated using the stochastic frontier approach) were used as the available time budgets.

The baseline utility parameters suggest that females are more likely (than males) to participate in shopping and pickup/drop-off activities but less likely to participate in active

21

recreation. With increasing age, social/recreational activities and pickup/drop-off activities reduce, while medical visits increase. As expected, licensed drivers are more likely to participate in all OH activities (i.e., they are likely to use a large proportion of their frontiers) and even more so for pickup/drop-off activities. Reflecting cultural differences, Whites are more likely to eat out than those from other races while those born in the US are more likely to eat, socialize and recreate out-of-home than immigrants. Individuals with a higher education attainment are more likely to undertake personal business (e.g., buy professional services) and active recreation. Those from households with children and households with more workers show lower participation in shopping and personal business but do more pickup/drop-off activities. Income, as expected, has a positive association with social/recreational activities, active recreation, and eating out. Several land-use variables were attempted to be included in the model, but only a few turned out marginally significant. Among these, accessibility to recreational land seems to encourage social recreation as well as active recreation; employment density (measured by # jobs within a mile of the household) and # cul-de-sacs within a quarter mile buffer (a surrogate for smaller amount of through traffic) are positively associated with active recreation. It remains to be seen, as explored later using policy simulations, if these variables have a practically significant influence on time-use. Finally, Monday is associated with smaller rates of social recreation and eat-out activities while Fridays attract higher rates of social recreation. Note that the baseline utility function for unspent time alternative (i.e., the outside good) does not have any observed explanatory variables in it, as the alternative was chosen as the base alternative for parameter identification in the utility functions of OH alternatives.

The satiation function parameters influence the continuous choice component; i.e., the amount of time allocation to each activity. The relative magnitudes of the satiation function

22

constants are largely consistent with that of the observed durations for different activities. For example, social recreational activities have a high satiation constant suggesting they are more likely to be pursued for longer durations. The unspent time alternative has the largest satiation constant reflecting that large proportions of the perceived OH-activity time frontiers in the sample are unspent. Females tend to allocate more time to shopping and social recreation but less time to active recreation, if they participate in these activities. People from middle age group tend to spend less time in social/recreation, while educational attainment is associated with larger time in active recreation. Mondays tend to have smaller time allocations for eating out, while Fridays are associated with larger time allocations to social/recreation and eating out. Finally, accessibility to recreational land has a positive influence on the time allocation to social/recreation and active recreation.

3.3.3 Comparison of Predictive Accuracy Assessments

This section presents a comparison of in-sample predictive accuracy assessments for the different MDCEV models estimated in this study based on different assumptions for OH activity time budgets. While it would be prudent to perform out-of-sample predictive assessments, we did not set aside a validation sample since the estimation sample size was not large. All predictions with the MDCEV model were undertaken using the forecasting algorithm proposed by Pinjari and Bhat (2011), using 100 sets of Halton draws to cover the error distributions for each individual in the data.

Table 3.4 presents the results for the observed and predicted activity participation rates with different assumptions on time budgets. The predicted participation rates for each activity were computed as the proportion of the instances the activity was predicted with a positive time allocation across all 100 sets of random draws for all individuals. In the row labeled "mean

23

absolute error," an overall measure of error in the aggregate prediction is reported. This measure is an average, across different activities, of the absolute difference between observed aggregate rate of participation and the corresponding aggregate predictions of rate of participation.

Several interesting observations can be made from these results. First, the MDCEV models that use budgets from the stochastic frontier model or the log-linear regression model exhibit a greater aggregate-level predictive accuracy than other MDCEV models. This is presumably because the budgets used for both the models are heterogeneous across individuals (based on their demographic characteristics), whereas other approaches do not systematically capture heterogeneity in the available time budgets across individuals. These results suggest the importance of capturing demographic heterogeneity in the available time budgets across different individuals for a better prediction of the daily activity participation by the MDCEV time-use model. Second, between the stochastic frontier and log-linear regression approaches, quality of the aggregate predictions is similar; albeit the predicted activity participation rates for the stochastic frontier approach are slightly better. Third, the predictive accuracy does not seem to differ significantly by the amount of total budget assumed if a constant amount is used as the budget for every individual in the sample. Specifically, the predictions were very similar between the models that assumed an equal amount of budget across all individuals – 875 minutes, 918 minutes, 1000 minutes, or 24 hours – albeit there seems to be deterioration in the predictions as the assumed budget amount increases.

To compare the observed and the predicted duration of participation in each activity type, the distributions of the observed and the predicted activity durations from different budget estimations were plotted in the form of box-plots in Figure 3.1. The predicted average duration for an activity was computed as the average of the predicted duration across all random draws for

24

all individuals with a positive time allocation. There are 8 sub-figures in Figure 3.1, one for each activity type. Comparing the different approaches, the results clearly show that log-linear regression approach and stochastic frontier regression approach perform better than the other approaches. This is probably because the budgets estimated from both the regression models consider heterogeneity in budgets across individuals (based on their demographic characteristics), whereas other approaches do not capture heterogeneity in the available time budgets across individuals. These results indicate the importance of capturing demographic heterogeneity in the available time budgets across different individuals for a better prediction of the daily activity time-allocations by the MDCEV time-use model. Between stochastic frontier regression and log-linear regression approaches to estimating time budgets, the log-linear approach resulted in better predictions of time allocation to different activity types. When an equal amount of time budget is assumed, there is no significant difference in the predictions of activity time-allocations, although there seems to be deterioration in the predictions as the assumed time budget increases.

3.3.4 Simulation of Land-use Effects on Time-use Patterns

This section presents the predictions of a hypothetical policy scenario using the different MDCEV models estimated in this study based on different approaches for time budgets. The policy scenario considered in this exercise is doubling of accessibility to recreational land-use.

To simulate the effects of this hypothetical policy, in the first step, time budgets were estimated for both the base-case and the policy-case (i.e., before-policy and after-policy, respectively).[3] However, since the corresponding variable – accessibility to recreational land –

---

[3] For the log-linear regression and stochastic frontier regression approaches, the time budgets were estimated by simply taking the expected value of the corresponding regression equations. For other approaches where deterministic amounts of time budgets were assumed for all individuals in the sample (i.e., approaches 1 to 5 in Section 3.1), those same assumptions were used for prediction as well.

25

does not appear in either the log-linear regression or the stochastic frontier regression equations, the estimated time budgets do not differ between the base-case and the policy-case. Similarly, the time-budget remains the same between the base-case and the policy-case when an arbitrarily assumed deterministic time-budget is used (i.e., approaches 1 to 5 in Section 3.1). In the second step, the time budgets from the first step were used as budgets for the corresponding MDCEV time-use models (along with the MDCEV parameter estimates) to simulate out-of-home time-use patterns in the base-case and policy-case. Subsequently, the policy effect was quantified as two different measures of differences in time-use patterns between the policy-case and base-case: (1) The percentage of individuals for whom the time allocation to different activities changed by more than minute[4], and (2) The average change in time allocation for whom the time allocation changed by more than a minute. Table 3.5 reports these measures for the different approaches/assumptions used in the study for estimating time budgets. Specifically, in each row (i.e., for each approach used to estimate time-budget) for each column (i.e., for an activity type), the % number represents the percentage of individuals for whom the time allocated to the corresponding activity changed by more than a minute. The number in the parenthesis adjacent to the % figure is the average change in time allocation (in minutes) for whom the time allocation to that activity changed by more than a minute. Several observations can be made from this table, as discussed next.

First, across all different approaches for arriving at time budgets, consistent with the MDCEV model parameter estimates, increasing accessibility to recreational land-use has increased the time allocation to OH social and active recreational activities. For example, with the stochastic frontier approach for time budgets, doubling accessibility to recreational land lead

---

[4] We report only those for whom the time allocation changed by more than a minute (and the average change in time allocation only for those individuals) as opposed to all individuals for whom the time allocation changed. This helps in avoiding the consideration of instances when changes in time allocation are negligible (i.e., less than a minute).

26

to an increased time allocation (by more than a minute) for 3% individuals in social recreation activities and for 2.2% individuals in active recreation activities; among these individuals, on average, the time spent in social recreation increased by 21 minutes and that in active recreation increased by 25 minutes, respectively.

Second, upon examining where the additional time for social and recreational activities comes from, the MDCEV model based on the log-linear regression approach for time budgets differs considerably from the other MDCEV models. Specifically, using estimated OH-ATEs from the log-linear regression as budgets leads to a simple reallocation of the time (i.e., the estimated OH-ATE) between different activity types. That is, all of the increase in time allocation to social and recreational activities must come from a decrease in the time allocation to other activities. This is a reason why the predicted increases in the social and recreational activity participation rates are the smallest (and for a smaller percentage of individuals) for the log-linear regression approach. On the other hand, the stochastic frontier approach provides a "buffer" in the form of an unspent time alternative from where the additional time for social and active recreational pursuits can be drawn. Therefore, the increase in the time allocation to social and active recreational activities comes partly from a reduction in the "unspent time" and partly from other OH activities. This reflects an overall increase in the total OH activity expenditure (OH-ATE) than a mere reallocation of the base-case OH-ATE. Such an increase in the total OH-ATE can be measured by the decrease in the time allocated for the "unspent time" alternative; for example, an average of 21 minutes for the stochastic frontier approach. Intuitively speaking, it is reasonable to expect that an increase in accessibility to recreational land would lead to an increase in social and active recreation activity and there by an overall increase in OH activity time among non-workers, as opposed to a mere reallocation of time across different OH

27

activities. This demonstrates the value of the stochastic frontier approach in allowing more reasonable effects of changes in alternative-specific explanatory variables in the MDCEV model.

Third, similar to the stochastic frontier approach, other approaches that assume an arbitrary budget greater than observed OH-ATEs also allow a "buffer" alternative. In fact, the policy forecasts from all these approaches are similar to (albeit slightly higher than) those from the stochastic frontier approach. But recall that their base-case predictions (against observed time-use patterns) were inferior compared to the stochastic frontier approach. Therefore, it might be better to use the stochastic frontier approach than making arbitrary assumptions on the time budgets.

3.4 Summary and Conclusion

This chapter presents an empirical case study of individuals' daily activity time-use analysis to evaluate different approaches to estimating budgets for the multiple discrete-continuous extreme value (MDCEV) models. Among the different approaches, the proposed stochastic frontier regression is used to estimate time budgets for individuals' daily out-of-home time-use analysis. Specifically, we use the notion of an out-of-home activity time *frontier* (OH-ATF) that represents the maximum amount of time that an individual is willing to allocate to out-of-home (OH) activities in a day. First, a stochastic frontier regression is performed on the observed total out-of-home activity time *expenditure* (OH-ATE) to estimate the unobserved out-of-home activity time *frontier* (OH-ATF). The estimated frontier is viewed as a subjective limit or maximum possible time individuals are willing to allocate to out-of-home activities and used to inform time budgets for a subsequent MDCEV model of activity time-use. The efficacy of the proposed approach is compared with the following other approaches to estimate budgets for the MDCEV model:

28

1. Using total OH-activity time expenditure (OH-ATE), estimated via log-linear regression, as the time budget, and

2. Various assumptions on the time budget, without necessarily estimating it as a function of individual's demographic and built environment characteristics.

The comparisons were based on predictive accuracy and reasonableness in the results of hypothetical scenario simulations, including changes in land-use accessibility. The overall findings from this empirical exercise are summarized below.

1. Employing time budgets obtained from the stochastic frontier approach (to estimate OH-ATF) and the log-linear regression approach (to estimate the OH-ATE) provide better predictions of OH activity and time-use patterns from the subsequent MDCEV models, when compared to employing arbitrarily assumed time budgets. This is presumably because the former approaches allow for the time budgets to vary systematically based on individual's demographic characteristics, while the latter approaches assume an arbitrary budget that does not allow demographic variation in the budgets.

2. Estimating budgets using the log-linear approach for a subsequent MDCEV model provided better predictions of the activity durations observed in the survey sample, when compared to estimating budgets using the stochastic frontier approach.

3. The stochastic frontier approach allows for the total OH activity time expenditure to increase or decrease due to changes in alternative-specific variables. On the other hand, using time budgets from the log-linear regression approach lead to a mere reallocation of time between the different OH activities without increasing the total time allocated for OH activities. This is an important advantage of the stochastic

29

frontier approach over the traditional log-linear regression approach to estimating activity time budgets.

4. When arbitrarily assumed time budgets were considered, the predictive accuracy and policy simulation outcomes (in terms of the changes in OH time allocation patterns) did not differ significantly between the different assumptions as long as an equal time budget was assumed for all individuals.

Table 3.1 Descriptive Statistics of the Estimation Sample

Person Characteristics		Household Characteristics	
Sample Size	6,218	Sample Size	4,766
Age 18 - 24 years 25 – 64 years 65+ years	 1.40% 33.80% 64.70%	Household Size 1 Person 2 Person 3+ Person	 24.60% 55.80% 19.60%
Gender Male Female	 42.80% 57.20%	Annual Income < $ 25 K $ 25 K - $50 K $ 51 K - $75 K > $ 75K	 29.00% 33.20% 15.30% 22.60%
Race White African American Other	 90.30% 5.30% 4.4.%	Vehicle Ownership 0 Vehicle 1 Vehicle 2 + Vehicle	 4.70% 39.10% 56.20%
Education Level High School or less Some College Bachelor/Higher	 40.80% 28.40% 30.80%	Number of Workers 0 Workers 1 Worker 2 Workers 3+ Workers	 69.50% 26.50% 3.30% 0.80%
Driver Status Driver Not a Driver	 91.70% 8.30%	Number of Drivers 0 Drivers 1 Driver 2 Drivers 3+ Drivers	 2.90% 31.80% 56.40% 8.90%
Internet Use Almost Everyday Several Times in a week Sometimes (once in a week or in a month) Never	 46.30% 10.30% 6.40% 37.00%	Number of Children 0 Children 1 Child 2 Children 3+ Children	 90.10% 4.90% 3.30% 1.60%
Average duration spent in out-of-home activities (minutes)	152.8	Residential Area Type Urban Rural	 78.90% 21.10%

Persons' Out-of-Home Activity Participation and Time-Use Characteristics									
	Total observed OH Activity Time	Shopping	Personal Business	Social/ Recreational	Active Recreation	Medical	Eat Out	Pick-Up/ Drop Off	Other
% Participation	100	63.3	39.1	37.6	26.3	29.9	32.5	20.1	7.7
Average Duration (min.)*	152.8	54.6	49.6	124.1	52.7	60.1	47.9	15.3	20.8

 * Average among those who participated in the activity.

31

Table 3.2 Parameter Estimates of the Out-of-Home Activity Time Frontier (OH-ATF) Model

Variables	Coefficients (t-stats)
Constant	6.03 (138.28)
Female	0.08 (3.97)
Young age; 18-29 years (mid age is base)	0.11 (1.89)
Old age; >75 years (mid age is base)	-0.08 (-3.48)
Black (white and others are base)	0.09 (2.12)
Licensed to drive	0.12 (3.46)
Uses internet at least once a week (no use is base)	0.08 (3.48)
Single person household	0.19 (4.96)
Low income < 25K/annum (medium income is base)	-0.07 (-2.92)
High income >75K/annum (medium income is base)	0.05 (2.00)
Zero-worker household	0.07 (2.73)
Urban residential location (rural is base)	0.04 (1.87)
Monday (Tuesday - Friday is base)	-0.09 (-3.74)
$\hat{\sigma}_u$	1.7164 (84.97)
$\hat{\sigma}_v$	0.2851 (23.37)
Log-likelihood at constants	-8739.99
Log-likelihood at convergence	-8675.99

32

Table 3.3 Parameter Estimates of MDCEV Out-of-Home Activity Time-Use Model

	Unspent Time	Shopping	Personal Business	Social/Rec.	Active Rec.	Medical	Eat Out	Pickup /Drop	Other
Baseline Utility Variables									
Constants	-	-1.03(-14.67)	-1.87(-26.23)	-2.10(-21.00)	-2.57(-31.11)	-2.39(-26.65)	-2.91(-23.32)	-2.92(-17.16)	-3.74(-48.29)
Female (Male is base)	-	0.05(1.24)	-	-	-0.11(-2.08)	-	-	0.09(1.59)	-
Age <30 years (30-54 is base)	-	-	-	0.59(4.82)	-	-	-	-	-
Age 55-64 years	-	-	-	-	-	0.10(1.21)	-	-0.30(-3.17)	-
Age 65-74 years	-	-	-	-	-	0.14(1.78)	-	-0.43(-4.60)	-
Age >= 75 years	-	-	-	-0.07(-1.33)	-	0.32(4.15)	-	-0.60(-6.17)	-
White (Non-white is base)	-	-	-	-	-	-	0.39(4.09)	-	-
Driver (Non-driver is base)	-	-	-	-	-	-	-	0.48(3.33)	-
Driver (All OH activities)	-	0.28(4.72)	0.28(4.72)	0.28(4.72)	0.28(4.72)	0.28(4.72)	0.28(4.72)	0.28(4.72)	0.28(4.72)
Some College (< college is base)	-	-	0.16(2.96)	-	-	-	-	-	-
Bachelor's degree or more	-	-	0.25(4.74)	-	0.28(4.87)	-	-	-	-
Born in US (others is base)	-	-	-	0.11(1.63)	-	-	0.30(3.86)	-	-
# Children aged 0-5 years	-	-0.12(-1.80)	-0.23(-2.68)	-	-	-	-	0.38(5.29)	-
# Children aged 6-15 years	-	-	-	-	-	-	-	0.46(9.01)	-
Total Number of Workers	-	-0.04(-1.25)	-	-	-	-	-	0.16(3.20)	-
Income 25- 50 K	-	-	-	0.06(1.12)	-	-	0.24(3.74)	-	-
Income 50-75 K	-	-	-	0.06(1.12)	0.21(2.84)	-	0.28(3.65)	-	-
Income >75 K	-	-	-	0.06(1.12)	0.41(6.47)	-	0.45(6.51)	-	-
Accessibility to recreational land	-	-	-	0.0059(1.84)	0.0052(1.45)	-	-	-	-
# Employments (1mile buffer)	-	-	-	-	0.0009(1.96)	-	-	-	-
# Cul-de-sacs (0.25 mile buffer)	-	-	-	-	0.007(1.29)	-	-	-	-
Monday (Tue.-Thurs.is base)	-	-	-	-0.14(-2.35)	-	-	-0.21(-3.22)	-	-
Friday (Tue.-Thurs.is base)	-	-	-	0.06(1.11)	-	-	-	-	-
Satiation Function Variables									
Constants	4.66(109.28)	2.83(63.91)	3.01(86.02)	4.42(88.96)	1.60(15.88)	3.27(76.43)	3.14(63.06)	1.45(30.21)	2.22(30.58)
Female (Male is base)	-	0.24(4.14)	-	0.12(2.02)	-0.13(-1.33)	-	-	-	-
30-54 years(<30 & >55 years-base)	-			-0.27(-2.52)					
Some College (< college is base)	-	-	-	-	0.45(3.66)	-	-	-	-
Bachelor's degree or more	-	-	-	-	0.76(6.46)	-	-	-	-
Monday (Tue.-Thurs. - base)	-	-	-	-	-	-	-0.19(-1.76)	-	-
Friday	-	-	-	0.12(1.22)	-	-	0.25(2.57)	-	-
Accessibility to recreational land	-	-	-	0.0051(0.91)	0.023(3.39)	-	-	-	-

33

Table 3.4 Predictive Performance of MDCEV Time-use Models with Different Assumptions on Time Budgets

Activities	Observed activity rate of participation and predicted activity rate of participation							
	Observed	Log-linear Regression	Stochastic Frontier	Budget = 875 min.	Budget = 918 min.	Budget = 1000 min.	Budget = 1440 min. (24 hrs.)	Budget = 24hrs-in home duration
Shopping	63.3%	67.1%	58.0%	56.0%	55.9%	55.7%	55.4%	53.7%
Personal Business	39.1%	45.9%	37.3%	35.9%	35.9%	35.9%	35.9%	34.2%
Social Recreation	37.6%	43.4%	34.7%	33.6%	33.5%	33.5%	33.4%	31.8%
Active Recreation	26.3%	27.8%	23.3%	22.5%	22.5%	22.5%	22.6%	21.1%
Medical	29.9%	32.1%	26.6%	25.5%	25.5%	25.5%	25.6%	24.0%
Eat Out	32.5%	35.9%	29.5%	28.4%	28.4%	28.4%	28.5%	27.0%
Pickup /Drop-off	20.1%	22.3%	18.4%	17.7%	17.7%	17.7%	17.9%	16.7%
Other Activities	7.7%	8.1%	6.8%	6.5%	6.5%	6.5%	6.6%	6.0%
Mean Absolute Error	-	3.28	2.74	3.80	3.83	3.86	3.83	5.25
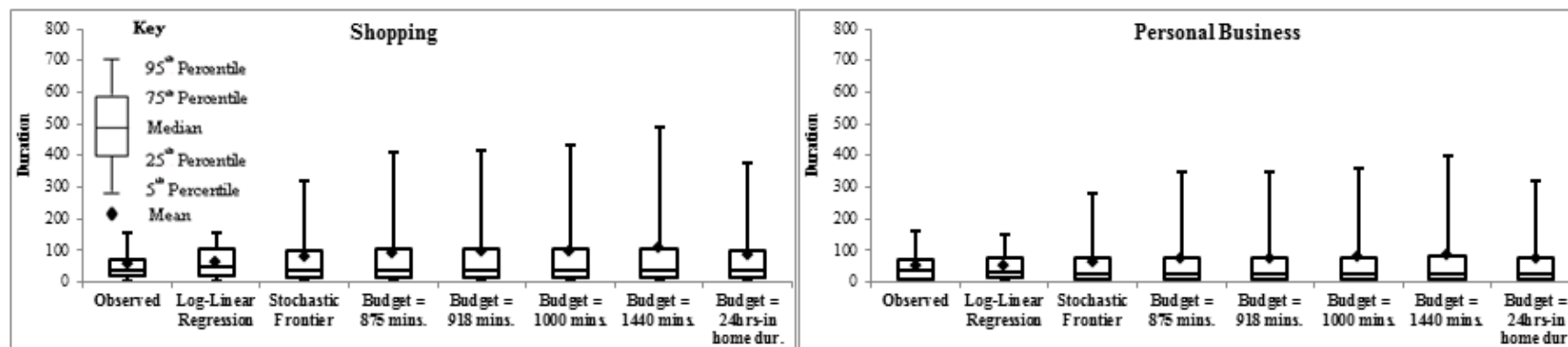
Figure 3.1 Observed and Predicted Distributions of Activity Durations with Different Approaches for Time Budgets

34
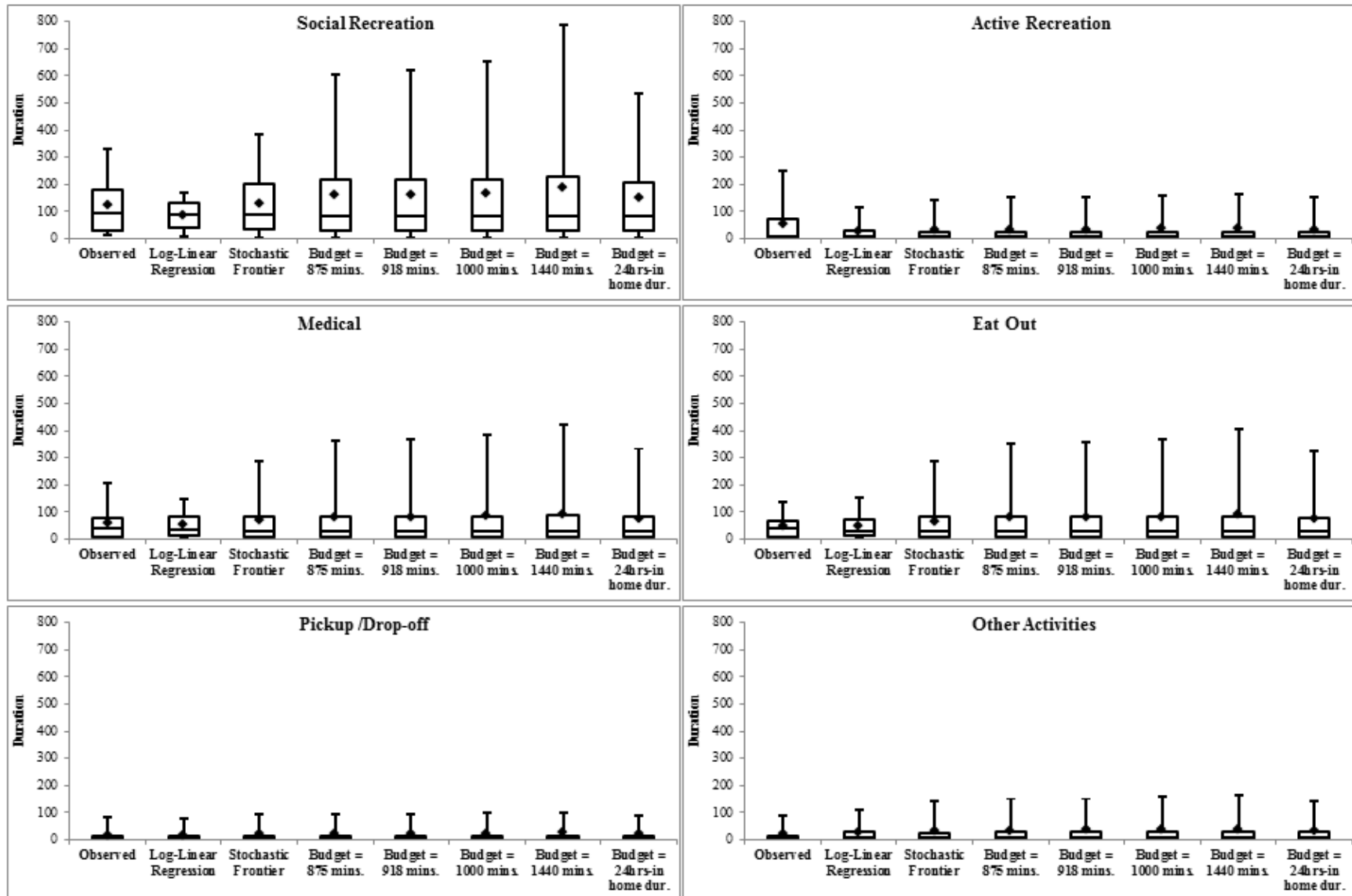
Figure 3.1 (Continued)

Table 3.5 Simulated Land-use Impacts on OH Time-use Patterns for MDCEV Models with Different Approaches for Time Budgets

MDCEV model with budget from…	Unspent Time	Shopping	Personal Business	Social Recreation	Active Recreation	Medical	Eat Out	Pickup /Drop-off	Other
Log-linear Regression	--	-2.5% (-9)	-1.7% (-8)	2.2% (13)	2.1% (18)	-1.1% (-9)	-1.3% (-8)	-0.6% (-4)	-0.2% (-4)
Stochastic Frontier Regression	-3.6% (-21)	-1.9% (-7)	-1.3% (-7)	3.0% (21)	2.2% (25)	-0.9% (-8)	-1.1% (-7)	-0.3% (-4)	-0.2% (-4)
Budget = 875 minutes	-4.8% (-24)	-1.7% (-7)	-1.1% (-6)	3.9% (21)	2.3% (28)	-0.8% (-7)	-0.9% (-7)	-0.2% (-5)	-0.1% (-4)
Budget = 918 minutes	-4.9% (-24)	-1.6% (-7)	-1.1% (-6)	4.0% (21)	2.3% (28)	-0.8% (-7)	-0.9% (-7)	-0.2% (-5)	-0.1% (-4)
Budget = 1000 minutes	-5.1% (-24)	-1.6% (-7)	-1.0% (-6)	4.1% (21)	2.4% (29)	0.8% (-7)	-0.9% (-7)	-0.2% (-5)	-0.1% (-5)
24hrs-in home duration	-4.3% (-21)	-1.6% (-7)	-1.0% (-6)	3.4% (21)	2.2% (27)	-0.8% (-7)	-0.8% (-7)	-0.2% (-5)	-0.1% (-4)

Note: In each cell, the % number indicates the % of individuals for whom the time allocated to an activity increased or decreased by more than a minute. A positive (negative) number indicates the % of individuals for whom the time allocated to the corresponding activity increased (decreased) by more than a minute. The numbers in the parentheses indicate the average change in the time allocated (minutes) for whom a change occurred in the time allocation to this activity by more than a minute. Positive number indicates an increase in the time allocation while a negative number indicates a decrease in the time allocation. For example, with time budgets estimated using log-linear regression, the MDCEV model predicts that doubling accessibility to recreational land leads to a decrease in the time allocated to shopping by more than a minute for 2.5% of the individuals in the sample. And the average decrease in time allocation to shopping activity for these same individuals is 9 minutes.

CHAPTER 4 ANALYSIS OF HOUSEHOLDS' VEHICLE OWNERSHIP AND UTILIZATION

4.1 Introduction

As previously discussed, the purpose of this thesis is to compare different approaches to estimate (or assume) budgets for MDCEV model. The different approaches tested include log-linear regression, stochastic frontier and various assumptions on the budgets. In this chapter, we present an empirical analysis of households' automobile ownership and utilization patterns in order assess the efficacy of those approaches.

In the U.S., household automobiles are the predominant mode of travel. According to Purcher and Renne (2003), 92% of households in the US owned at least one vehicle in 2001 compared to 80% in the 1970s, and 87% of daily trips were made by personal-use motorized vehicles. Therefore, analyzing household vehicle ownership (*i.e.,* number and types of vehicles owned) and utilization (*e.g.,* miles driven per year) patterns can be valuable for forecasting vehicle travel demand and for devising relevant policies.

A large body of literature exists on household vehicle ownership and utilization patterns in the United States. Several of these studies have analyzed household vehicle holding based on body type (Lave and Train, 1979; Kitamura et al., 2000; Choo and Mokhtarian, 2004; Bhat and Sen, 2006), body type and vintage (Berkovec and Rust, 1985; Mohammadian and Miller, 2003a; You et al, 2014), make/model (Manski and Sherman, 1980; Mannering and Winston, 1985), make/model and vehicle acquisition type (Mannering et al., 2002), vehicle make/model/vintage and vehicle ownership level (Berkovec, 1985; Hensher et al., 1992), and joint vehicle

make/model and vehicle type/vintage (Bhat et al, 2009). Household vehicle holdings can be analyzed using the four most common modeling structures: multinomial logit model (Lave and Train, 1979; Manski and Sherman, 1980; Mannering and Winston, 1985; Kitamura et al., 2000), nested logit (Hocherman et al., 1983; Berkovec and Rust, 1985; Berkovec, 1985; Mannering et al., 2002) and multiple discrete-continuous extreme value (Bhat and Sen, 2006; Ahn et al., 2008; Jaggi et al., 2011; Bhat et al, 2009; You et al, 2014), and reduced-form discrete-continuous choice models (Fang, 2008). Multinomial logit model (MNL) and nested logit models only analyze situations where the decision-makers are allowed to choose a single alternative from a set of available alternatives. Whereas, MDCEV model formulation recognizes that households may simultaneously own and use multiple vehicle types to meet various functional needs of the household. For instance, households may use vans for family vacations and use smaller vehicles for work, grocery shopping, etc. Chapter 2 provides a review of the formulation of the MDCEV model.

Several studies used the MDCEV model to analyze household vehicle holding and usage. Among these, Bhat and Sen (2006) analyzed household vehicle holdings and usage using data from the 2000 San Francisco Bay Area Travel Survey. A multiple discrete-continuous extreme value (MDCEV) model is used to perform the analysis. They analyze the impact of household demographics (number of children, household size, and number of employed adults), residence location variables and vehicle operating cost of the type of vehicles that households own and use. In the paper, Bhat and Sen demonstrated the application of the model by analyzing the influence of an increase in operating cost due to an increase in fuel cost (from \$1.40/gallon to \$2.00/gallon). They found that the increase in operating cost resulted in a marginal decrease in vehicle ownership of passenger cars and a significant decrease in the ownership of SUVs and

38

minivans. In addition, they found that households would use passenger cars (compact, subcompact, large sedans, etc.) more than other vehicle types as a result of this change.

In another study, Bhat et al. (2009) formulated and estimated a nested model structure that includes a multiple discrete-continuous extreme value (MDCEV) component for the analysis of vehicle type/vintage holdings and usage in the upper level and multinomial logit model (MNL) component to analyze the choice of vehicle make/model for a given vehicle type/vintage in the lower level. The data used for this analysis is the 2000 San Francisco Bay Survey (BATS). Their results suggested that high income households have a low preference for older vehicles and are unlikely to use non-motorized vehicles as a mode of transportation. In addition, they found that household location attributes and built environment characteristics have significant impacts on vehicle ownership and usage. Bhat et al. applied the model to demonstrate the effect of increasing bike lane density, street block density and fuel cost by 25%. They observed that the increase in bike lane density and street block density have negative impacts on the holdings and usage of vehicle types. Also, they found that the increase in fuel cost leads to a shift from the ownership of larger vehicles to the ownership of smaller and more fuel efficient vehicles.

Both studies contributed to a better understanding of the variables that impact vehicle holdings and usage. However, those studies used the observed total annual mileage as the budget to model household vehicle holdings and use. This approach does not allow an increase or decrease in the total available mileage expenditure due to changes in alternative-specific attributes. The policy simulations used in the above-discussed studies only lead to a reallocation of the total annual mileage expenditure among different vehicle types. For instance, the models do not allow that an increase in operating cost might lead to a decrease in the total mileage expenditure among household vehicles. To address this issue, Bhat et al (2009) included a non-

39

motorized alternative in their model; the mileage for this non-motorized alternative was aggregated across all household members that spent time walking and biking on the two days of the survey and projected to an annual level. While the presence of the non-motorized alternative allows for the total mileage on motorized household vehicles to decrease as a result of increases in operating costs, the model necessarily implies an equal amount of increase in non-motorized mileage. This may not necessarily hold in reality.

4.2 Contribution and Organization of the Chapter

In this chapter, stochastic frontier approach and other approaches are used to estimate mileage budgets for analyzing household vehicle ownership (by type and vintage) and usage in Florida. In the stochastic frontier approach, the concept of a total annual mileage frontier (AMF) is used to represent the maximum amount of miles a household is willing to travel in a year. First, a stochastic frontier regression is performed on the observed total annual mileage *expenditure* to estimate the unobserved annual mileage *frontier* (*i.e.,* the AMF*).* The estimated *frontier* is viewed as a subjective limit or the maximum possible annual vehicle miles that a household is willing to travel. The estimated frontier is used as the mileage budget for a subsequent MDCEV model of vehicle usage. Second, in the MDCEV model we used several attributes to analyze vehicle ownership and usage: a) vehicle body type, b) vehicle age (*i.e.,* vintage), c) vehicle make and model, and d) vehicle usage (*i.e.*, miles driven per year). The combination of vehicle body type and vintage was used to create choice alternatives for the MDCEV model. However, it is difficult to include vehicle specific attributes such as purchase price, horsepower, engine size, fuel type and other variables in the MDCEV model. This is because for each body type and vintage category, the household can have several different make and model options to choose from. A descriptive analysis of the data indicated that for

40

any vehicle body type and vintage most households own only one make and model. Therefore, for each vehicle type/vintage chosen, we use a multinomial logit structure to analyze the choice of a single vehicle make and model (Bhat et al., 2009). We use logsum variables to connect the MNL model (lower level of the nest) to the MDCEV model (upper level of the nest). The logsum variables carry the information on vehicle specific attributes from the MNL model to the MDCEV model. In the MDCEV model, several sets of determinants of vehicle holdings and usage decisions were tested: household demographics, individual characteristics, and built environment characteristics. Finally, policy simulations are conducted to demonstrate the value of the stochastic frontier approach in allowing the total annual mileage *expenditure* to either expand or shrink within the limit of the *frontier* implied by the stochastic frontier model.

As mentioned earlier, the stochastic frontier approach is compared with several other approaches to estimate budgets for the MDCEV model. The following approaches were tested:

1. The stochastic frontier regression model is used to estimate annual mileage frontiers (AMF).

2. A log-linear regression model is used to estimate the annual mileage expenditure (AME). Log-linear approach is used to ensure that the estimated mileage budgets are positive.

3. Non-motorized mileage was calculated for each household to be used as the unspent mileage alternative (*i.e.*, outside good). To calculate the non-motorized mileage, we arbitrarily assumed a walking distance of 0.5 miles per day for all household members (> 4 years old) for 100 days a year. The budget for this scenario is the sum of non-motorized annual mileage and total observed annual mileage expenditure.

41

4. An arbitrary budget of 119505 miles is assumed for every household, which is equal to the maximum observed annual mileage expenditure (AME) in the dataset (119405 miles) plus 100 miles. This budget is uniform across households.

It is worth noting here that the budget estimated using the log-linear regression approach is an estimate of the total annual mileage expenditure (AME), all of which is utilized for vehicle types/vintages. On the other hand, the other approaches listed above (3 and 4) specify an arbitrary budget amount greater than the observed AME. Therefore, similar to the stochastic frontier approach, the analyst can specify an outside good in the vehicle-use model to represent the difference between the arbitrary budget and the AME. The outside good, in turn, allows for AME to increase or decrease due to changes in alternative-specific attributes. In this chapter, we present an empirical analysis to compare the efficacy of all the above approaches both in terms of prediction accuracy and the reasonableness of the changes in vehicle use patterns due to changes in alternative-specific variables.

The rest of the chapter is organized as follows. Section 4.3 presents the methodology. Section 4.4 presents the data sources used in this analysis, the sample formation and data description. Section 4.5 presents the empirical results, and Section 4.6 concludes the chapter.

4.3 Data

4.3.1 Data Sources

The primary data source used for this analysis is the Florida add-on of the 2009 US National Household Travel Survey (NTHS). The survey collected detailed information on vehicle fleet compositions for over 15,000 households. The information collected on household ownership are the make/model for all vehicles in the households, the year of the manufacture for each vehicle, the miles driven per year, the year of possession for each vehicle, etc. Additional

42

vehicle information such as fuel economy, fuel cost, and annual mileage was added on version 2 of the 2009 NHTS.   In addition, the survey also collected information on individual demographics (age, gender, race, education, etc.), household demographics (income, number of children, etc.) and activity travel characteristics (purpose, mode of transportation, start and end time, etc.).

Several other secondary sources were used to derive the dataset for the analysis. First, vehicle specific attributes such as engine horsepower, vehicle weight (pounds), engine size (liters) and cylinders, type of wheel drive (all-wheel, front-wheel, 4-wheel and rear-wheel), transmission type (manual and automatic), seat capacity, number of doors and fuel type (regular, premium, diesel and electric) were obtained for each vehicle make/model from CarqueryAPI.com (carqueryAPI, 2014). Additional vehicle attributes such as purchase price, luggage volume (non-trucks) and payload capacity (for tucks only) were obtained for each vehicle make/model from Motortrend.com (Motor trend, 2014).

4.3.2 Sample Formation

In order to perform the analysis, two datasets were prepared: (1) a dataset for the MNL model of vehicle make/model choice, and (2) another dataset for the MDCEV model of vehicle type/vintage holdings and utilization.

4.3.2.1 Data Formation for MNL

In this sub-section, the procedures that were undertaken to prepare the dataset for the multinomial logit model of vehicle make/model are described. The following steps were taken:

1. First, the vehicles in the vehicle file were categorized into nine distinct vehicle types. The nine vehicle types are: (1) Compact (2) Subcompact (3) Large Sedan (4) Mid-size Sedan (5) Two-seater (6) Van (minivan and cargo van) (7) Sports Utility Vehicle

43

(SUV) (8) Pickup Truck and (9) Motorcycle. Other vehicle types such as Recreational Vehicles (RVs) and other vehicle types were removed from the dataset.

2. Second, three vintages were created using vehicle age, which is the difference between the year of the survey (2009) and the year of the manufacture of the vehicle. The three vintages were: (1) 0 to 5 years (2) 6 to 11 years and (3) 12 years or older.

3. Next, secondary data sources were added to the files. Households with missing vehicle attributes (for e.g., vehicle age, purchase price, horsepower, weight, etc.) and missing socio-demographics information (for e.g. income) were removed.

4. Due to the dissimilarities in motorcycle characteristics to the other vehicle types, we excluded motorcycles from this analysis. Only 8 vehicles types and the 3 vintages, for a total of 24 vehicle type/vintage classes, were used in the MNL model. Within the 24 vehicle type/vintage classes, households have large number of makes/models choice sets. Therefore, similar to Bhat et al (2009), for each vehicle class we collapsed the makes/models into commonly held distinct makes/models and grouped the other makes/models into a single 'other" make/model category. We defined "commonly held vehicle" if a vehicle make/model is more than 0.5% of the total vehicles in that vehicle type/vintage category.

5. Next, for the "other" make/model category we used an average vehicle attributes for the all vehicle make/models that belonged to that vehicle type/vintage.

6. The sample size for the MNL model comprised of 19,749 vehicles from 11,488 households.

44

4.3.2.2 Data Formation for MDCEV

This sub-section describes the procedures to prepare the dataset for the MDCEV model. The steps of the data setup are as follows:

1. Using the nine vehicle classes (including motorcycles) and the three vintages mentioned in MNL data formation, we have a total of 27 vehicle type/vintage categories.

2. Due to formulation constraint, we only retained households that own no more than 1 vehicle type/vintage.

3. The "BESTMILE" variable in the vehicle file from the NHTS 2009 was used as the annual mileage for each vehicle.[5] The total annual mileage was calculated for each household by taking the sum of annual miles driven for all the vehicle type/vintage categories own by that household.

4. For practical reasons, we removed households that have vehicles with an annual mileage greater than 50k. We also removed households with a total annual mileage of fewer 100 miles.

5. Since the analysis is at the household level, in order to include individuals' characteristics we assumed that the head of the household make vehicle decisions in the household[6].

6. Finally, we cleaned the person file to obtain information on socio-demographic characteristics about the head of the household, such as age, gender, ethnicity, education and employment status. The final sample comprises of 10,294 records with each record represents households with at least one vehicle. We randomly selected

---

[5] See NHTS 2009 user guide BESTMILE for detailed information about the computation of the BESTMILE variable.
[6] Similar to Bhat et al (2009), the head is assigned as the employed individual in one-worker household. If all the adults in a household were unemployed, or if more than 1 adult was employed, the oldest member was defined as the household head.

45

8,500 households from the 10,294 for model estimation and we kept the rest 1,794 households for data validation.

4.3.3 Sample Description

Table 4.1 shows the sample characteristics for the MDCEV dataset used for the analysis. The dataset consists of 8,500 households with at least one vehicle. The first part of the table shows the descriptive statistics of the head of household characteristics. The results show that there is a larger proportion of household heads who are males (59.3%) compared to females (40.7%). The table also shows that 44.9% of the household heads are elders (>65 years old) while only 1.8% of them are between the ages of 18 to 29. This is partly because Florida is an attractive location for elderly individuals and partly because of the NHTS survey sample might be skewed more toward elderly (simply because of better survey response rates from elderly individuals than from the younger demographic segments). A very large proportion of the household heads are white (90.0%). 39.2% of householders have a Bachelor's degree or higher, while 31.4% have a high school diploma or less. The second part of the table shows the household characteristics and household location characteristics. The sample size comprises mostly one or two household members, 25.6% and 50.7%, respectively. About 30.1% of the households make more than $75,000 per year while 21.6% make less than $25,000. There are also a larger proportion of households that own 2 vehicles (44.4%), and households that have no workers (44.4%). A very large proportion of the households has no children (85.4%) and lives in urban areas (78.9%).

Table 4.2 shows the number of vehicles in each of the 24 vehicle type/vintage categories (except the motorcycles) used in the model. An MNL model was estimated for each of these vehicle type/vintage categories to analyze the vehicle make and model choice. The third column

46

shows the number of distinct make and model alternatives (for the MNL model) for each vehicle type/vintage category. It is observed that there are more SUVs of 0 to 5 years (11.4%), mid-size sedans of 0 to 5 years (9.3%) and mid-size sedans of 6 to 11 years (8.7%) in the dataset. Two-seaters comprise only 1.5% of the sample. It can be observed that there is a preference for newer vehicles in Florida. There are more vehicle make/model choices available for SUVs compared to other vehicle types. There are 52 vehicle makes/models for SUVs of 0 to 5 years while there are only 14 to 17 vehicle makes/models for pickup trucks.

Table 4.3 shows the descriptive statistics of household vehicle type/vintage holdings and utilization. The second and third columns describe the total number of household owning vehicle type/vintage and the average annual mileage for the vehicle type/vintage, respectively. First, households in Florida have a higher preference for SUVs of 0 to 5 years old (17.8% of households), mid-size sedans of 0 to 5 years old (15.3% of households), mid-size sedans of 6 to 11 years old (14.4% of households) and SUVs of 6 to 11 years old (12.6% of households). This suggests a high baseline preference for SUVs of 0 to 5 years, mid-size sedans of 0 to 5 years, mid-size sedans of 6 to 11 years and SUVs of 6 to 11 years. Second, there is a low percentage of households owning motorcycles (1.8% of households own motorcycles of 0 to 5 years, 1.5% own motorcycles of 6 to 11 years and 1.2% own motorcycles older than 12 years) and two-seaters (1.2% own two-seaters of 0 to 5 years, 1.1% own two-seaters of 6 to 11 years and 1.1% own two-seaters of 12 years or older). It is also seen that these two vehicle types (motorcycles and two-seaters) have a low annual mileage usage rate. This suggests a low baseline preference and high satiation for motorcycles and two-seaters. The results further depict that there is a high utilization rate for vans of 0 to 5 years (13,184 miles per year), pickup trucks of 0 to 5 years (13,046 miles per year) and SUVs of 0 to 5 years (12,851 miles per year). This suggests a low

47

satiation for these vehicle types/vintages. The results also indicate that households tend to use new vehicle types more compared to older vehicle types. For example, the average annual mileage expenditure for vans of 0 to 5 years, 6 to 11 years and 12 years or older are 13,184 miles, 11,222 miles and 8,898 miles, respectively.

The last three columns in table 4.3 show households owning one vehicle, households owing 2 vehicles and households with 3 or more vehicles for each vehicle type/vintage category. The results show that there are 624 households that own and use large sedans of 0 to 5 years. Out of the 624 households, 251 of households (40.2%) own and use large sedans of 0 to 5 years only, 278 (44.6%) own and use large sedans of 0 to 5 years plus another vehicle type/vintage, and 95 (15.2%) own and use large sedans of 0 to 5 years plus two or more other vehicle types/vintages. The results further indicate that households that own and use SUVs and vans are more inclined to own and use at least another vehicle type/vintage. This might be because SUVs and vans are mostly used for family obligations (for e.g., taking kids to school, family vacation, etc.) and those household members tend to use other vehicle types for their personal trip (for e.g., work). The results also indicate that households that own and use motorcycles are more likely to own and use two or more other vehicle types. This is perhaps motorcycles are mostly used for leisure and personal trips, but cannot be used for trips like shopping, taking kids to school and family vacations. Finally, households that own and use two-seaters tend to have other vehicle types which can be mainly be due to seating capacity.

4.4 Methodology

This section presents the methodology used to analyze households' vehicle holdings and usage. Stochastic frontier approach and other different approaches are used to estimate mileage budgets for the MDCEV models. The modeling structures for stochastic frontier and MDCEV

48

are presented in chapter 2. Two modeling components were used, including a multiple discrete-continuous extreme value (MDCEV) component to analyze the choice of vehicle type/vintage and usage in the upper level and a multinomial logit (MNL) component to analyze the choice of vehicle make/model in the lower level. Logsum variables are used to carry the impacts of vehicle-specific attributes of the MNL model to the MDCEV model. The MNL model structure and the structure of the logsum variables are presented in the sub-sections below.

4.4.1 MNL Model Structure

Multinomial logit model, one of most conventional discrete choice models, is based on a random utility maximization approach. In the case of vehicle make/model choice, given a vehicle type/vintage this approach assumes that a household will select the vehicle make/model that provides the maximum utility from a set of available vehicle make/model alternatives. In this approach, the utility function comprised of two components, including the observed component which can be measured as a function vehicle make/model attributes and the unobserved component which cannot be measured. This utility function can be expressed as:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{1}$$

where, U_{in} is the total utility of vehicle make/model i to household n , V_{in} is the observed portion of the utility of vehicle make/model i to household n and ε_{in} is the error component. The error components are assumed to follow a Gumbel distribution (i.e., identically and independently distributed type 1 extreme values) and also assumed to be independent from the irrelevant alternative (IIA property of the MNL model). Based on these assumptions, the probability expression of the MNL model can be written as:

49

$$P_{in} = \frac{e^{V_{in}(\beta)}}{\sum e^{V_{jn}(\beta)}} \qquad (2)$$

where P_{in} is the probability of household *n* choosing vehicle make/model *i* and β is a vector of coefficients of the household characteristics and vehicle make/model specific attributes.

4.4.2 Log-sum Variables

Logsum terms were constructed to carry out the effect of different vehicle makes/models from the MNL model to the MDCEV model. The logsum term, representing the maximum expected utility from the MNL model, is a natural log of the sum of exponents of deterministic utility terms from the MNL model. Specifically, the logsum terms are computed using the following expression:

$$\text{Logsum} = \ln\left(\sum_{j} e^{Vjn} \right) \qquad (3)$$

where V_{jn} is utility of vehicle make/model *j* for household *n*. Since we have 24 vehicle type/vintage alternatives from the MDCEV model used in the MNL model, then we have 24 different logsum variables.

4.5 Empirical Results

4.5.1 Stochastic Frontier Model of Annual Mileage Frontier (AMF)

Table 4.4 presents the results for the parameter estimates of the annual mileage frontier (AMF) models. The results indicate that households with a male householder tend to have larger annual mileage frontiers (AMFs) than households with female householder. The results also suggest that households with a householder between the age of 18 to 29 years, and 30 to 54 years have larger AMFs relative to households with older householders. The model results also indicate that AMFs for lower income households tend to increase with income. Further, the results indicate that the number of drivers, number of workers and presence of children in the

50

households have larger AMFs, presumably because more household members will create more travel needs. An increase in fuel cost ($/gallon), as expected, tends to decrease households' AMFs. Households located in rural areas tend to have larger AMFs compared to households located in urban areas. This is expected since household members living in rural areas have to drive longer distances to different activity locations. Finally, households located in high employment density and high residential neighborhoods have lower AMFs, possibly due greater accessibility to employment and other activity opportunities.

4.5.2 Log-Linear Model of Total Annual Mileage Expenditure (AME)

The specification used for the total annual mileage frontiers (AMFs) is also used for the log-linear regression (see Table A.1). It is good to emphasize that the stochastic frontier model predicts the unobserved total annual mileage frontier that households are assumed to perceive, whereas log-linear regression model predicts the observed total annual mileage expenditure (AME). The results are interpreted the same way as in the stochastic model. However, we use the concept of annual mileage expenditure (AME) as opposed to annual mileage frontier (AMF). For instance, the results indicate that as household income increases, the total annual mileage expenditures (AME) also increase. It is also shown that the number of drivers, number of workers and presence of children in the households have positive impacts on total annual mileage expenditures. Households with two or more members have larger annual mileage expenditures than single person households. Fuel cost, employment density and residential density have a negative impact on AMEs. Finally, Households in rural areas tend to drive more than households in urban areas.

The log-linear regression model results can be used to estimate the annual mileage expenditure for each household in the survey sample. This estimated mileage is used to generate

51

a distribution of expected annual mileage expenditure. Such a distribution is plotted in Figure 4.1, along with the distribution of the observed total annual mileage expenditures in the sample. The expected (or estimated) annual mileage is shown in red dotted line and the observed annual mileage expenditure (AME) is in blue solid line. It is seen that the expected annual mileage expenditures closely follow the distribution of the observed total annual mileage. The average observed total annual mileage is about 18,010 miles whereas that for the expected annual mileage is 20,163 miles.

4.5.3 Multinomial Logit Model Results for Vehicle Make/Model Choice

Table 4.5 presents the multinomial logit model results for vehicle make/model choices conditional on the choice of vehicle type/vintage category. For cost variables, the results suggest that households prefer vehicle makes/models that are less expensive to purchase and operate (see, Lave and Train, 1979, Hocherman et al., 1983, Berkovec and Rust, 1985 Mannering and Winston, 1985, Bhat et al., 2009, for similar results). For households that own pickup trucks, the results indicate that these households have a higher preference for pickup trucks with high standard payload capacity (see, Bhat et al., 2009, for similar results); possibly because pickup trucks are mainly used for heavy duty work such as hauling of construction material. Next, the vehicle engine performance was captured by engine size and the ratio of engine horsepower to vehicle weight. The results show that households have a greater preference for vehicle makes/models with greater performance. It is also found that households in Florida have a higher preference for vehicle makes/models with all-wheel-drive compared to rear-wheel-drive, if an all-wheel-drive model is available in a specific vehicle type category. Finally, Table 4.5 shows that households are less likely to prefer vehicle makes/models that use premium fuel compared to

regular fuel (see, Bhat et al., 2009, for similar results); this is intuitive, since premium fuel is more expensive than regular fuel.

4.5.4 MDCEV Model Results for Vehicle Type/Vintage Holdings and Utilization

Several different MDCEV models of vehicle type/vintage holdings and usage were estimated with different assumptions for mileage budgets. Overall, the parameter estimates from all the models were found to be intuitive and consistent (in interpretation) with each other and previous studies. This section presents and discusses only the results of the model in which the expected total annual mileage frontiers (estimated using the stochastic frontier approach) were used as mileage budgets. The results are presented in Table 4.6.

4.5.4.1 Baseline Utility

The household income effect suggests that high income (> \$75,000) and mid-income (\$50,000 to \$75,000) households have lower baseline preferences for older vehicle types (12 years or older) relative to low income households. The results further indicate that high income households also have lower baseline preferences for mid-age (6 to 11 years) vehicle types. Also, high income households have a higher baseline preference for two-seaters regardless of vintage. The results also show that high and mid-income households have a higher baseline preference for new SUVs (Kitamura et al., 2000; Choo and Mokhtarian, 2004; Bhat et al., 2009, for similar results) relative to low income households. The results also suggest that lower income households (< \$25,000) tend to own and use older vehicle types relative to low income households (\$25,000 to \$50,000).

Households with senior adults (>65 years old) have a higher baseline preference for compact, large and mid-size vehicles and a lower baseline preference for older subcompact vehicles relative to households with no senior adults. This can be because senior adults prefer

53

vehicle types that they can easily get in and out of. The model results also suggest that households with more children are more likely to own and use vans; this is expected since vans are more convenient to transport families. Larger households in Florida have a higher baseline preference for mid-size and older SUVs compared to smaller households. Also, the results indicate that households with more workers are less likely to own and use large sedans and new vans. This is perhaps because households with more workers prefer to drive alone, leading to preferences for vehicles with less seating capacity.

For householder characteristics, the results suggest that males (*i.e.*, head of the household is male) are more likely to own and use pickup trucks, motorcycles and old vans compared to females. Older households (*i.e.*, age of the head of household) have higher baseline preferences for large sedans of 6 to 11 years and vans of 6 to 11 years. Head of households that are between the ages 31 to 45 years are more likely to own and use motorcycles. Ethnicity variables were found to have impacts on vehicle type/vintage holdings and usage. The results suggest that blacks tend to use old large sedans, mid-age and old mid-size sedans and are less likely to prefer trucks compared to other ethnic groups. Also, the results suggest that Hispanics prefer large sedans and Asians are less likely to own and use trucks and more likely to prefer old compacts.

Finally, several household location characteristics were tested in the model; only rural area (urban is base), employment density and residential density were found to have significant impacts on vehicle type/vintage holdings and usage. Households located in rural areas have higher baseline preferences for pickup trucks compared to households located in urban areas. The results also indicate that households located in high residential density neighborhoods prefer vans, SUVs and pickup trucks compared to households located in less dense neighborhoods. The results further suggest that households located in high employment density neighborhoods have

54

low baseline preferences for pickup trucks. These results are intuitive since high density areas have spatial constraints for parking which lead to a preference for smaller vehicles.

4.5.4.2 Log-sum Parameter

The logsum variables were created separately for each vehicle type/vintage (except for motorcycles since they were not included the MNL model). The logsum variables help to carry the effect of vehicle-specific attributes from the lower level (MNL of vehicle make/model choice) to the upper level (MDCEV of vehicle holdings and utilization). Logsum parameters were estimated for multiple combinations of vehicle type/vintage but the estimates were found to be more than 1. Therefore, to be consistent with utility maximization, the logsum parameter for all vehicle type/vintage categories was fixed to 1.

4.5.4.3 Baseline Constants

The baseline constants are presented in the second part of table 4.6. The baseline constants provide an indication of preferences for various vehicle types, and the marginal utility at zero consumption for different alternatives. The results suggest that households have higher baseline preferences for mid-size, SUVs and compacts compared to other vehicle types. New mid-size sedans and SUVs (0 to 5 years) have the highest baseline utility which suggests a preference for new mid-size and SUVs. These results are consistent with the survey data set.

4.5.4.4 Satiation Parameters

The satiation parameters represent the diminishing marginal utility with increasing consumption of various alternatives and the extent to which households are inclined to drive various vehicle types. A high satiation parameter for a vehicle type/vintage means that those households are more likely to drive that vehicle type/vintage (less satiated). The results suggest that households tend to drive new vans and SUVs (0 to 5 years) more than other vehicle

55

types/vintages. This is intuitive, because vans and SUVs are used to transport families for vacations, activities, etc. Also, the results indicate that high income households tend to allocate less miles to new SUVs. Overall, the satiation parameters are higher for new vehicle types which suggest that households use new vehicles more than older vehicles.

4.5.5 Comparison of Predictive Accuracy Assessments Using Data Validation

This section presents a comparison of predictive accuracy assessments for different MDCEV models estimated using different approaches for estimating mileage budgets. As mentioned in the data formation for the MDCEV models, in the final sample formation we randomly selected 8,500 households out of 10,294 households. We kept the rest of the 1,794 households for validation data. We estimated (or assumed) the budgets for the validation data using the different approaches mentioned earlier. Subsequently, the corresponding MDCEV model parameter estimates were used on the households in the validation dataset to predict their vehicle type/vintage holdings and utilization (i.e., mileage) patterns. All predictions with the MDCEV model were undertaken using the forecasting algorithm proposed by Pinjari and Bhat (2011), using 100 sets of Halton draws to cover the error distributions.

Table 4.7 presents the results for the observed and predicted market share of each vehicle type/vintage using different approaches used for estimating mileage budgets on the validation data (see Table A.2 for predictions on the estimation data in Appendix A). The predicted holdings for each vehicle type/vintage were computed as the proportion of the instances the vehicle type/vintage was predicted with a positive mileage allocation across all 100 sets of random draws for all households. In the row labeled "mean absolute error," an overall measure of error in the aggregate prediction is reported. This measure is an average, across different vehicle types/vintages, of the absolute difference between the observed aggregate percentage of

56

vehicle type/vintage holdings and the corresponding aggregate predictions of the percentage of vehicle/vintage holdings. Overall, all the approaches resulted in similar results except for the arbitrarily assumed budget approach (119,505 miles) – although, log-linear regression resulted in slightly better predictions.

Using the different approaches for estimating mileage budgets, we predicted the mileage allocation for each vehicle type/vintage. The predicted average mileage for a vehicle type/vintage was computed as the average of the predicted mileage across all random draws for all households with a positive mileage allocation. To compare the different approaches used to estimate mileage budgets, we plotted the distributions of the observed mileage and the predicted mileage for each vehicle type/vintage using different approaches for the mileage budgets. The distributions were plotted in the form of box-plots in Figure 4.2 for the validation data (see Figure A.1 for predictions for the estimation data in Appendix A). There are 28 sub-figures in Figure 4.2, one for each vehicle type/vintage and one for the unspent mileage alternative (difference between the used budget and the observed annual mileage expenditure). In the unspent mileage sub-figure, log-linear regression approach is not present because log-linear regression models the annual mileage expenditures, all in which are used in mileage allocations to different vehicle types/vintages. The results show that the non-motorized unspent mileage is significantly less than stochastic frontier and assumed budget equal to 119,505 miles. This is because the average unspent annual mileage for the non-motorized approach is about 102 miles, whereas the stochastic frontier and the assumed budget approach (119,505 miles) are 18,445 miles and 101,495 miles, respectively. For all vehicle types/vintages, the results show that log-linear regression model performs better in predicting annual mileage expenditures compares to all other approaches. In addition, the results show that stochastic frontier approach and the

57

assumed budget approach (119,505 miles) over-predict the allocation of annual mileage expenditures. Overall, the results indicate that log-linear linear performed better in prediction mileage allocation to different vehicle types/vintages.

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

The MDCEV models (estimated from different approaches for mileage budgets) can be used to determine the change in the holdings and usage of vehicle types/vintages due to changes in the independent variables. Here, we compare the different approaches by examining the effect of increasing fuel economy (miles/gallon) of different vehicle types/vintages on the holdings and mileage allocation patterns of vehicle types/vintages. Specifically, we increase fuel economy for new (0 to 5 years) compact, subcompact, large and mid-size vehicles by 25%. This change reflects in the "Fuel Cost ($/year)/Income ($/year)" variable in the MNL model for vehicle make/model choice. The logsum variables were used to carry this change to the MDCEV models.

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). Using different approaches to estimate mileage budgets, we were able to simulate vehicle holdings and usage for the base-case and the policy-case. Then, the policy effect was quantified as two different measures of differences between the policy-case and base-case: 1) The percentage change of holdings of vehicle type/vintage and 2) the average change of mileage for the households in which a change occurred in the mileage allocation. Table 4.8 presents the results of the simulation for different approaches used to estimate mileage budgets. For each approach used for mileage budget, there

58

are two columns: 1) The "% Change in Holdings" column shows the percentage change in the holdings of the corresponding vehicle type/vintage, and 2) The "Change in Mileage" column indicates the average change in mileage for households in which a change occurred in mileage allocation for each vehicle type/vintage.

Several observations can be made from those results. First, the results show that an increase in fuel economy of new (0 to 5 years) compact, subcompact, large and mid-size leads to an increase in the holdings of the new compact, subcompact, large and mid-size across all approaches. For instance, with the stochastic frontier approach for mileage budgets, the increase in fuel economy leads to an increase in the holdings of new compact, new subcompact, new large and new mid-size vehicles by 1.28%, 0.95%, 1.02% and 1.12%, respectively. The results also indicate a decrease in the holdings of almost all other vehicle type/vintages across almost all approaches used for mileage budgets. Overall, this is intuitive since an increase in fuel economy reduces operating cost and households prefer vehicles that are less costly to operate (consistent with MNL results). Comparing the different approaches, it is observed that stochastic frontier predicts a higher percentage change in the holdings of new compact, subcompact, and mid-size vehicles. Specifically, with log-linear regression approach as the mileage budget, the increases in the holdings of the vehicles mentioned above are less compared to the increases of other approaches. Second, for the average change in mileage, the results show that an increase in fuel economy for new compact, subcompact, large and mid-size vehicles leads to an increase in the usage of the corresponding vehicle type/vintage across all approaches used for mileage budgets. For example, with stochastic frontier approach, the average change in mileage for new compact, subcompact, large and mid-size vehicles is 431 miles, 243 miles, 322 miles and 325 miles, respectively. Also, the results indicate a decrease in the average mileage for all other vehicle

59

type/vintages. However, within a vehicle type it is observed that there is a higher decrease in the usage of older vehicle types as compared to newer vehicle types. For example, with stochastic frontier approach, the decrease in the mileage allocation of SUVs of 0 to 5 years, SUVs of 6 to 11 years and SUVs older than 12 years is 107 miles, 138 miles and 171 miles, respectively. This is intuitive since older vehicles tend to have lower fuel economy compared to newer vehicles, which makes older vehicle types more expensive to operate. Third, when examining where the additional mileage for new compact, subcompact, large and mid-size vehicles comes from, log-linear regression approach significantly differs from other approaches. In the log-linear regression approach, the mileage budget (i.e., estimated budget from log-linear regression) is simply reallocated between the different vehicle types/vintages. That is, all of the increases in mileage allocation to new compact, subcompact, large and mid-size must come from a decrease in the mileage allocation to other vehicle types/vintages. On the other hand, stochastic frontier approach and the other approaches provide a "buffer" in the form of an unspent mileage alternative from where the additional mileage can be drawn. As a result, the increase in mileage allocation to new compact, subcompact, large and mid-size are mainly from a decrease in the "unspent mileage" alternative and some decreases from other vehicle types/vintages. For the non-motorized approach, however, it is seen that the additional mileage is mainly from other vehicle types/vintages. That is because the unspent mileage is so low that there are not enough unspent miles available to be drawn from. Finally, the presence of an "unspent mileage" alternative allows the AME to increase as a result of improvement in fuel economy for certain vehicle types/vintages. In Table 4.8, the last row labelled as "Change in total expenditure" indicates the average change in the total expenditure (*i.e.*, only mileage allocated to inside goods) for the households in which a change occurred in the mileage allocation. The results suggest that

60

the mileage expenditures for stochastic frontier approach, non-motorized approach, and assumed budget approach (119,505 miles) increase by 258 miles, 10 miles and 554 miles, respectively.

4.6 Summary and Conclusion

This chapter presents an analysis of households' vehicle holdings and usage to compare different approaches to estimate budgets for Kuhn-Tucker demand systems (specifically for MDCEV model). The empirical case study analyzes the vehicle ownership and usage patterns of households with at least one vehicle in a survey sample from the state of Florida. In this study, we have two modeling components – a multiple discrete-continuous extreme value (MDCEV) component to analyze the choice of vehicle type/vintage and usage and a multinomial logit (MNL) component to analyze the choice of vehicle make/model for each vehicle type/vintage alternative. Several different approaches were used to estimate mileage budgets for the MDCEV models including stochastic frontier approach (to estimate annual mileage frontier), log-linear regression (to estimate annual mileage expenditure), non-motorized mileage (budget equals to the sum of non-motorized mileage and observed expenditure) and maximum observed annual mileage plus 100 miles (budget equals to 119,405 miles + 100 miles). The different approaches are compared based on the predictive accuracy (of the corresponding MDCEV models) and the results of hypothetical policy scenario of increasing fuel economy for new compact, subcompact, large and mid-size vehicles. The overall findings from this empirical exercise are summarized below.

In terms of the prediction of aggregate vehicle type/vintage holding and usage patterns, all the approaches resulted in similar results except for the arbitrarily assumed budget approach (119,505 miles). However, estimating budgets from log-linear regression resulted in slightly better predictions. For the stochastic frontier approach, the MDCEV model resulted in over-

predictions of annual mileage for different vehicle type/vintage alternatives, when compared to the predictions from the log-linear approach.

In the context of policy simulation results, using budgets estimated from the log-linear regression approach does not allow for increases or decreases in total annual mileage on household vehicles due to changes in alternative-specific characteristics. It only allows a reallocation of the total annual mileage among different vehicle type/vintage alternatives. On the other hand, the stochastic frontier approach allows for the total annual mileage expenditure to increase or decrease due to changes in alternative-specific variables. This is an important advantage of the stochastic frontier approach over the traditional log-linear regression approach to estimating budgets.

Overall, the results show that for the predictive accuracy of vehicle holdings and usage log-linear is approach better than stochastic frontier. However, stochastic frontier approach provides better results when it comes to simulating the effects of changes in alternative-specific attributes.

Table 4.1 Sample Characteristics

Sample Size	8500
Head Household Characteristics	
Gender	
Male	59.3%
Female	40.7%
Age	
18-29	1.8%
30-54	30.1%
55-65	23.2%
65+	44.9%
Race	
White	90.0%
African-American	5.1%
Other	4.9%
Education	
High school or less	31.4%
Some college	29.4%
Bachelor or higher	39.2%
Household Characteristics	
Household Size	
1	25.6%
2	50.7%
3+	23.7%
Household Income	
< $ 25 K	21.6%
$ 25 K - $50 K	30.4%
$ 51 K - $75 K	17.8%
> $ 75K	30.1%
Vehicle Ownership	
1	39.9%
2	44.4%
3+	16.7%
Number of Workers	
0 Workers	44.4%
1 Worker	34.0%
2 Workers	19.5%
3+ Workers	2.1%
Number of Children	
0 Children	85.4%
1 Child	8.0%
2 Children	5.2%
3+ Children	1.4%
Residential Type Area	
Urban	78.9%
Rural	21.1%

Table 4.2 Classification of the Vehicle Type/Vintage for the MNL Models

Vehicle Type/Vintage	Number of Vehicles (%)	Number of Make/Models Alternatives
Compact 0 to 5 years	1309 (6.6%)	36
Compact 6 to 11 years	1085 (5.5%)	45
Compact 12 years or older	556 (2.8%)	29
Subcompact 0 to 5 years	408 (2.1%)	23
Subcompact 6 to 11 years	334 (1.7%)	21
Subcompact 12 years or older	355 (1.8%)	27
Large 0 to 5 years	850 (4.3%)	25
Large 6 to 11 years	732 (3.7%)	19
Large 12 years or older	480 (2.4%)	20
Mid-size 0 to 5 years	1844 (9.3%)	32
Mid-size 6 to 11 years	1726 (8.7%)	35
Mid-size 12 years or older	585 (3.0%)	35
Two-seater 0 to 5 years	134 (0.7%)	21
Two-seater 6 to 11 years	116 (0.6%)	14
Two-seater 12 years or older	130 (0.7%)	13
Van 0 to 5 years	711 (3.6%)	20
Van 6 to 11 years	704 (3.6%)	22
Van 12 years or older	280 (1.4%)	20
SUV 0 to 5 years	2260 (11.4%)	52
SUV 6 to 11 years	1519 (7.7%)	41
SUV 12 years or older	412 (2.1%)	24
Pickup Truck 0 to 5 years	1214 (6.1%)	17
Pickup Truck 6 to 11 years	1212 (6.1%)	16
Pickup Truck 12 years or older	793 (4.0%)	14
Total	19749(100.0%)	-

64

Table 4.3 Descriptive Statistics of Vehicle Type/Vintage Holdings and Usage

Vehicle Type/Vintage	Total number (%) of household owning	Average Annual Mileage	Number of households (%) who own		
			only one vehicle type/vintage (1 vehicle household)	vehicle type/vintage + another vehicle type/vintage (2 vehicle households)	vehicle type/vintage + 2 other vehicle type/vintage (3+ vehicle households)
Compact 0 to 5 years	887 (10.4%)	11363	323 (36.4%)	410 (46.2%)	154 (17.4%)
Compact 6 to 11 years	802 (9.4%)	10471	275 (34.3%)	378 (47.1%)	149 (18.6%)
Compact 12 years or older	391 (4.6%)	8254	130 (33.2%)	177 (45.3%)	84 (21.5%)
Subcompact 0 to 5 years	301 (3.5%)	11104	69 (22.9%)	145 (48.2%)	87 (28.9%)
Subcompact 6 to 11 years	246 (2.9%)	9998	49 (19.9%)	125 (50.8%)	72 (29.3%)
Subcompact 12 years or older	251 (3.0%)	8276	45 (17.9%)	125 (49.8%)	81 (32.3%)
Large 0 to 5 years	624 (7.3%)	10754	251 (40.2%)	278 (44.6%)	95 (15.2%)
Large 6 to 11 years	566 (6.7%)	9573	227 (40.1%)	253 (44.7%)	86 (15.2%)
Large 12 years or older	336 (4.0%)	8282	141 (42.0%)	141 (42.0%)	54 (16.1%)
Mid-size 0 to 5 years	1299 (15.3%)	11079	469 (36.1%)	624 (48.0%)	206 (15.9%)
Mid-size 6 to 11 years	1223 (14.4%)	10183	462 (37.8%)	571 (46.7%)	190 (15.5%)
Mid-size 12 years or older	417 (4.9%)	7921	149 (35.7%)	193 (46.3%)	75 (18%)
Two-seater 0 to 5 years	101 (1.2%)	8625	10 (9.9%)	57 (56.4%)	34 (33.7%)
Two-seater 6 to 11 years	97 (1.1%)	8345	10 (10.3%)	50 (51.5%)	37 (38.1%)
Two-seater 12 years or older	93 (1.1%)	8193	12 (12.9%)	31 (33.3%)	50 (53.8%)
Van 0 to 5 years	522 (6.1%)	13184	147 (28.2%)	278 (53.3%)	97 (18.6%)
Van 6 to 11 years	522 (6.1%)	11222	140 (26.8%)	270 (51.7%)	112 (21.5%)
Van 12 years or older	195 (2.3%)	8898	59 (30.3%)	89 (45.6%)	47 (24.1%)
SUV 0 to 5 years	1512 (17.8%)	12851	363 (24.0%)	829 (54.8%)	320 (21.2%)
SUV 6 to 11 years	1067 (12.6%)	11920	240 (22.5%)	591 (55.4%)	236 (22.1%)
SUV 12 years or older	279 (3.3%)	9428	52 (18.6%)	124 (44.4%)	103 (36.9%)
Pickup Truck 0 to 5 years	852 (10.0%)	13046	102 (12.0%)	496 (58.2%)	254 (29.8%)
Pickup Truck 6 to 11 years	818 (9.6%)	11598	112 (13.7%)	475 (58.1%)	231 (28.2%)
Pickup Truck 12 years or older	540 (6.4%)	8948	69 (12.8%)	301 (55.7%)	170 (31.5%)
Motorcycle 0 to 5 years	153 (1.8%)	4305	7 (4.6%)	33 (21.6%)	113 (73.9%)
Motorcycle 6 to 11 years	126 (1.5%)	3461	7 (5.6%)	25 (19.8%)	94 (74.6%)
Motorcycle 12 years or older	99 (1.2%)	2194	6 (6.1%)	29 (29.3%)	64 (64.6%)
Total Observed Annual Mileage	-	18010	-	-	-

65

Table 4.4 Parameter Estimates of the Total Annual Mileage Frontier (AMF) Model

Variables	Coefficients	t-stats
Constant	10.53	18.07
Head Household Characteristics		
Male	0.06	4.22
Age 18 to 29 (age 55 to 74 is base)	0.23	4.23
Age 30 to 54 (age 55 to 74 is base)	0.14	7.46
Age >75 (age 55 to 74 is base)	-0.20	-10.51
Household Characteristics		
Income < 25k/year (Income 25k to 50k is base)	-0.04	-2.16
Income 50k and < 75k (Income 25 k to 50k is base)	0.13	6.03
Income >75k/year (Income 25k to 50k is base)	0.21	10.58
Number of drivers	0.27	15.49
Number of workers	0.15	13.08
Presence of children	0.04	1.71
2+ household members	0.18	7.32
Fuel Cost ($/gallon)		
Fuel cost	-0.35	-1.82
Household Location Attributes		
Rural(Urban is base)	0.15	8.14
Employment density	-0.002	-3.77
Residential density	-0.008	-2.62
$\hat{\sigma}_u$	1.108	
$\hat{\sigma}_v$	0.354	
Log-likelihood at constants	-11721.93	
Log-likelihood at convergence	-9281.75	
Number of observations	8500	

Table 4.5 Multinomial Logit Model Results for Vehicle Make/Model Choice

Variables	Coefficients	t-stats
Cost variables		
Purchase Price (in $)/Income (in $/yr.) [x 10]	-0.04	-4.94
Fuel Cost (in $/yr.) /Income (in $/yr.) [x 10]	-0.81	-9.23
Internal Dimension		
Standard Payload Capacity (for Pickup Trucks only) (in 1000s lbs.)	0.29	10.08
Performance		
Horsepower (in HP) /Vehicle Weight (in lbs.)	0.29	5.91
Engine Size (in liters)	-0.02	-2.79
Type of Drive Wheels		
Dummy Variable for All-Wheel-Drive (base: Rear-Wheel-Drive)	-0.11	-6.58
Fuel Type		
Dummy Variable for Premium Fuel	-0.36	-17.16
Log-likelihood	-65749.69	

Table 4.6 Parameter Estimates of MDCEV Model for Vehicle Ownership and Usage Using Stochastic Frontier

Baseline Utility		Baseline Utility		Baseline Utility	
Explanatory Variables	**Coef. (t-stat)**	**Explanatory Variables**	**Coef. (t-stat)**	**Explanatory Variables**	**Coef. (t-stat)**
Logsum	1.00(fixed)	**Large 12 years or older**		**Van 0 to 5 years**	
Compact 0 to 5 years		Income <25k	0.50(4.05)	Number of kids	0.31(6.09)
Income 50k to 75k	-0.17(-1.67)	Income 50k to 75k	-0.33(-5.23)	Residential density	-0.02(-1.98)
Income > 75k	-0.32(-3.78)	Income > 75k	-0.69(-11.55)	Number of workers	-0.30(-7.43)
Presence of senior	0.21(4.45)	Presence of senior	0.17(2.47)	**Van 6 to 11 years**	
Compact 6 to 11 years		Number of workers	-0.38(-4.48)	Income > 75k	-0.26(-6.74)
Income 50k to 75k	-0.25(-2.45)	Hispanic	-0.82(-1.79)	Number of kids	0.30(5.65)
Income > 75k	-0.26(-6.74)	Black	0.33(1.53)	Age 31 to 45 years old	0.55(1.35)
Presence of senior	0.20(3.97)	**Mid-size 0 to 5 years**		Age > 45 years	0.59(1.51)
Compact 12 years or older		Presence of senior	0.21(4.45)	Residential density	-0.03(-2.73)
Income <25k	0.28(2.30)	Household size	-0.11(-3.57)	**Van 12 years or older**	
Income 50k to 75k	-0.33(-5.23)	**Mid-size 6 to 11 years**		Income <25k	0.75(4.65)
Income > 75k	-0.69(-11.55)	Income > 75k	-0.26(-6.74)	Income 50k to 75k	-0.33(-5.23)
Presence of senior	0.17(2.47)	Presence of senior	0.20(3.97)	Income > 75k	-0.69(-11.55)
Asian	0.49(1.26)	Household size	-0.10(-3.02)	Number of kids	0.39(4.49)
Subcompact 0 to 5 years		Black	0.22(1.97)	Residential density	-0.03(-2.73)
Income > 75k	0.52(4.32)	**Mid-size 12 years or older**		Male	0.44(2.72)
Subcompact 6 to 11 years		Income <25k	0.34(2.89)	Age 31 to 45 years old	0.74(2.98)
Income > 75k	-0.26(-6.74)	Income 50k to 75k	-0.33(-5.23)	**SUV 0 to 5 years**	
Subcompact 12 years or older		Income > 75k	-0.69(-11.55)	Income 25 to 75 k	0.19(1.63)
Income 50k to 75k	-0.33(-5.23)	Presence of senior	0.17(2.47)	Income > 75k	0.33(2.76)
Income > 75k	-0.69(-11.55)	Household size	-0.04(-0.81)	Residential density	-0.02(-1.98)
Presence of senior	-0.35(-3.58)	Black	0.22(1.97)	**SUV 6 to 11 years**	
Large 0 to 5 years		**Two-seater 0 to 5 years**		Income > 75k	-0.26(-6.74)
Presence of senior	0.21(4.45)	Income > 75k	1.26(5.64)	Residential density	-0.03(-2.73)
Number of workers	-0.30(-7.43)	**Two-seater 6 to 11 years**		**SUV 12 years or older**	
Hispanic	-0.82(-1.79)	Income > 75k	0.54(3.65)	Income 50k to 75k	-0.33(-5.23)
Large 6 to 11 years		**Two-seater 12 years or older**		Income > 75k	-0.69(-11.55)
Income > 75k	-0.26(-6.74)	Income > 75k	0.54(3.65)	Household size	0.15(3.26)
Presence of senior	0.20(3.97)			Residential density	-0.03(-2.73)
Number of workers	-0.34(-5.71)				
Age > 45 years	0.99(4.68)				
Hispanic	-0.82(-1.79)				

Table 4.6 (Continued)

Baseline Utility		Baseline Constants		Satiation Parameters	
Explanatory Variables	**Coef. (t-stat)**	**Vehicle type/Vintage**	**Coef. (t-stat)**	**Vehicle Type/Vintage**	**Coef. (t-stat)**
Pickup Truck 0 to 5 years		**Compact 0 to 5 years**	-4.39(-64.27)	**Unspent Mileage**	8.28(233.66)
Income 25k to 75 k	0.19(2.30)	**Compact 6 to 11 years**	-4.55(-74.52)	**Compact 0 to 5 years**	8.81(133.93)
Male	0.19(2.45)	**Compact 12 years or older**	-4.84(-59.18)	**Compact 6 to 11 years**	8.60(125.49)
Black	-0.17(-1.34)	**Subcompact 0 to 5 years**	-5.39(-56.54)	**Compact 12 years or older**	8.61(90.42)
Asian	-0.47(-1.49)	**Subcompact 6 to 11 years**	-5.03(-63.40)	**Subcompact 0 to 5 years**	9.10(84.51)
Rural	0.24(2.81)	**Subcompact 12 years or older**	-4.87(-58.33)	**Subcompact 6 to 11 years**	9.12(77.49)
Employment density	-0.01(-2.49)	**Large 0 to 5 years**	-4.43(-66.11)	**Subcompact 12 years or older**	8.72(77.34)
Residential density	-0.02(-1.98)	**Large 6 to 11 years**	-5.20(-23.87)	**Large 0 to 5 years**	8.98(113.6)
Pickup Truck 6 to 11 years		**Large 12 years or older**	-4.69(-41.32)	**Large 6 to 11 years**	8.91(106.17)
Income > 75k	-0.26(-6.74)	**Mid-size 0 to 5 years**	-3.98(-44.01)	**Large 12 years or older**	8.83(81.33)
Male	0.18(2.26)	**Mid-size 6 to 11 years**	-4.00(-43.95)	**Mid-size 0 to 5 years**	8.59(157.41)
Black	-0.17(-1.34)	**Mid-size 12 years or older**	-4.78(-31.65)	**Mid-size 6 to 11 years**	8.48(150.43)
Asian	-0.47(-1.49)	**Two-seater 0 to 5 years**	-6.74(-35.31)	**Mid-size 12 years or older**	8.47(91.62)
Rural	0.20(2.29)	**Two-seater 6 to 11 years**	-6.26(-47.54)	**Two-seater 0 to 5 years**	8.92(50.28)
Employment density	-0.01(-3.12)	**Two-seater 12 years or older**	-6.30(-47.59)	**Two-seater 6 to 11 years**	8.94(48.93)
Residential density	-0.03(-2.73)	**Van 0 to 5 years**	-4.48(-65.80)	**Two-seater 12 years or older**	8.87(45.82)
Pickup Truck 12 years or older		**Van 6 to 11 years**	-5.26(-13.46)	**Van 0 to 5 years**	9.40(108.29)
Income 50k to 75k	-0.33(-5.23)	**Van 12 years or older**	-6.39(-22.51)	**Van 6 to 11 years**	8.99(107.2)
Income > 75k	-0.69(-11.55)	**SUV 0 to 5 years**	-4.47(-41.09)	**Van 12 years or older**	8.98(66.25)
Male	0.19(1.94)	**SUV 6 to 11 years**	-4.27(-82.12)	**SUV 0 to 5 years**	9.38(43.59)
Black	-0.17(-1.34)	**SUV 12 years or older**	-5.25(-37.65)	*Income 25k to 75 k	-1.22(-5.37)
Asian	-0.47(-1.49)	**Pickup Truck 0 to 5 years**	-4.56(-50.42)	*Income > 75k	-0.81(-3.52)
Rural	0.34(3.43)	**Pickup Truck 6 to 11 years**	-4.40(-50.83)	**SUV 6 to 11 years**	8.62(145.48)
Employment density	-0.01(-3.12)	**Pickup Truck 12 years or older**	-4.54(-45.46)	**SUV 12 years or older**	8.88(80.93)
Residential density	-0.03(-2.73)	**Motorcycle 0 to 5 years**	-5.78(-48.20)	**Pickup Truck 0 to 5 years**	8.89(135.91)
Motorcycle 0 to 5 years		**Motorcycle 6 to 11 years**	-5.88(-46.89)	**Pickup Truck 6 to 11 years**	8.76(132.77)
Male	0.37(3.23)	**Motorcycle 12 years or older**	-5.92(-44.28)	**Pickup Truck 12 years or older**	8.53(106.28)
Age 31 to 45 years old	0.34(2.66)			**Motorcycle 0 to 5 years**	7.98(57.75)
Motorcycle 6 to 11 years				**Motorcycle 6 to 11 years**	7.55(49.74)
Income > 75k	-0.26(-6.74)			**Motorcycle 12 years or older**	6.92(39.82)
Male	0.37(3.23)				
Age 31 to 45 years old	0.34(2.66)			**Log-likelihood at constants**	-190968.65
Motorcycle 12 years or older				**Log-likelihood at convergence**	-160054.15
Income 50k to 75k	-0.33(-5.23)			**Number of parameters estimated**	113
Income > 75k	-0.69(-11.55)			**Observations**	8500
Male	0.37(3.23)				
Age 31 to 45 years old	0.34(2.66)				

*Explanatory variables for SUVs of 0 to 5 years.

68

Table 4.7 Observed and Predicted Vehicle Type/Vintage Holding Using Validation Data

Vehicle Type/Vintage	Observed	Log-Linear Regression	Stochastic Frontier	AME + Non-Motor	Budget = 119505 miles
Unspent Mileage	-	-	90.9%[7]	100.0%	100.0%
Compact 0 to 5 years	11.4%	11.8%	9.4%	10.9%	13.2%
Compact 6 to 11 years	9.1%	9.1%	9.8%	10.9%	14.0%
Compact 12 years or older	4.6%	5.1%	3.7%	4.3%	5.9%
Subcompact 0 to 5 years	3.1%	3.8%	3.4%	4.2%	4.8%
Subcompact 6 to 11 years	3.2%	7.1%	2.2%	2.8%	3.5%
Subcompact 12 years or older	3.2%	3.1%	2.2%	2.8%	3.5%
Large 0 to 5 years	7.6%	8.0%	6.0%	7.2%	9.4%
Large 6 to 11 years	5.9%	5.6%	6.2%	7.6%	10.5%
Large 12 years or older	3.8%	4.1%	4.2%	5.2%	7.2%
Mid-size 0 to 5 years	15.0%	15.6%	12.9%	14.6%	17.6%
Mid-size 6 to 11 years	15.4%	15.2%	12.7%	14.4%	18.3%
Mid-size 12 years or older	5.5%	5.7%	5.0%	5.8%	7.8%
Two-seater 0 to 5 years	1.3%	1.3%	0.9%	1.2%	1.3%
Two-seater 6 to 11 years	0.8%	0.9%	0.7%	0.9%	1.1%
Two-seater 12 years or older	1.1%	1.0%	1.1%	1.1%	1.6%
Van 0 to 5 years	6.6%	7.0%	8.1%	9.6%	12.2%
Van 6 to 11 years	5.2%	5.0%	4.4%	5.3%	6.4%
Van 12 years or older	2.5%	2.5%	1.8%	2.3%	3.1%
SUV 0 to 5 years	18.3%	18.8%	16.2%	17.6%	19.6%
SUV 6 to 11 years	12.3%	12.2%	11.5%	13.0%	15.4%
SUV 12 years or older	4.2%	3.9%	2.8%	3.5%	4.3%
Pickup Truck 0 to 5 years	10.6%	10.6%	9.0%	10.2%	11.7%
Pickup Truck 6 to 11 years	11.6%	11.9%	10.6%	12.5%	14.6%
Pickup Truck 12 years or older	6.6%	6.8%	6.0%	7.1%	9.0%
Motorcycle 0 to 5 years	1.6%	2.0%	1.2%	1.6%	2.0%
Motorcycle 6 to 11 years	1.3%	1.2%	1.0%	1.3%	1.6%
Motorcycle 12 years or older	1.0%	1.2%	0.8%	1.0%	1.4%
Mean Absolute Error		0.4	0.9	0.6	1.8

---

[7] In the stochastic frontier model, for some households the mileage frontier is estimated to be less than the observed mileage expenditure. In those cases, the observed annual mileage expenditure is used as budget. Therefore, those households do not have an unspent mileage.
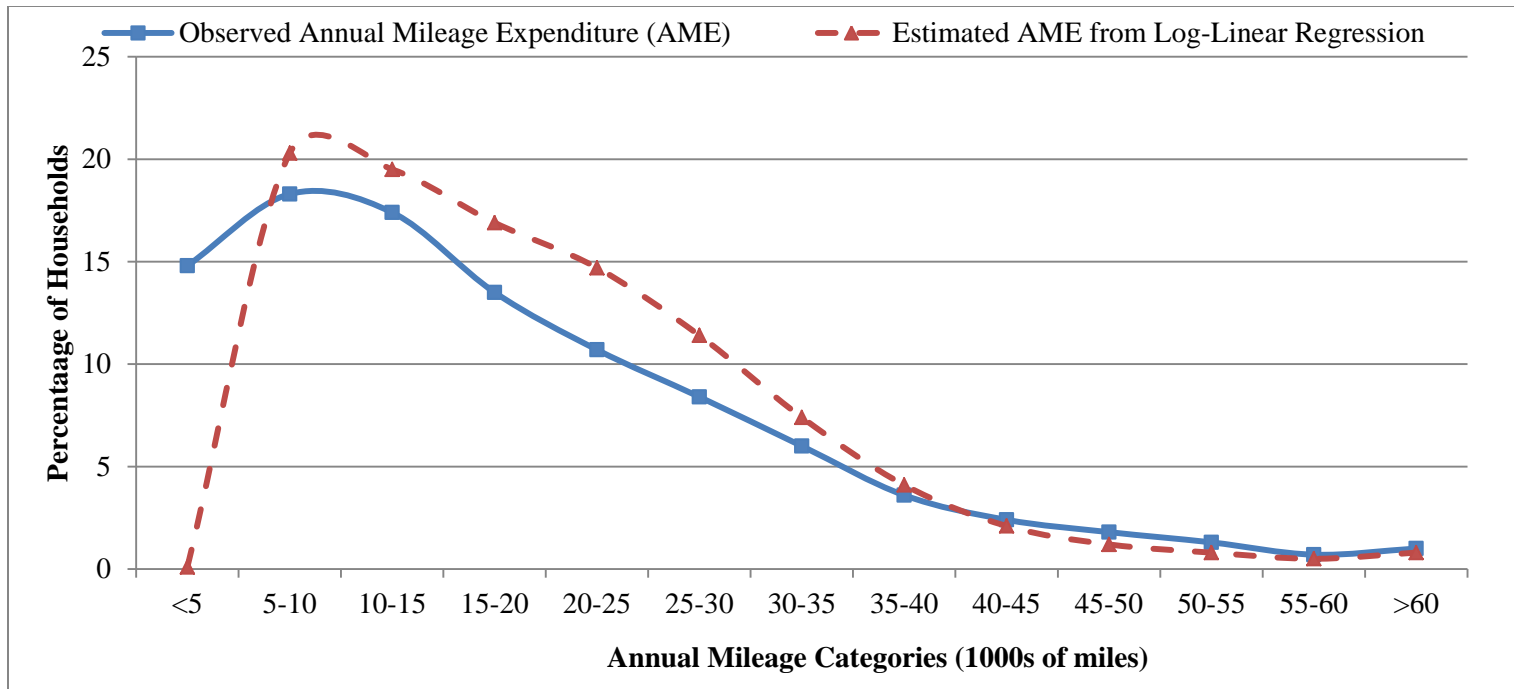
69

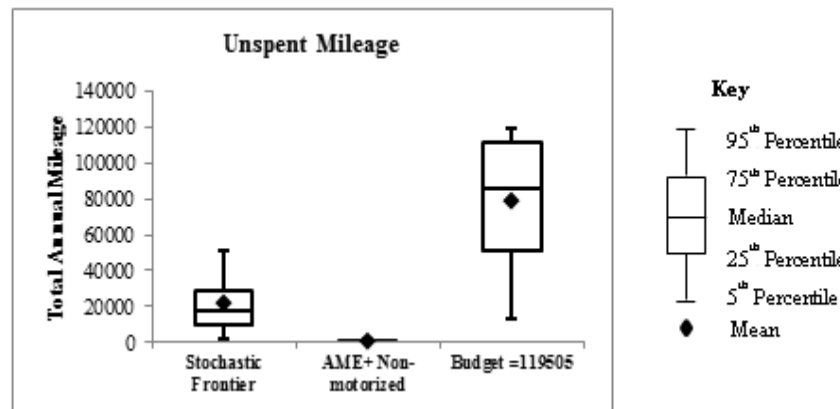**Figure 4.1 Distributions of Observed and Expected Budget from Log-Linear Regression**

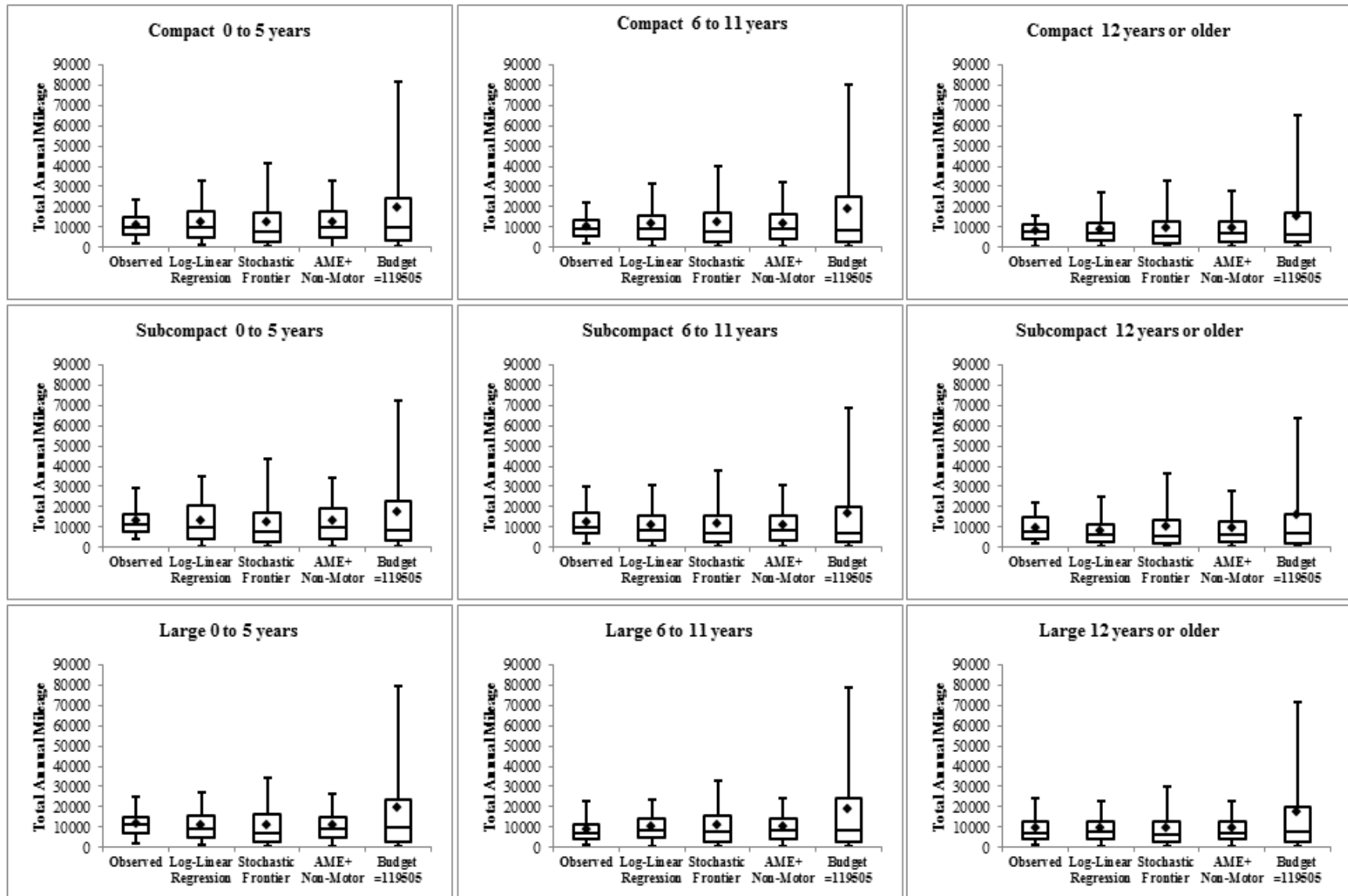**Figure 4.2 Observed and Predicted Distributions of Total Annual Mileage by Vehicle Type/Vintage Using Validation Data**
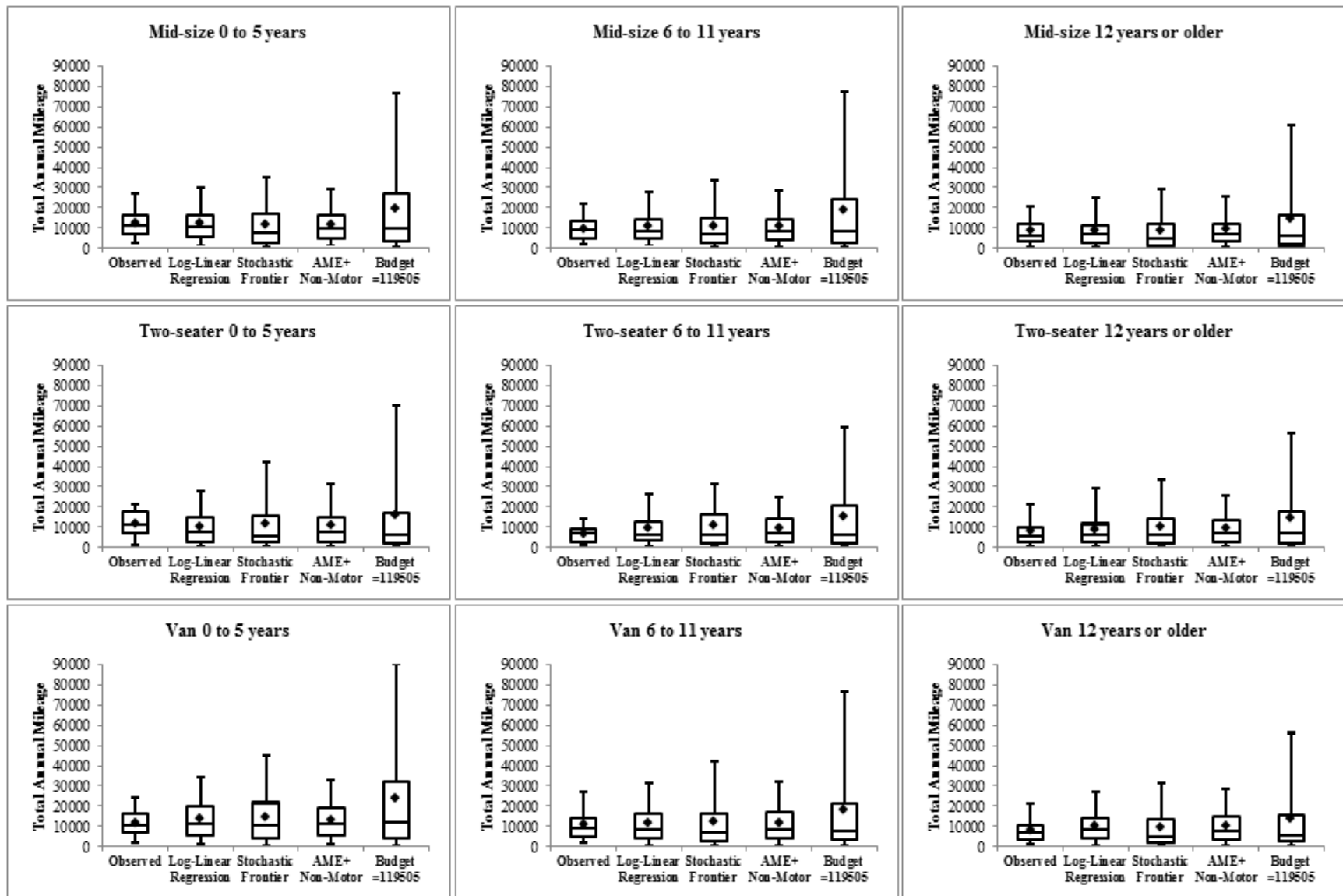
70

Figure 4.2 (Continued)
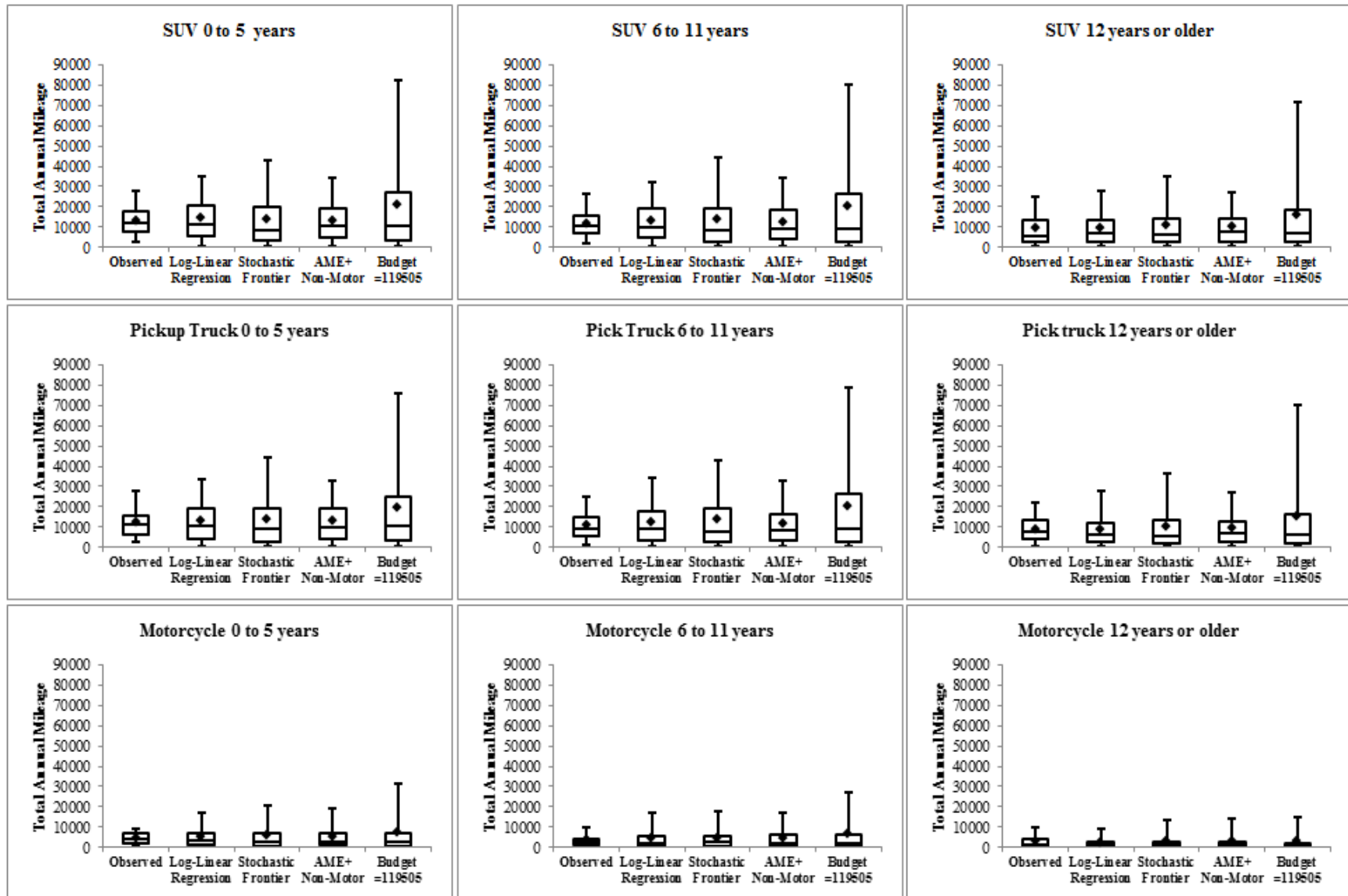
71

Figure 4.2 (Continued)

Figure 4.2 (Continued)

73

Table 4.8 Impact of Increasing Fuel Economy for New (0-5 years) Compact, Subcompact, Large and Mid-size Vehicles

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

*These numbers indicate the average change in the mileage allocated for households that a change in the mileage allocation occurred to this vehicle type/vintage.

74

www.manaraa.com

CHAPTER 5 CONCLUSION

5.1 Summary and Conclusions

This thesis compares different approaches to estimating budgets for Kuhn-Tucker (KT) demand systems, more specifically for the multiple discrete-continuous extreme value (MDCEV) model. The approaches tested include: (1) The log-linear regression approach (2) The stochastic frontier regression approach, and (3) arbitrarily assumed budgets that are not necessarily modeled as functions of socio-demographic characteristics of decision makers and choice-environment characteristics.

The log-linear regression approach has been used in the literature to model the observed total expenditure as way of estimating budgets for the MDCEV models. This approach allows the total expenditure to depend on the characteristics of the choice-maker and the choice environment. However, this approach does not offer an easy way to allow the total expenditure to change due to changes in choice alternative-specific attributes, but only allows a reallocation of the observed total expenditure among the different choice alternatives. To address this issue, we propose the stochastic frontier regression approach when the underlying budgets driving a choice situation are unobserved, but only the expenditures on the choice alternatives of interest are observed. The approach is based on the notion that consumers operate under latent budgets that can be conceived (and modeled using stochastic frontier regression) as the maximum possible expenditure they are willing to incur. The estimated stochastic frontier, or the subjective limit, or the maximum amount of expenditure consumers are willing to allocate can

75

be used as the budget in the MDCEV model. Since the frontier is by design larger than the observed total expenditure, the MDCEV model needs to include an outside alternative along with all the choice alternatives of interest to the analyst. The outside alternative represents the difference between the frontier (i.e., the budget) and the total expenditure on the choice alternatives of interest. The presence of this outside alternative helps in allowing for the total expenditure on the inside alternatives to increase or decrease due to changes in decision-maker characteristics, choice environment attributes, and more importantly the choice alternative attributes. The other assumptions used for the budgets also follow the same logic as the stochastic frontier except that their budgets are not estimated as function of socio-demographics or built environment.

To compare the efficacy of the above-mentioned approaches, we performed two empirical assessments: (1) The analysis of out-of-home activity participation and time-use (with a budget on the total time available for out-of-home activities) for a sample of non-working adults in Florida, and (2) The analysis of household vehicle type/vintage holdings and usage (with a budget on the total annual mileage) for a sample of households in Florida. A comparison of the MDCEV model predictions (based on budgets from the above mentioned approaches) to the observed discrete-continuous distributions in the data suggests that the log-linear regression approach and the stochastic frontier approach performed better than using arbitrarily assumed budgets. This is because both approaches consider heterogeneity in budgets due to socio-demographics and other explanatory factors rather than arbitrarily imposing uniform budgets on all consumers. Between the log-linear regression and the stochastic frontier regression approaches, the log-linear regression approach resulted in relatively better predictions from the MDCEV model. However, policy simulations suggest that the stochastic frontier approach

76

allows the total expenditures to either increase or decrease as a result of changes in alternative-specific attributes. While the log-linear regression approach allows the total expenditures to change as a result of changes in relevant socio-demographic and choice-environment characteristics, it does not allow the total expenditures to change as a result of changes in alternative-specific attributes. This is an important advantage of the stochastic frontier approach over the traditional log-linear regression approach to estimating budgets for the MDCEV model.

5.2 Future Research

Based on the findings from this thesis, there are at least a couple of avenues for further research, as discussed below.

5.2.1 Heteroskedastic Extreme Value Distribution of the Random Utility Components in MDC Models

Based on the comparison of the predictive assessments of households' vehicle type/vintage holdings and usage in chapter 4 , the results suggested that the MDCEV models using budgets from the stochastic frontier and log-linear regression approaches performed well in predicting the aggregate-level discrete choices observed in the validation data (i.e., the percentage of holding for each vehicle type/vintage). However, for the aggregate allocation of annual mileage expenditures, the MDCEV models using budgets from the log-linear regression approach performs relatively better than the MDCEV models using budgets from the stochastic frontier approach. Specifically, the MDCEV model using budgets from the stochastic frontier approach over-predicts the annual mileage expenditures. It is possible that this problem in prediction is due to the fat right tail of the extreme value distributions assumed in the MDCEV model. This can be rectified to a considerable extent by using heteroskedastic extreme value distributions in the model structure. Specifically, one can use the multiple discrete-continuous

77

heteroskedastic extreme value (MDHCEV) model proposed by Sikder and Pinjari (2014) to recognize the differences in the variation of unobserved influences on the preferences for different vehicle types/vintages[8].

The MDCHEV model, when used in conjunction with the budgets from the stochastic frontier approach can address the issue of over-prediction in the allocation of annual mileage expenditures to different vehicle types. To test this hypothesis, we estimated the MDCHEV model for the household vehicle holdings and utilization data discussed in Chapter 4. In the MDCHEV model, we also estimate one scale parameter for all vehicle types/vintages (*i.e.*, inside goods) and fixed the scale parameter for the unspent mileage (*i.e.*, outside good) to 1. The estimated scale parameter for all vehicle types/vintages was 0.70 suggesting that the outside good's utility function has higher variance than that of the inside goods.

Using the MDCHEV model, we predicted the annual mileage expenditure for each vehicle type/vintage. The distributions of the predicted annual mileage expenditures are plotted in Figure A.2 using the validation data. When comparing the results of the MDCEV model and the MDCHEV models using stochastic frontier as budgets, it clearly shows a reduction in the over-prediction of annual mileages for different vehicle types/vintages. By doing so, the predictions from the MDCHEV model (with stochastic frontier estimated budgets) are closer to those of the MDCEV model (with log-linear estimated budgets).

These preliminary results demonstrate the value of using a heteroskedastic extreme value distribution for the random utility components in MDC choice models. Of course, additional empirical testing is needed in the context of different geographical contexts and different empirical applications before reaching conclusions on this.

---

[8] For the structure of the MDCHEV model, please refer to Sikder and Pinjari (2014).

5.2.2 Other Future Research

1. In this study, the regression models for budgets (i.e., the stochastic frontier regression model and the log-linear regression model) were estimated separately from the corresponding MDCEV models. In future research, it will be useful to integrate the budget regression model equations with the MDCEV models into an integrated model system using latent variable modeling approaches. That way, the budget estimation would be endogenous to the MDCEV model.

2. While the current empirical applications are in the context of time-use and mileage-use in Florida, it will be useful to test the performance of different approaches (to estimate budgets) for other empirical applications and other geographical contexts involving MDC choices, including long-distance vacation time and money budgets, and market basket analysis.

79

REFERENCES

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

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

Augustin, B., A.R. Pinjari, V. Sivaraman, A. Faghih Imani, N. Eluru and R. Pendyala (2014). Stochastic Frontier Estimation of Budgets for Kuhn-Tucker Demand Systems: Application for Time-use Analysis. In review, *Transportation Research Part A*.

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

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

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

Bhat, C.R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B*, 42(3), 274-303.

Bhat, C. R. and S. Sen, (2006). Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. *Transportation Research Part B*, 40(1), 35-53.

Bhat, C.R., S. Sen, and N. Eluru (2009). The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B*, 43(1), 1-18.

Bhat, C.R., K.G. Goulias, R.M. Pendyala, R. Paleti, R. Sidharthan, L. Schmitt, and H-H. Hu (2013a). A Household-Level Activity Pattern Generation Model with an Application for Southern California. *Transportation*, 40(5), 1063-1086.

Bhat, C.R., M., Castro and A. R. Pinjari (2013b). Allowing for non-additively separable and flexible utility forms in multiple discrete-continuous models. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

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

Castro, M., C.R. Bhat, R.M. Pendyala, and S. Jara-Diaz (2012). Accommodating multiple constraints in the multiple discrete–continuous extreme value (MDCEV) choice model. *Transportation Research Part B*, 46(6), 729-743.

Choo, S. and P. L. Mokhtarian, (2004). What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A*, 38(3), 201-222.

Chikaraishi, M., J. Zhang, A. Fujiwara and K.W Axhausen (2010). Exploring variation properties of time use behavior based on a multilevel multiple discrete-continuous extreme value model, *Transportation Research Record*, No. 2156, pp. 101-110.

Chintagunta, P.K., and H. Nair (2011). Marketing Models of Consumer Demand. *Marketing Science*, 30(6), 977-996.

Deaton, A., and J. Muellbauer (1980). Economics and Consumer Behavior. Cambridge University Press, Cambridge.

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

Farooq, B., E. J. Miller, and M. A. Haider (2013) Multidimensional Decisions Modelling Framework for Built Space Supply. *Journal of Transport and Land Use*, 6(3), 61-74.

Habib, K.M.N. and E.J. Miller (2008). Modeling daily activity program generation considering within-day and day-to-day dynamics in activity-travel behaviour. *Transportation*, 35(4), 467-484.

Hanemann, M.W. (1978). A methodological and empirical study of the recreation benefits from water quality improvement. Ph.D. dissertation, Department of Economics, Harvard University.

Hensher, D.A., P.O. Barnard, N.C. Smith and F.W. Milthorpe, (1992). Dimensions of Automobile Demand. A Longitudinal Study of Automobile Ownership and Use.North-Holland, Amsterdam.

Hocherman, I., J.N. Prashker and M. Ben-Akiva, (1983). Estimation and use of dynamic transaction models of automobile ownership. *Transportation Research Record* 944, 134–141.

81

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

Kaza, N., C. Towe and X. Ye. (2012). A hybrid land conversion model incorporating multiple end uses. *Agricultural and Resource Economics Review* 40(3), 341-359.

Kim, J., G.M. Allenby, and P.E. Rossi (2002). Modeling consumer demand for variety. *Marketing Science*, 21(3), 229-250.

Kitamura, R., T.F. Golob, T. Yamamoto and G. Wu, (2000). Accessibility and auto use in a motorized metropolis. TRB ID Number 00-2273, *Paper presented at the 79th Transportation Research Board Annual Meeting*, Washington, DC.

Kitamura, R., T. Yamamoto, K. Kishizawa, and R.M. Pendyala (2000). Stochastic Frontier Models of Prism Vertices. In *Transportation Research Record*, No. 1718, pp. 18-26.

Kumbhakar, S. and C.A.K. Lovell (2000). *Stochastic Frontier Analysis.* Cambridge University Press, Cambridge, UK.

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

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

Mannering, F., C. Winston and W. Starkey,(2002). An exploratory analysis of automobile leasing by US households. *Journal of Urban Economics* 52 (1), 154–176.

Manski, C. F. and L. Sherman, (1980). An empirical analysis of household choice among motor vehicles. *Transportation Research Part A*, 14(6), 349-366.

Mohammadian, A. and E.J. Miller, (2003). An empirical investigation of household vehicle type choice decisions. *Forthcoming, Transportation Research*.

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

Pinjari, A.R., and C.R. Bhat (2011). computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: application to residential energy consumption analysis. Technical paper, Department of Civil & Environmental Engineering, University of South Florida.

Pinjari, A.R. (2011). Generalized Extreme Value (GEV)-Based Error Structures for Multiple Discrete-Continuous Choice Models. *Transportation Research Part B*, 45(3), 474-489.

Pinjari, A.R., B. Augustin, A. Faghih Imani, V. Sivaraman, N. Eluru and R. Pendyala (2014). Stochastic Frontier Estimation of Budgets for Kuhn-Tucker Demand Systems: Application for Time-use Analysis. *Proceedings of the Transportation Research Board (TRB) Annual Meeting*, Washington D.C., January 2014.

Pinjari, A.R., C. R. Bhat and D. A. Hensher (2009). Residential self-selection effects in an activity time-use behavior model. *Transportation Research Part B*, 43(7), 729-748.

Pucher, J. and J. L. Renne, (2003). Socioeconomics of urban travel: evidence from 2001 NHTS. *Transportation Quarterly,* 57(3), 49-77.

Sikder, S., and A.R. Pinjari* (2014). The Benefits of Allowing Heteroscedastic Stochastic Distributions in Multiple Discrete-Continuous Choice Models. Forthcoming*, Journal of Choice Modelling.*

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

Van Nostrand, C., V. Sivaraman., and A.R. Pinjari (2013). Analysis of long-distance vacation travel demand in the United States: a multiple discrete-continuous choice framework. *Transportation*, 40(1), 151-171.

von Haefen, R.H. (2010). Incomplete demand systems, corner solutions, and welfare measurement. *Agricultural and Resource Economics Review* 39(1), 22-36.

von Haefen, R.H. and D.J. Phaneuf (2005), Kuhn-Tucker demand system approaches to nonmarket valuation, in R. Scarpa and A. Alberini (eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, Dordrecht, The Netherlands: Springer, pp. 135-158.

Wales, T.J., and A.D. Woodland (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics*, 21(3), 263-285.

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

APPENDICES

Appendix A: Additional Tables

Table A.1 Log-Linear Regression for Total Annual Mileage Expenditure (AME)

Variables	Coefficients	t-stats
Constant	10.77	13.29
Head Household Characteristics		
Male	0.08	4.36
Age 18 to 29 (age 55 to 74 is base)	0.28	4.46
Age 30 to 54 (age 55 to 74 is base)	0.18	7.94
Age >75 (age 55 to 74 is base)	-0.25	-11.01
Household Characteristics		
Income < 25k/year (Income 25k to 50k is base)	-0.09	-3.86
Income >=50 and < 75 (Income 25 k to 50k is base)	0.15	5.90
High Income >=75k/year (Income 25k to 50k is base)	0.24	10.30
Number of drivers	0.31	15.77
Number of workers	0.18	13.11
Presence of children	0.05	1.78
2+ household members	0.25	9.29
Fuel Cost ($/gallon)		
Fuel Cost	-0.76	-2.86
Household Location Attributes		
Rural(Urban is base)	0.17	7.60
Employment Density	-0.003	-3.96
Residential Density	-0.008	-2.20
$\hat{\sigma}_v$	0.773	
R-squared	0.354	
Adjusted R-squared	0.353	
Number of observations	8500	

Table A.2 Observed and Predicted Vehicle Type/Vintage Holding Using Estimation Data

| Vehicle Type/Vintage	Observed	Log-Linear Regression	Stochastic Frontier	AME + Non-Motorized	Budget = 119505 miles
Unspent Mileage	-	-	90.2%	100.0%	100.0%
Compact 0 to 5 years	10.4%	11.2%	9.5%	10.4%	9.6%
Compact 6 to 11 years	9.4%	9.1%	9.7%	10.4%	9.7%
Compact more than 12 years	4.6%	5.0%	3.9%	4.6%	3.9%
Subcompact 0 to 5 years	3.5%	3.9%	3.1%	3.7%	3.0%
Subcompact 6 to 11 years	2.9%	2.8%	2.3%	2.8%	2.2%
Subcompact more than 12 years	3.0%	3.2%	3.9%	4.7%	3.9%
Large 0 to 5 years	7.3%	8.0%	6.5%	7.2%	6.4%
Large 6 to 11 years	6.7%	7.6%	6.0%	7.0%	6.0%
Large more than 12 years	4.0%	4.5%	3.3%	4.0%	3.2%
Mid-size 0 to 5 years	15.3%	16.2%	15.3%	16.5%	15.3%
Mid-size 6 to 11 years	14.4%	12.9%	13.2%	14.3%	13.2%
Mid-size more than 12 years	4.9%	5.5%	4.2%	5.0%	4.2%
Two-seater 0 to 5 years	1.2%	1.2%	1.3%	1.5%	1.2%
Two-seater 6 to 11 years	1.1%	0.9%	1.1%	1.4%	1.1%
Two-seater more than 12 years	1.1%	1.2%	1.0%	1.0%	1.0%
Van 0 to 5 years	6.1%	6.6%	5.4%	6.3%	5.2%
Van 6 to 11 years	6.1%	12.6%	5.1%	6.0%	5.0%
Van more than 12 years	2.3%	2.5%	2.4%	2.9%	2.3%
SUV 0 to 5 years	17.8%	18.6%	16.8%	17.5%	16.5%
SUV 6 to 11 years	12.6%	11.6%	13.7%	14.7%	13.6%
SUV more than 12 years	3.3%	3.5%	2.7%	3.3%	2.5%
Pickup Truck 0 to 5 years	10.0%	10.4%	10.4%	11.8%	10.3%
Pickup Truck 6 to 11 years	9.6%	9.2%	8.5%	9.6%	8.1%
Pickup Truck more than 12 years	6.4%	6.5%	5.3%	6.2%	5.1%
Motorcycle 0 to 5 years	1.8%	1.9%	1.4%	1.7%	1.3%
Motorcycle 6 to 11 years	1.5%	1.3%	1.4%	1.9%	1.4%
Motorcycle more than 12 years	1.2%	1.2%	1.3%	1.7%	1.2%
Mean Absolute Error		0.7	0.6	0.4	0.6
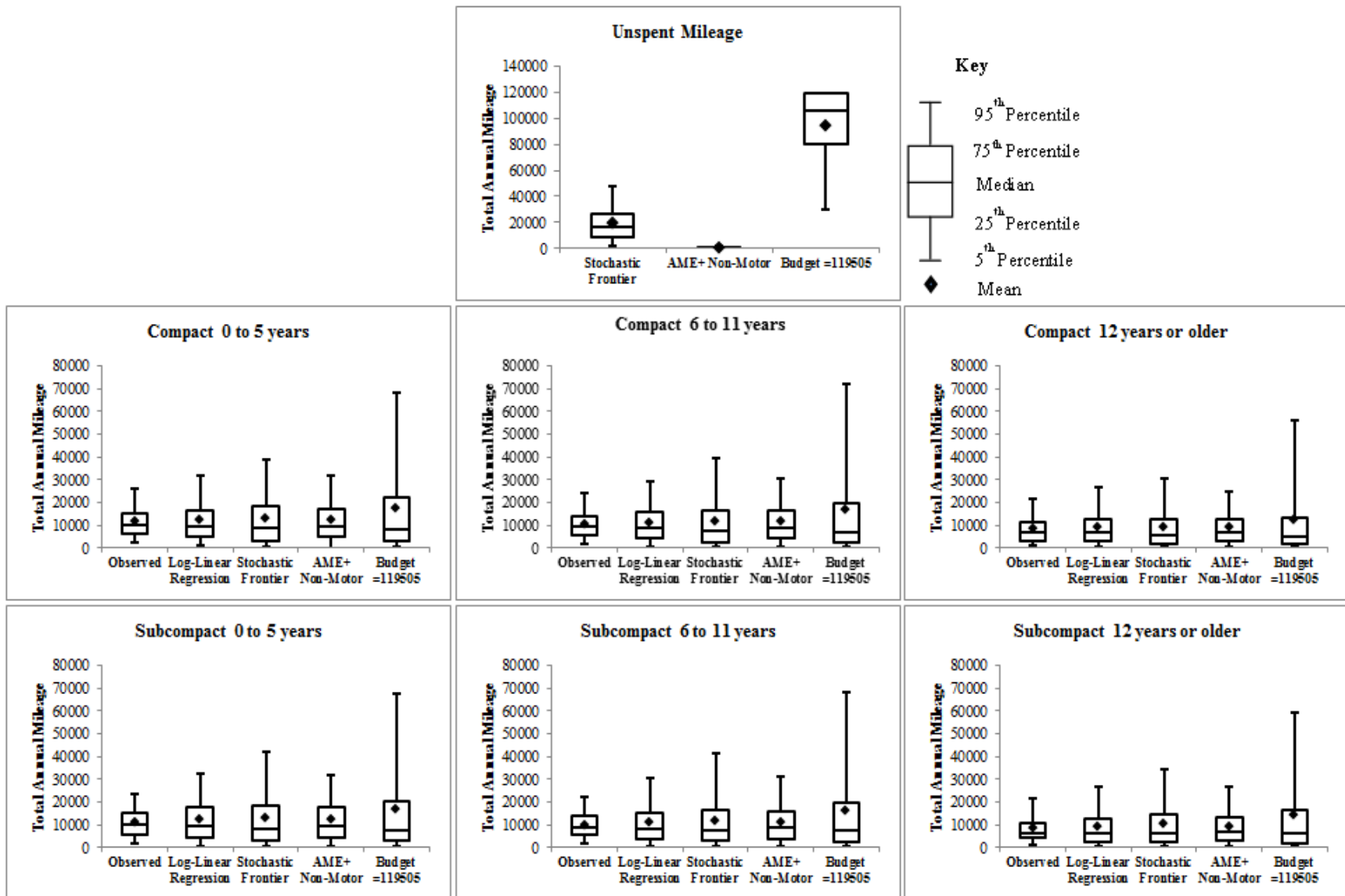
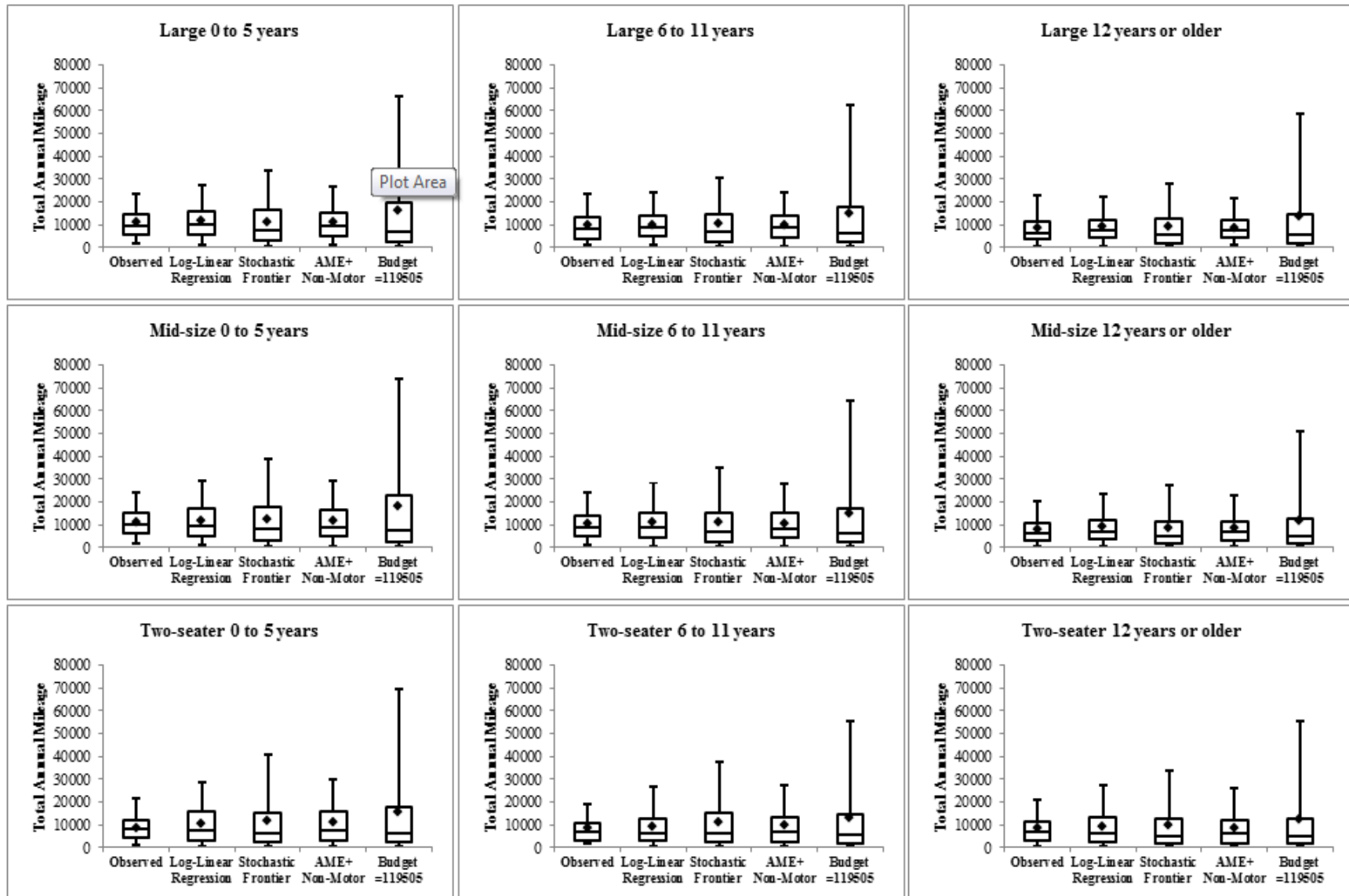**Figure A.1 Observed and Predicted Distributions of Total Annual Mileage by Vehicle Type/Vintage Using Estimation Data**
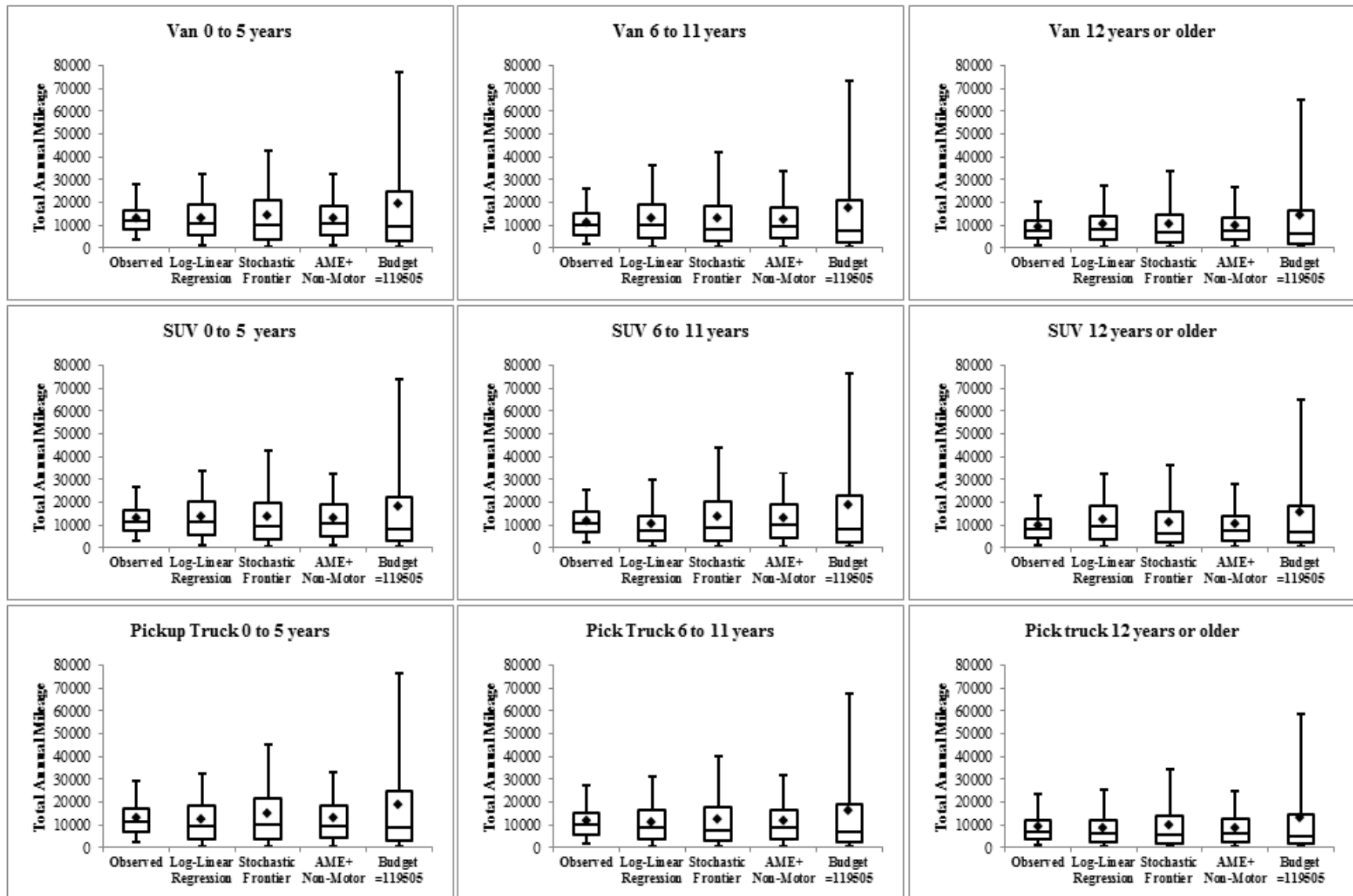
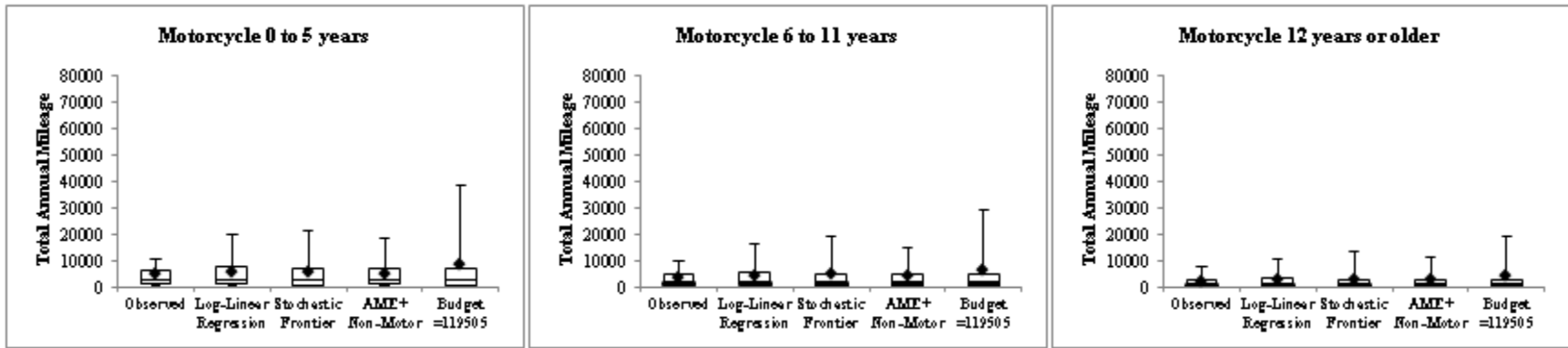**Figure A.1 (Continued)**
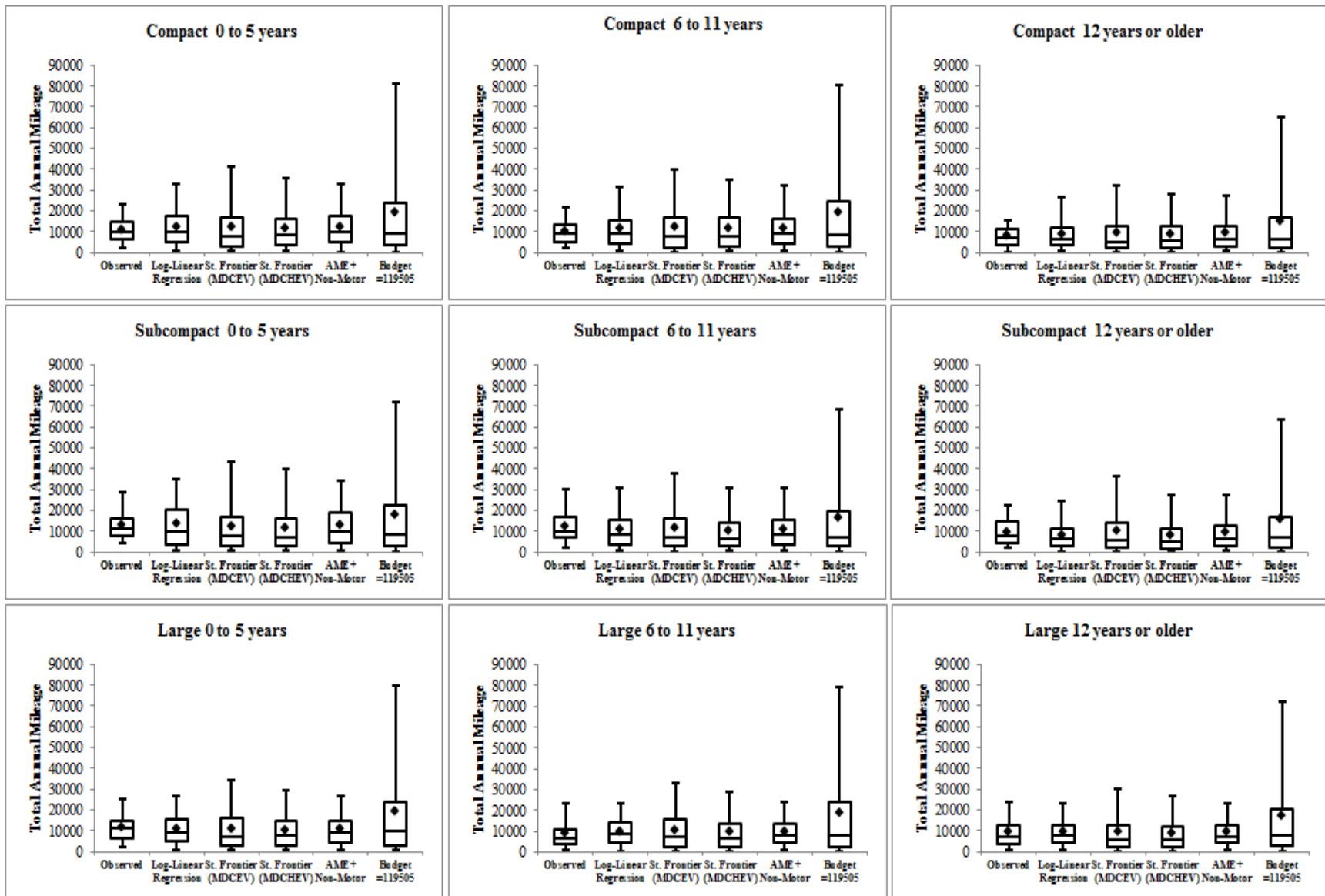
Figure A.1 (Continued)

89

Figure A.1 (Continued)

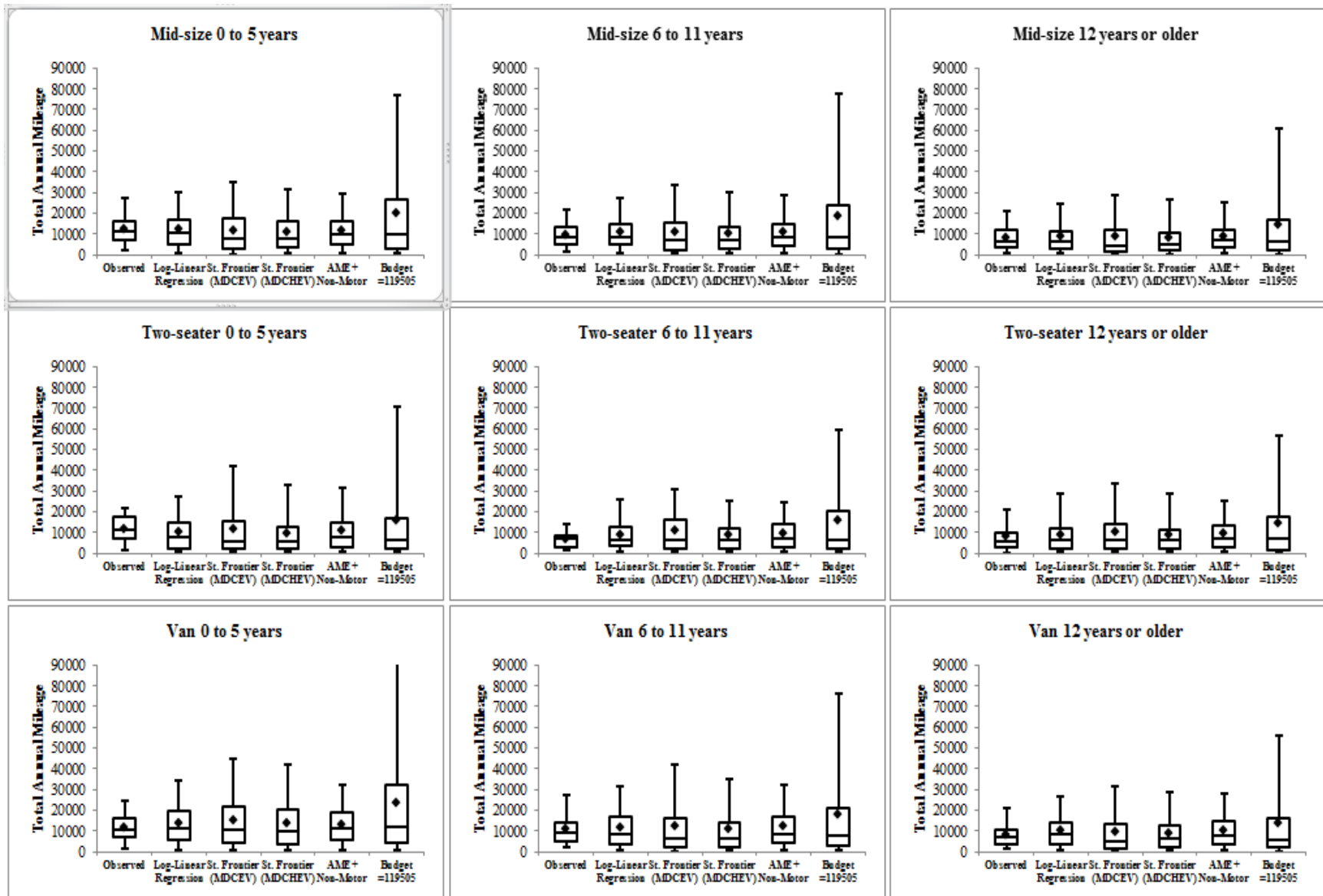**Figure A.2 Observed and Predicted Distributions of Total Annual Mileage by Vehicle Type/Vintage to MDCHEV Model**
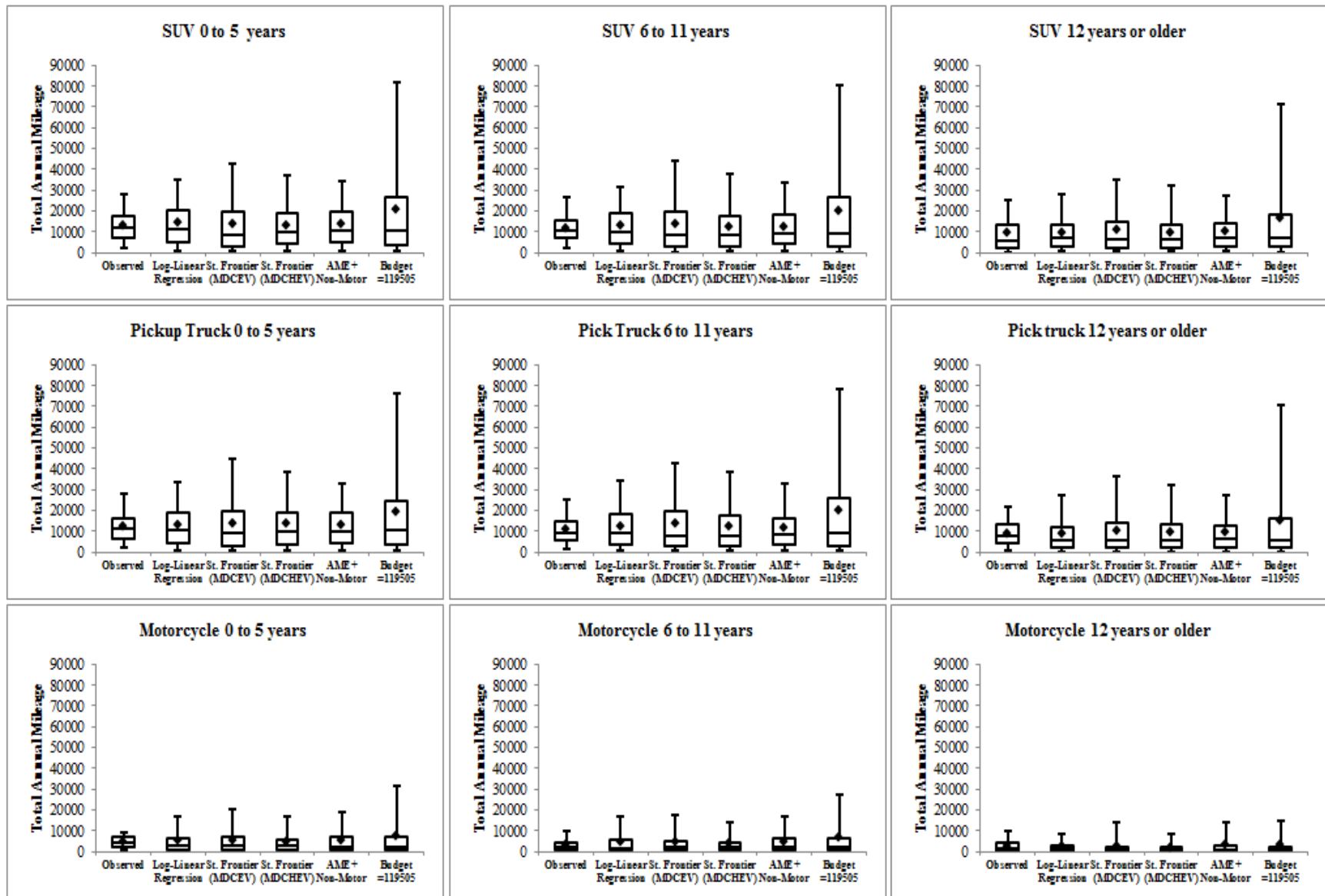
91

Figure A.2 (Continued)

92

Figure A.2 (Continued)

93

Appendix B: Copyright Permission for Chapter 1, 2 and 3

ELSEVIER

Type here to search on Elsevier.com   Advanced search   Follow us: Help & Contact

Journals & books   Online tools   Authors, editors & reviewers   About Elsevier   Community   Store

For Authors

Journal authors' home

Author Rights

Ethics

Agreements

Open access

Author services

Journal performance

Early career researchers

Authors' Update

Book authors' home

Getting your paper noticed

Author Rights

Elsevier supports the need for authors to share, disseminate and maximize the impact of their research. We take our responsibility as stewards of the online record seriously, and work to ensure our policies and procedures help to protect the integrity of scholarly works.

Author's rights to reuse and post their own articles published by Elsevier are defined by Elsevier's copyright policy. For our proprietary titles, the type of copyright agreement used depends on the author's choice of publication:

For subscription articles: These rights are determined by a copyright transfer, where authors retain scholarly rights to post and use their articles.

For open access articles: These rights are determined by an exclusive license agreement, which applies to all our open access content.

In both cases, the fundamental rights needed to publish and distribute an article remain the same and Elsevier authors will be able to use their articles for a wide range of scholarly purposes.

Details on how authors can reuse and post their own articles are provided below.

Help and support
For reuse and posting not detailed below, please see our posting policy, or for authors who would like to:

- Include material from other sources in your work being published by Elsevier, please visit: Permission seeking guidelines for Elsevier authors.
- Obtain permission to re-use material from Elsevier books, journals, databases, or other products, please visit: Obtaining permission to reuse Elsevier material
- Or if you are an Elsevier author and are contacted by a requestor who wishes to re-use all or part of your article or chapter, please also refer them to our Obtaining Permission to Re-Use Elsevier Material page.
- See our FAQ on posting and copyright queries.
- Contact us directly, please email our Permissions Help Desk.

Author Use	Author Posting	Definitions
How authors can use their own journal articles		

Authors can use their articles for a wide range of scholarly, non-commercial purposes as outlined below. These rights apply for all Elsevier authors who publish their article as either a subscription article or an open access article.

We require that all Elsevier authors always include a full acknowledgement and, if appropriate, a link to the final published version hosted on Science Direct.

For open access articles these rights are separate from how readers can reuse your article as defined by the author's choice of Creative Commons user license options.

Authors can use either their accepted author manuscript or final published article for:

✓	Use at a conference, meeting or for teaching purposes
✓	Internal training by their company
✓	Sharing individual articles with colleagues for their research use* (also known as 'scholarly sharing')
✓	Use in a subsequent compilation of the author's works
✓	Inclusion in a thesis or dissertation
✓	Reuse of portions or extracts from the article in other works

95

✔ Preparation of derivative works (other than for commercial purposes)

*Please note this excludes any systematic or organized distribution of published articles.

96