Development of a Household Vehicle Ownership and Fleet Composition Model

Final Report

2015

Maricopa Association of Governments

MARICOPA ASSOCIATION of GOVERNMENTS

Acknowledgements

This MAG report is based on project deliverables provided to MAG by ASU. MAG staff would like to specifically acknowledge Dr. Ram M. Pendyala, Dr. Venu Garikapati and Dr. Daehyun You, who provided the contribution to the project.

Table of Contents

List of Tables

List of Figures

1. Introduction

It has long been recognized that household vehicle ownership is a key determinant of travel demand, many travel demand models lack a rigorous statistical model to forecast vehicle ownership under a wide range of scenarios – even though vehicle ownership is an explicit dimension of interest in the trip generation (and possibly destination choice and mode choice) component of the four-step model. Many four step travel models employ rather simplistic trend line approaches to forecast future car ownership distributions that do not offer sensitivity to socio-economic, demographic, and built environment attributes. More recently, there has been growing recognition of the absence models in practice for forecasting the mix of vehicle types owned by households and the degree of utilization of each vehicle type. The vehicle fleet mix or composition is a critical input to emissions analysis and modeling (such as MOBILE6 and MOVES); however, rather than model the vehicle fleet composition as a function of various influencing factors, the profession has historically used default vehicle fleet mix distributions (or ad-hoc customizations of the default distributions) to estimate emissions inventories. Simple procedures that may be adopted to customize a default vehicle fleet mix for purposes of emissions analysis are provided by FHWA[1]. However, virtually all of these procedures are insensitive to policy inputs such as pricing signals, tax rebates, household and person socio-economic transitions (such as the aging of the population), and new vehicle technology availability. As a result, the ability of metropolitan planning organizations (MPOs) to accurately forecast GHG emissions attributable to personal travel under a variety of policy, socio-economic, and technology scenarios is limited in the current modeling context.

The amount of greenhouse gas (GHG) emissions and energy consumption is directly related to the type(s) of vehicle(s) that households drive and the amount that they use different vehicle types. As the state-of-the-practice of travel demand modeling is moving into an era of activity-based travel microsimulation, the profession now has the ability to simulate activity-travel patterns arising from individual person and household choice processes. Although traditional four-step travel demand models have sometimes included vehicle ownership model components, they have not included models of vehicle type choice primarily for two reasons. First, in the absence of a microsimulation framework for model development and application, it is not possible to effectively incorporate and use a household-level vehicle fleet composition model within the four-step modeling paradigm that generally uses zonal level inputs for forecasting purposes. Second, while the profession has made great strides in implementing the classic single discrete choice modeling methods (such as multinomial logit, ordered probit or logit, and nested logit) where travelers are assumed to choose a single alternative from among a set of choices, very little progress has been made in the development of multiple discrete choice models where households or individuals are able to choose multiple alternatives from a choice set. This latter situation is encountered in the vehicle fleet composition context because households may choose to own multiple vehicles of different types, and often end up doing so. Recent work by (Pinjari and Bhat, 2009), introducing the multiple discrete-continuous extreme value (MDCEV) model and its variants, has provided the ability to model this type of multiple discrete choice behavior in a computationally manageable manner. Essentially, there are six motivating factors for the development and implementation of household ownership and vehicle fleet composition models within transportation model systems:

[1] http://www.fhwa.dot.gov/environment/conformity/ emission/emismeth7.htm

- Four-step travel demand models depend heavily on accurate forecasts of vehicle ownership for trip generation estimation,
- The modeling paradigm has shifted to an activity-based microsimulation framework, thus providing the ability to utilize a disaggregate household-level vehicle ownership and fleet composition model,
- Recent data on vehicle type choices is now available through the 2008 National Household Travel Survey data set and its Maricopa Association of Governments (MAG) region add-on,
- A modeling methodology that accommodates the multiple discreteness of vehicle type choice behavior in a computationally manageable manner is now available,
- There is considerable policy interest in reducing greenhouse gas emissions and the analysis of the effect of policy on greenhouse gas emissions can be best supported with a model system that incorporates a vehicle fleet composition model, and
- The field of emissions modeling has also moved into a microsimulation domain with the new MOVES model, which is capable of utilizing very disaggregate activity-trip level information for estimating emissions. This capability permits the association of vehicle type with every trip simulated in a region.

2. Household Vehicle Ownership Model Update

The purpose of household vehicle ownership model is to estimate the number of automobiles owned by households in each zone. The household vehicle ownership model used for 4-step travel demand model was updated based on the 2008 National Household Travel Survey (NHTS) add-on dataset for the MAG modeling area. The NHTS dataset includes 4,707 households. Since 396 households in the NHTS are missing values of household income or home type, these households were removed from the model estimation. A total of 4,311 households were used in the household vehicle ownership model estimation. The obtained model is calibrated using the Public Use Microdata Sample (PUMS) dataset based on ACS 2006-2010.

2.1 Model Structure

Multinomial Logit (MNL) model structure is utilized in formulating the household vehicle ownership model. In this formulation, the number of automobiles owned by a household is modeled based on the socio-economic characteristics of the household and the land use characteristics of the Traffic Analysis Zone (TAZ) in which household resides. Following is the list of independent variables considered to estimate this model:

<u>Explanatory variables</u>

INCQTL1: Household income quintile 1
INCQTL2: Household income quintile 2
INCQTL3: Household income quintile 3
INCQTL4: Household income quintile 4
INCQTL5: Household income quintile 5

HHSIZE1: Household Size 1
HHSIZE2: Household Size 2
HHSIZE3: Household Size 3
HHSIZE4: Household Size 4+

HHWRKR0: 0 workers per household

HHWRKR1: 1 worker per household
HHWRKR2: 2 workers per household
HHWRKR3: 3+ workers per household

REMP30T: Percent of regional employment within 30 minutes total un-weighted transit time.
EMPDEN: Employment Density (Total Employment /(Total Square Miles of TAZ*100))
HHDEN: Household Density (Total Households/(Total Square Miles of TAZ*100))
Single family housing or Multi-Family housing unit

<u>Dependent Variable</u>
HHVEH0: 0 Autos per household
HHVEH1: 1 Autos per household
HHVEH2: 2 Autos per household
HHVEH3: 3 Autos per household
HHVEH4: 4+ Autos per household

2.2 Model Estimation

MNL structure is used to estimate the models. Two MNL models are estimated for (i) balancing process[2] (BP) and (ii) no balancing process[3] (NBP). The model with NBP was estimated based on the model with BP. Hence, most variables used in the MNL model for BP are included in the MNL model for NBP. Statistical modeling packages LIMDEP was used for estimating the models based on the add-on NHTS 2008 dataset in MAG region. MNL models are also estimated using R code. Since there are limitations in estimation constraints for MNL model in R, the final models were estimated in LIMDEP. Tables 2.1 and 2.2 show the results of MNL models for household vehicle ownership model estimated in LIMDEP for both BP and NBP.

[2] It is to assign the number of households by household unit type (Single and Multiple Family Housing Unit) using the seed table (4 household size groups, 5 income groups, 3 worker size groups, retirement zone flag (0/1), 2 household types, and Percent), # of workers per household by 0, 1, 2, and 3+ workers for each TAZ, and socioeconomic data.

[3] It is to take the number of households by household unit type generated by Population Synthesizer but this option is not implemented in the official model.

Table 2.1 Estimated Multinomial Logit Model with Balancing Process

Autos	Variables	Coef.	Std. Err.	t-stat
Zero	Constant	-2.491	0.374	-6.658
	INCQTL1*HHWRKR0	3.141	0.427	7.361
	INCQTL1*HHWRKR1	0.786	0.557	1.411
	INCQTL2*HHWRKR0	1.554	0.474	3.280
	INCQTL3*HHWRKR0	1.048	0.583	1.798
	INCQTL4	-0.685	0.579	-1.182
	INCQTL5	-0.344	0.816	-0.422
	INCQTL1*HHDEN	0.038	0.009	4.259
	EMPDEN	0.023	0.004	5.441
One	Constant	-0.665	0.143	-4.652
	INCQTL1	2.320	0.130	17.869
	INCQTL2	1.837	0.126	14.622
	INCQTL3*HHWRKR0	1.324	0.187	7.097
	INCQTL3*HHWRKR1	1.000		
	INCQTL4*HHWRKR0	0.255	0.257	0.990
	HHSIZE1	2.500	0.130	19.195
	HHSIZE2	0.593	0.111	5.366
	INCQTL5*HHDEN	0.060	0.013	4.660
Two	Constant	1.739	0.075	23.180
	HHWRKR2	0.904	0.147	6.139
	INCQTL4*HHDEN	-0.050	0.005	-9.188
	INCQTL5*HHDEN	0.029	0.010	2.763
	INCQTL4*HHWRKR1	1.100		
	INCQTL5*HHWRKR1	1.100		
	INCQTL1*HHWRKR2	1.000		
	INCQTL1*HHWRKR3	1.000		
Three	Constant	0.000		
	INCQTL1	-0.320		
	INCQTL2	-0.200		
	INCQTL5	1.674	0.174	9.635
	INCQTL4*HHWRKR0	-0.264	0.297	-0.889
	HHWRKR2	1.022	0.163	6.268
	HHWRKR3	2.598	0.272	9.551
	HHSIZE1	-0.644	0.231	-2.781
	REMP30T	-25.400	9.699	-2.619
Four +	Constant	0.000		
	INCQTL1	-0.340		
	INCQTL2	-0.200		
	INCQTL5	1.800	0.198	9.087
	INCQTL4*HHWRKR0	-0.614	0.530	-1.160
	HHWRKR2	1.268	0.191	6.633
	HHWRKR3	3.581	0.283	12.657
	Single Family Housing Unit	-0.687	0.155	-4.425
	HHDEN	-0.027	0.007	-3.917

Table 2.2 Estimated Multinomial Logit Model with No Balancing Process

Autos	Variables	Coef.	Std. Err.	t-stat
Zero	Constant	-3.186	0.379	-8.401
	INCQTL1*HHWRKR0	2.975	0.428	6.959
	INCQTL1*HHWRKR1	0.609	0.558	1.092
	INCQTL2*HHWRKR0	1.401	0.474	2.956
	INCQTL3*HHWRKR0	1.193	0.583	2.047
	INCQTL4	-0.285	0.579	-0.491
	INCQTL5	-0.337	0.816	-0.412
	INCQTL1*HHDEN	0.038	0.009	4.310
	EMPDEN	0.024	0.004	5.528
One	Constant	-1.131	0.148	-7.618
	INCQTL1	1.922	0.137	13.978
	INCQTL2	1.447	0.130	11.165
	INCQTL3*HHWRKR0	1.231	0.187	6.579
	INCQTL3*HHWRKR1	1.000		
	INCQTL4*HHWRKR0	0.332	0.255	1.302
	HHSIZE1	2.545	0.130	19.565
	HHSIZE2	0.597	0.111	5.399
	INCQTL5*HHDEN	0.050	0.013	3.921
Two	Constant	1.026	0.099	10.402
	HHWRKR2	0.898	0.146	6.137
	INCQTL4*HHDEN	-0.027	0.006	-4.872
	INCQTL5*HHDEN	0.031	0.010	2.999
	INCQTL4*HHWRKR1	1.100		
	INCQTL5*HHWRKR1	1.100		
	INCQTL1*HHWRKR2	1.000		
	INCQTL1*HHWRKR3	1.000		
Three	Constant	0.000		
	INCQTL1	-0.218	0.177	-1.233
	INCQTL2	-0.148	0.139	-1.065
	INCQTL5	1.093	0.179	6.092
	INCQTL4*HHWRKR0	-0.664	0.298	-2.231
	HHWRKR2	0.816	0.165	4.954
	HHWRKR3	2.393	0.274	8.73
	HHSIZE1	-0.521	0.232	-2.243
	REMP30T	-21.114	9.536	-2.214
Four +	Constant	0.000		
	INCQTL1	-0.069	0.231	-0.299
	INCQTL5	1.304	0.198	6.588
	INCQTL4*HHWRKR0	-0.941	0.53	-1.775
	HHWRKR2	1.091	0.192	5.682
	HHWRKR3	3.401	0.285	11.951
	Single Family Housing Unit	-0.834	0.162	-5.132
	HHDEN	-0.023	0.007	-3.214

2.3 Model Calibration

Since the estimated models over predicted 3 or more vehicles and under predicted less than 3 vehicles in the households, the constants in the model were adjusted to prevent overestimation/ underestimation in predicting the number of vehicles. The latest updated 2010 input data was used for this calibration of no balancing process. The table given below shows the same estimation results as the tables above, but with adjusted constant/coefficient values. In zero autos, coefficients of four interaction and two income group variables (INCQTL1*HHWRKR0, INCQTL1*HHWRKR1, INCQTL2*HHWRKR0, INCQTL3*HHWRKR0, INCQTL4, and INCQTL5) were adjusted instead of constant. This is done because zero household vehicle ownerships were not being predicted in the rural areas which are purported to be highly auto dependent. The household vehicle ownership characteristics of people in the lower income quintiles in these regions will be different from household vehicle ownership characteristics of a similar household in an urban area with better transportation facilities. To account for this variation, the coefficients of aforementioned variables were adjusted. In addition, two coefficients of high income quintile group were adjusted because the households in these groups prefer to have at least one auto. This prevents overestimation for zero autos in the households with high income.

Table 2.3 shows the calibrated Multinomial Logit Model with balancing process with adjusted constants and coefficients based on the 2008 socio-economic (2012 projection) and skim data.

Table 2.4 shows the calibrated Multinomial Logit Model with no balancing process with adjusted constants and coefficients based on the 2010 socio-economic (2012 projection) and skim data.

Table 2.5 shows the calibrated Old Multinomial Logit Model with no balancing process with adjusted constants and coefficients based on the 2010 socio-economic (2008 projection) and skim data.

Table 2.3 Calibrated Multinomial Logit Model with Balancing Process

Autos	Variables	Coef.	Std. Err.	t-stat
Zero	Constant	-2.491	0.374	-6.658
	INCQTL1*HHWRKR0	4.420		
	INCQTL1*HHWRKR1	1.384		
	INCQTL2*HHWRKR0	3.422		
	INCQTL3*HHWRKR0	1.648		
	INCQTL4	-1.939		
	INCQTL5	-2.641		
	INCQTL1*HHDEN	0.038	0.009	4.259
	EMPDEN	0.023	0.004	5.441
One	Constant	-0.815		
	INCQTL1	2.320	0.130	17.869
	INCQTL2	1.837	0.126	14.622
	INCQTL3*HHWRKR0	1.324	0.187	7.097
	INCQTL3*HHWRKR1	1.000		
	INCQTL4*HHWRKR0	0.255	0.257	0.990
	HHSIZE1	2.500	0.130	19.195
	HHSIZE2	0.593	0.111	5.366
	INCQTL5*HHDEN	0.060	0.013	4.660
Two	Constant	1.129		
	HHWRKR2	0.904	0.147	6.139
	INCQTL4*HHDEN	-0.050	0.005	-9.188
	INCQTL5*HHDEN	0.029	0.010	2.763
	INCQTL4*HHWRKR1	1.100		
	INCQTL5*HHWRKR1	1.100		
	INCQTL1*HHWRKR2	1.000		
	INCQTL1*HHWRKR3	1.000		
Three	Constant	-0.572		
	INCQTL1	-0.320		
	INCQTL2	-0.200		
	INCQTL5	1.674	0.174	9.635
	INCQTL4*HHWRKR0	-0.264	0.297	-0.889
	HHWRKR2	1.022	0.163	6.268
	HHWRKR3	2.598	0.272	9.551
	HHSIZE1	-0.644	0.231	-2.781
	REMP30T	-25.400	9.699	-2.619
Four +	Constant	-1.010		
	INCQTL1	-0.340		
	INCQTL2	-0.200		
	INCQTL5	1.800	0.198	9.087
	INCQTL4*HHWRKR0	-0.614	0.530	-1.160
	HHWRKR2	1.268	0.191	6.633
	HHWRKR3	3.581	0.283	12.657
	Single Family Housing Unit	-0.687	0.155	-4.425
	HHDEN	-0.027	0.007	-3.917

Table 2.4 Calibrated Multinomial Logit Model with No Balancing Process

Autos	Variables	Coef.	Std. Err.	t-stat
Zero	Constant	-3.186	0.379	-8.401
	INCQTL1*HHWRKR0	3.845		
	INCQTL1*HHWRKR1	1.189		
	INCQTL2*HHWRKR0	3.091		
	INCQTL3*HHWRKR0	1.153		
	INCQTL4	-1.525		
	INCQTL5	-1.557		
	INCQTL1*HHDEN	0.038	0.009	4.310
	EMPDEN	0.024	0.004	5.528
One	Constant	-1.041		
	INCQTL1	1.922	0.137	13.978
	INCQTL2	1.447	0.13	11.165
	INCQTL3*HHWRKR0	1.231	0.187	6.579
	INCQTL3*HHWRKR1	1.000		
	INCQTL4*HHWRKR0	0.332	0.255	1.302
	HHSIZE1	2.545	0.130	19.565
	HHSIZE2	0.597	0.111	5.399
	INCQTL5*HHDEN	0.050	0.013	3.921
Two	Constant	0.391		
	HHWRKR2	0.898	0.146	6.137
	INCQTL4*HHDEN	-0.027	0.006	-4.872
	INCQTL5*HHDEN	0.031	0.010	2.999
	INCQTL4*HHWRKR1	1.100		
	INCQTL5*HHWRKR1	1.100		
	INCQTL1*HHWRKR2	1.000		
	INCQTL1*HHWRKR3	1.000		
Three	Constant	-1.102		
	INCQTL1	-0.218	0.177	-1.233
	INCQTL2	-0.148	0.139	-1.065
	INCQTL5	1.093	0.179	6.092
	INCQTL4*HHWRKR0	-0.664	0.298	-2.231
	HHWRKR2	0.816	0.165	4.954
	HHWRKR3	2.393	0.274	8.73
	HHSIZE1	-0.521	0.232	-2.243
	REMP30T	-21.114	9.536	-2.214
Four +	Constant	-1.682		
	INCQTL1	-0.069	0.231	-0.299
	INCQTL5	1.304	0.198	6.588
	INCQTL4*HHWRKR0	-0.941	0.530	-1.775
	HHWRKR2	1.091	0.192	5.682
	HHWRKR3	3.401	0.285	11.951
	Single Family Housing Unit	-0.834	0.162	-5.132
	HHDEN	-0.023	0.007	-3.214

Table 2.5 Calibrated Old Multinomial Logit Model with No Balancing Process

Autos	Variables	Coef.	Std. Err.	t-stat
Zero	Constant	-3.186	0.379	-8.401
	INCQTL1*HHWRKR0	3.975		
	INCQTL1*HHWRKR1	1.309		
	INCQTL2*HHWRKR0	3.201		
	INCQTL3*HHWRKR0	1.263		
	INCQTL4	-1.385		
	INCQTL5	-1.437		
	INCQTL1*HHDEN	0.038	0.009	4.310
	EMPDEN	0.024	0.004	5.528
One	Constant	-1.041		
	INCQTL1	1.922	0.137	13.978
	INCQTL2	1.447	0.130	11.165
	INCQTL3*HHWRKR0	1.231	0.187	6.579
	INCQTL3*HHWRKR1	1.000		
	INCQTL4*HHWRKR0	0.332	0.255	1.302
	HHSIZE1	2.545	0.130	19.565
	HHSIZE2	0.597	0.111	5.399
	INCQTL5*HHDEN	0.050	0.013	3.921
Two	Constant	0.416		
	HHWRKR2	0.898	0.146	6.137
	INCQTL4*HHDEN	-0.027	0.006	-4.872
	INCQTL5*HHDEN	0.031	0.010	2.999
	INCQTL4*HHWRKR1	1.100		
	INCQTL5*HHWRKR1	1.100		
	INCQTL1*HHWRKR2	1.000		
	INCQTL1*HHWRKR3	1.000		
Three	Constant	-1.160		
	INCQTL1	-0.218	0.177	-1.233
	INCQTL2	-0.148	0.139	-1.065
	INCQTL5	1.093	0.179	6.092
	INCQTL4*HHWRKR0	-0.664	0.298	-2.231
	HHWRKR2	0.816	0.165	4.954
	HHWRKR3	2.393	0.274	8.730
	HHSIZE1	-0.521	0.232	-2.243
	REMP30T	-21.114	9.536	-2.214
Four +	Constant	-1.730		
	INCQTL1	-0.069	0.231	-0.299
	INCQTL5	1.304	0.198	6.588
	INCQTL4*HHWRKR0	-0.941	0.530	-1.775
	HHWRKR2	1.091	0.192	5.682
	HHWRKR3	3.401	0.285	11.951
	Single Family Housing Unit	-0.834	0.162	-5.132
	HHDEN	-0.023	0.007	-3.214

The estimated MNL models are applied to 2008 input data with balancing process and 2010 input data with no balancing process to obtain the predicted household vehicle ownership shares. The following table shows input data. Input data are different based on whether balancing process is applied or not. The PUMS data based on ACS 2006-2010 and 2008 NHTS add-on are utilized to obtain the observed household vehicle ownership shares in the MAG region. The predicted household vehicle ownership shares from new MNL with BP and NBP are compared with the observed shares from PUMS and NHTS data. The results are shown in the following graph. From the chart, it can be observed that predicted shares with both BP and NBP are very close to the observed shares from ACS 2006-2010. This chart also includes the previous result from the model with no balancing process based on the previous version of input data.

Table 2.6 Input Data with and without Balancing Process

Balancing Process	Input Data Description	File Name
Yes	TAZ data file	Tazdata.asc (2008)
	Housing type factors	Phewgts.dat
	Number of households by income, household size and worker size	Worker.asc
	Transit skim matrix (In-vehicle time, First wait, Transfer wait, Access walk, Egress walk)	Tskm_pk_wlk_com_wlk.mtx
No	TAZ data file	Tazdata.asc (2010)
	Number of residential households by combination (PopSyn output)	Htouthh1.asc
	Household type factors for non-residential households	Htoutfactor.asc
	Transit skim matrix (In-vehicle time, First wait, Transfer wait, Access walk, Egress walk)	Tskm_pk_wlk_com_wlk.mtx

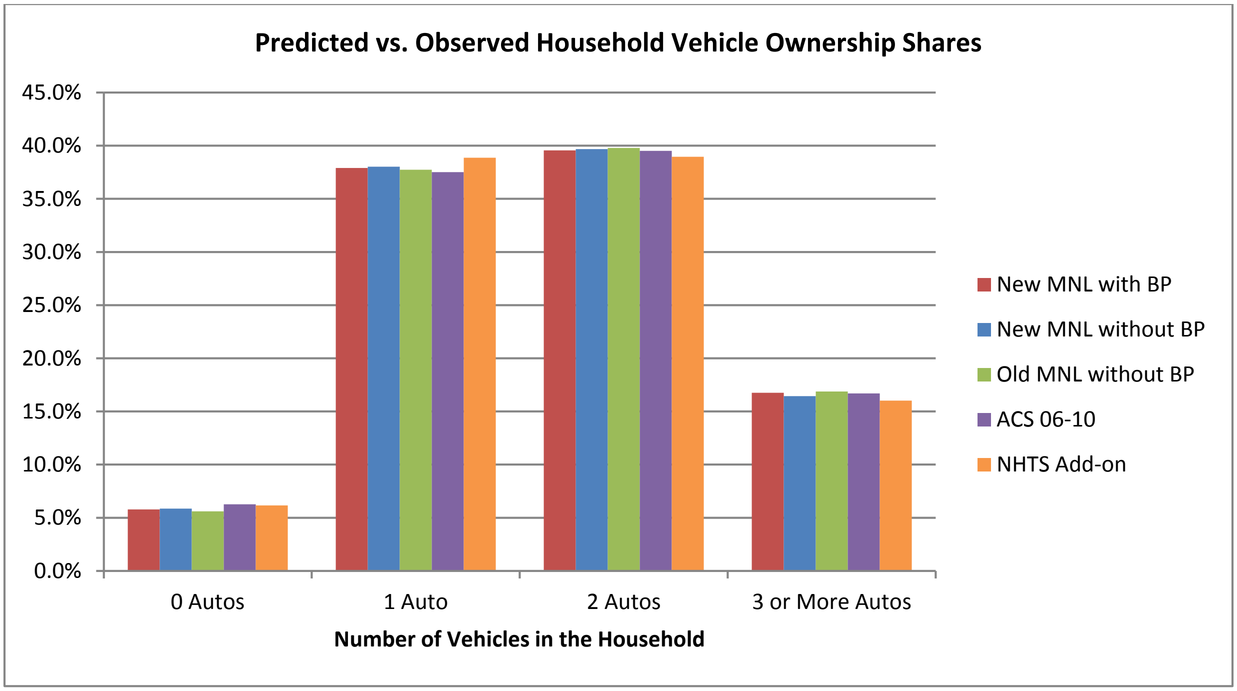


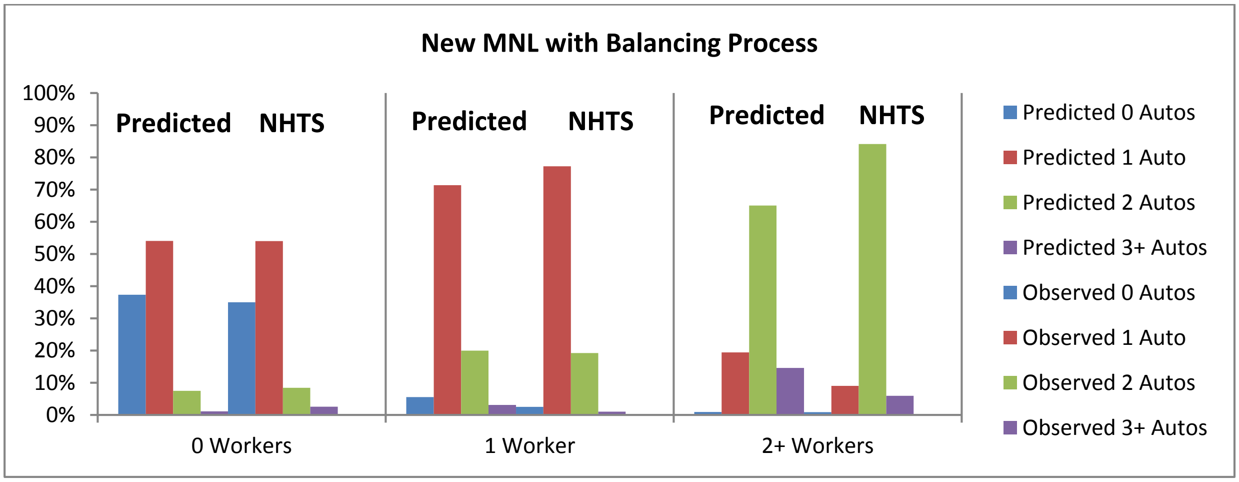
Figure 2.1 Predicted versus Observed Household Vehicle Ownership Shares

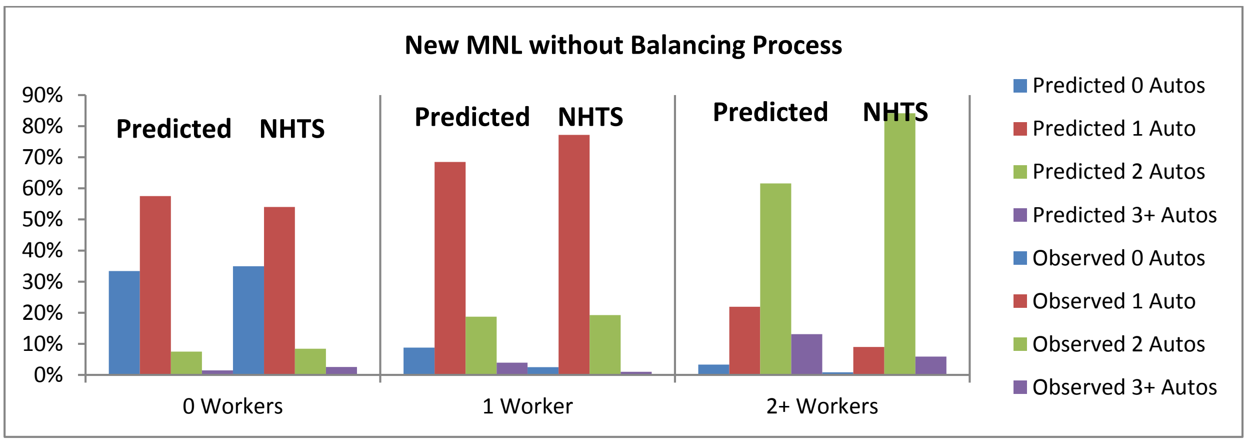
2.4 Model Validation

For the household vehicle ownership model validation, a various source of data such as 2008 NHTS add-on data and 2006-10 ACS data were used. This section summarizes the different level of comparison using those data.

2.4.1 2008 NHTS Add-On Data Aggregate Level Comparison

A series of graphs were created by stratifying household income and number of workers in the household. Four categories of household vehicle ownership (0, 1, 2, and 3+ autos) are shown in these graphs. As sample size of the households with 3 or more workers in 2008 NHTS add-on data is small, 2 and more workers in the household are combined into one category. Hence, 0, 1, and 2+ workers are represented in the graphs for comparison between the observed and the predicted shares. Figures 2.2-2.6 show the aggregate level comparison by income group.





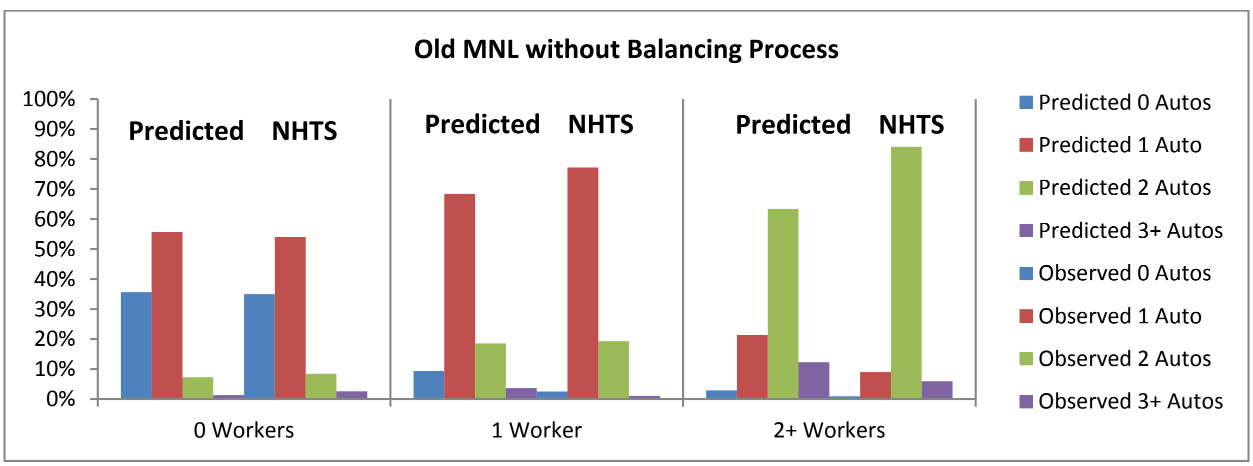
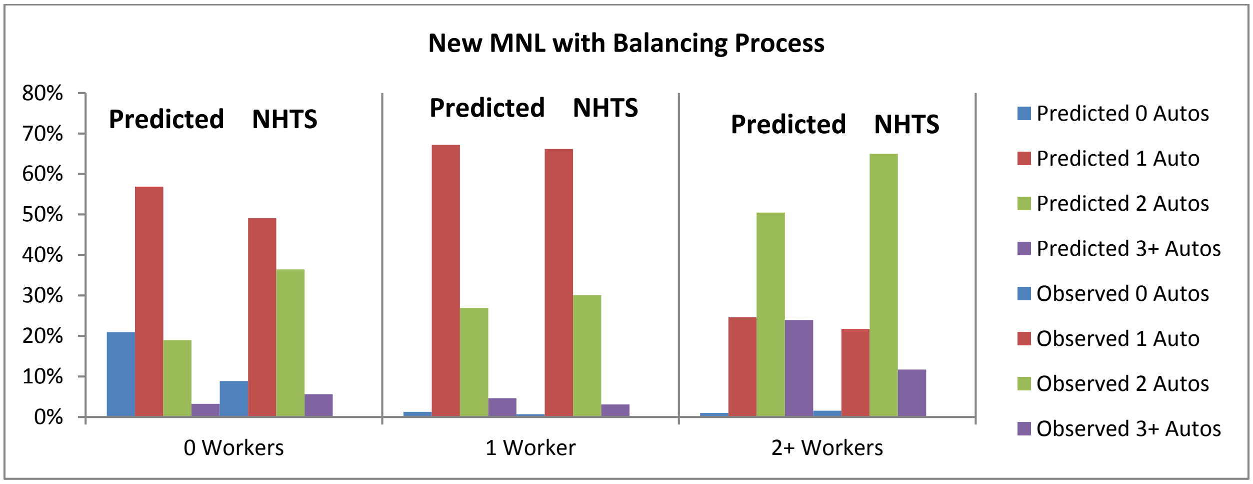
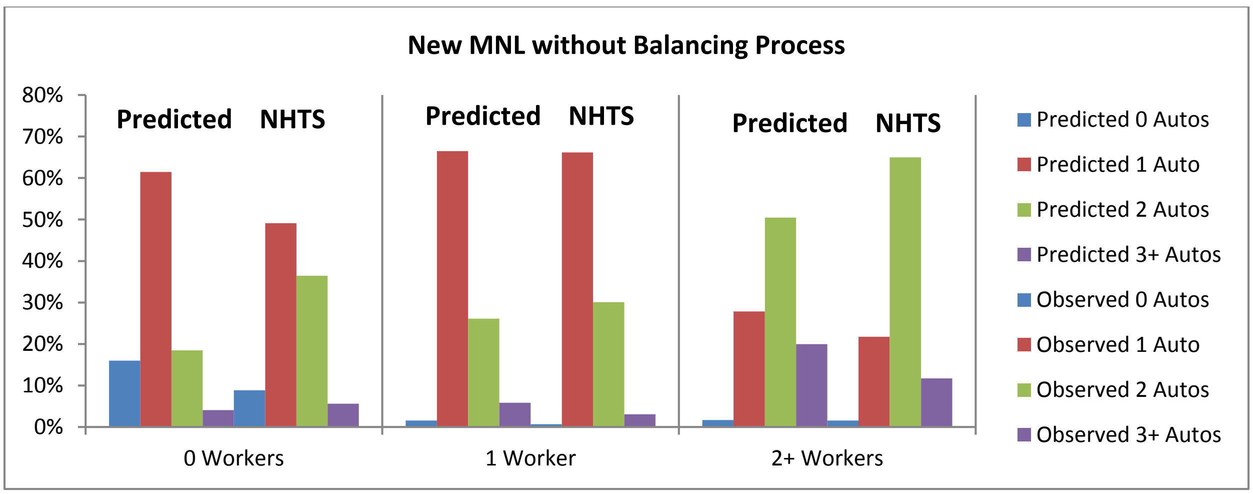


Figure 2.2 2008 NHTS Add-On Data Aggregate Level Comparison: Income Quintile 1





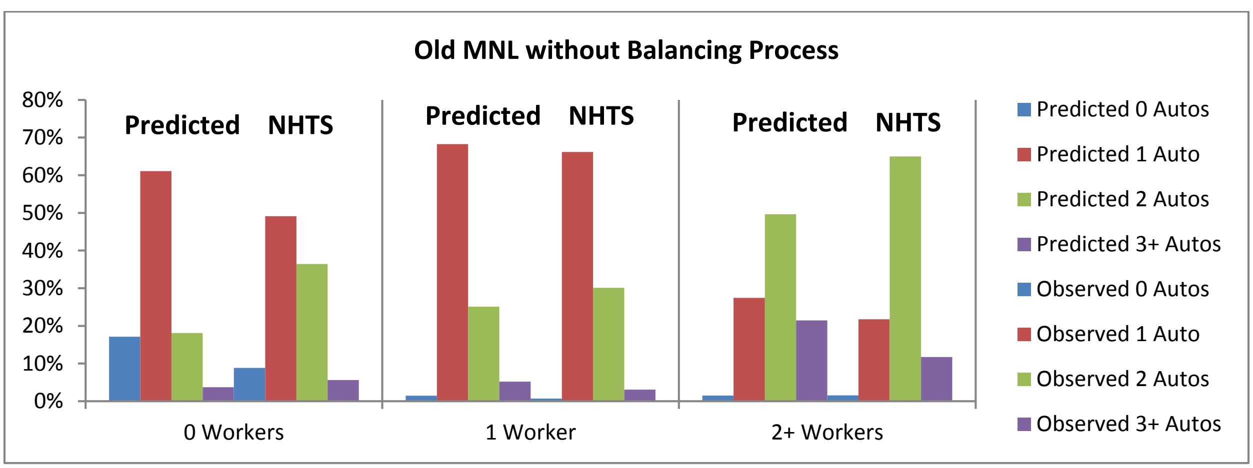


Figure 2.3 2008 NHTS Add-On Data Aggregate Level Comparison: Income Quintile 2

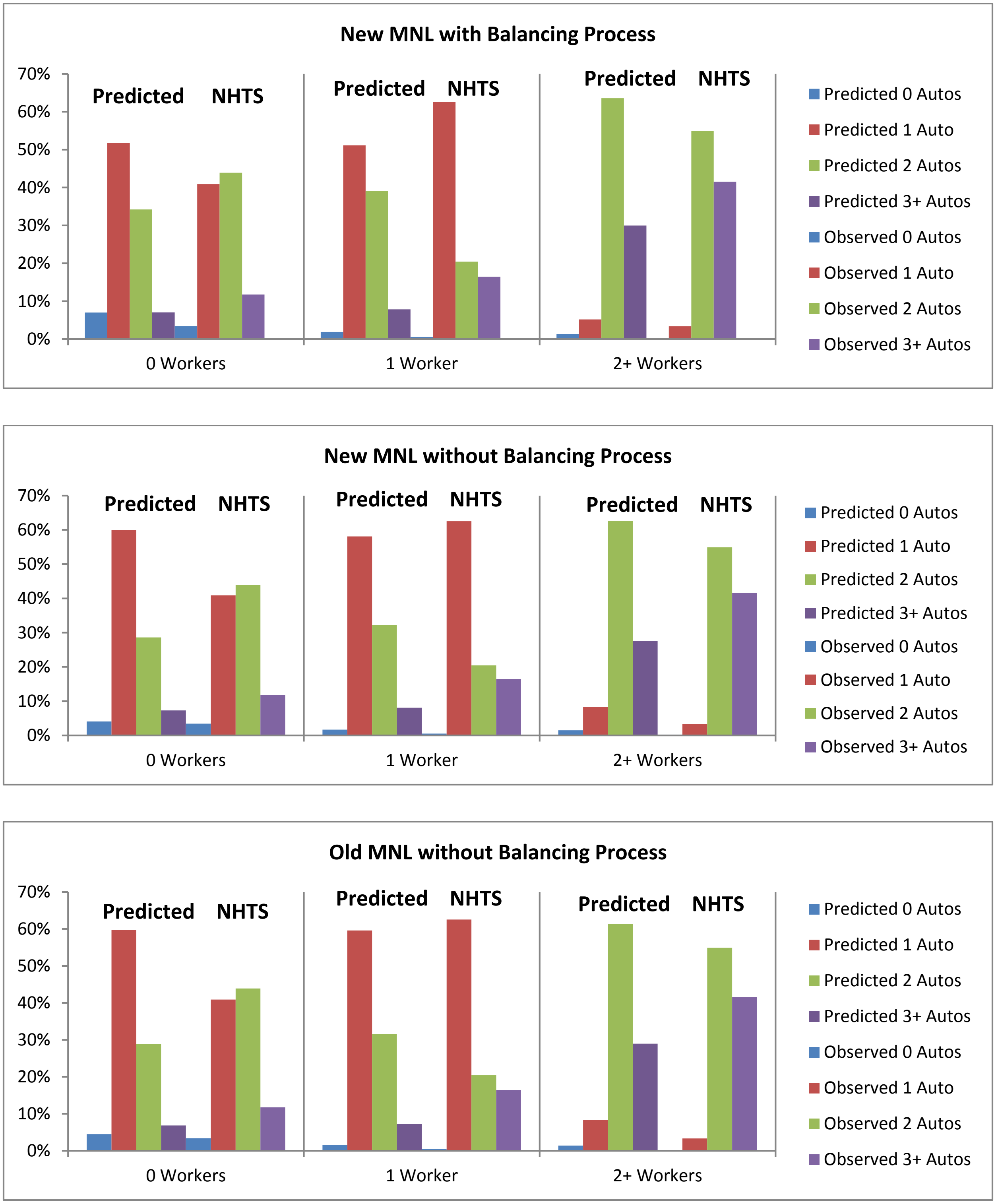


Figure 2.4 2008 NHTS Add-On Data Aggregate Level Comparison: Income Quintile 3

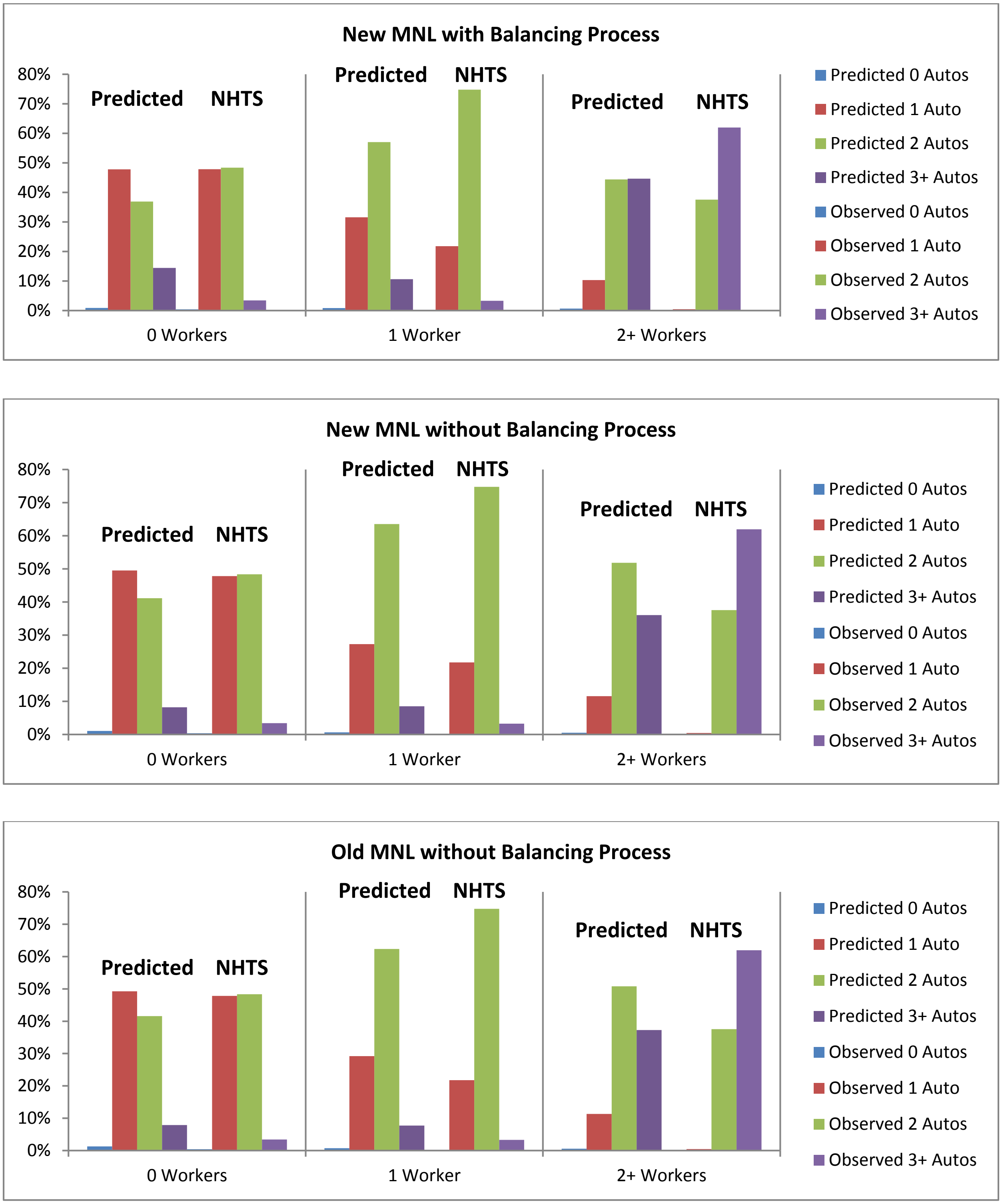
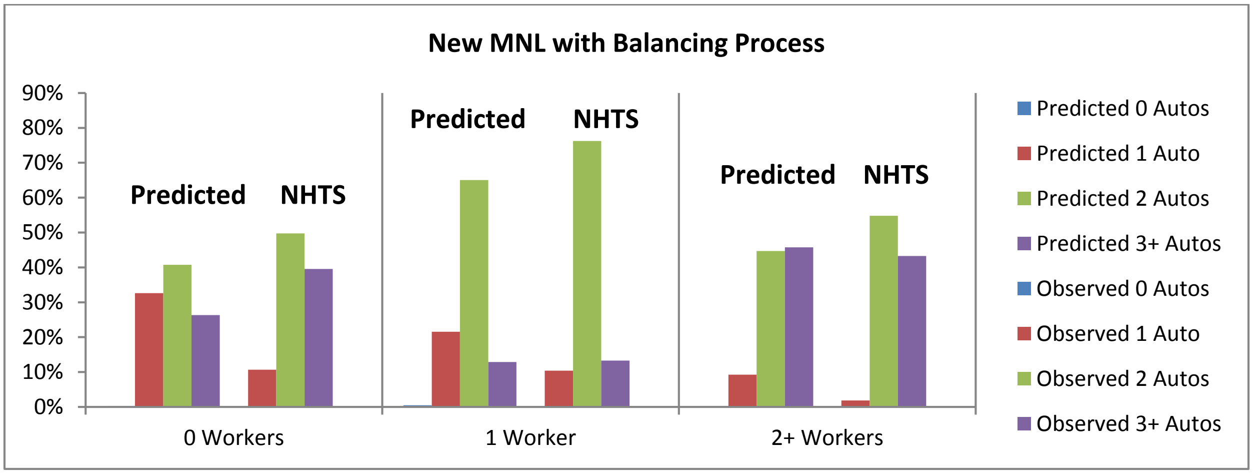
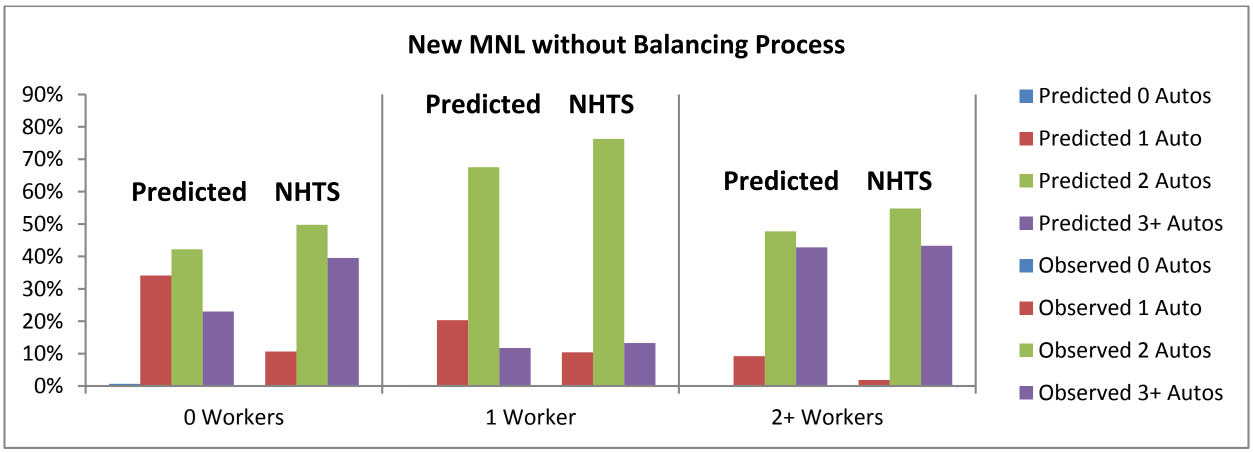


Figure 2.5 2008 NHTS Add-On Data Aggregate Level Comparison: Income Quintile 4





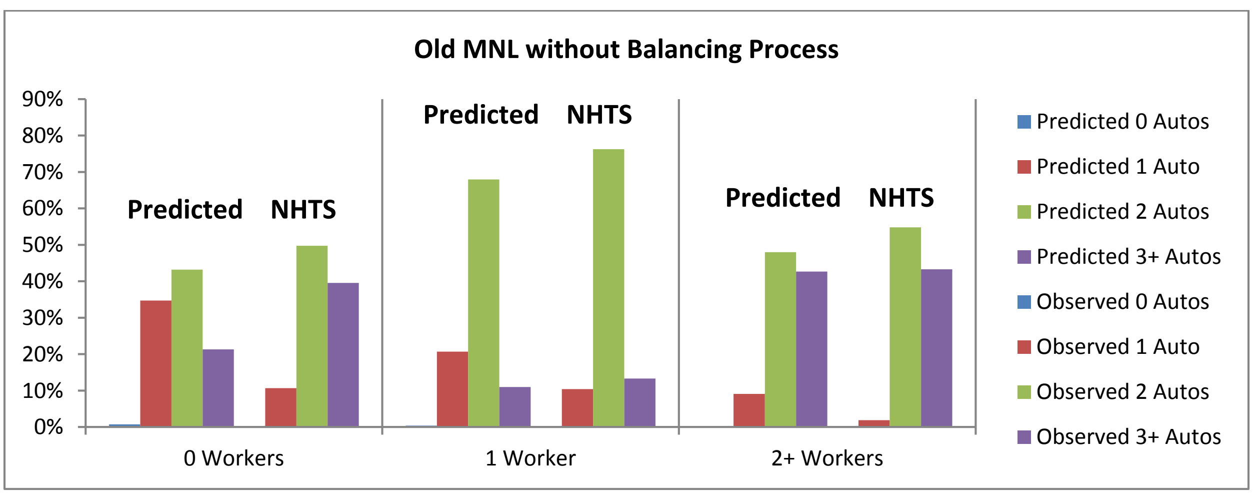


Figure 2.6 2008 NHTS Add-On Data Aggregate Level Comparison: Income Quintile 5

2.4.2 2006-2010 ACS Data PUMA Level Comparison

ACS 2006-2010 data are used to compare the predicted household vehicle ownership with observed shares by stratifying household income in the PUMA level. Since ACS 2006-2010 data do not include the number of workers in the households, comparisons were made using household income in the PUMA level. The following graphs show 45 degree line with minimal dispersion in case of both BP and new NBP. It means that the MNL models for both BP and NBP can accurately predict household vehicle ownership shares in the PUMA level. Figures 2.7-2.11 show the PUMA level comparison by income group.

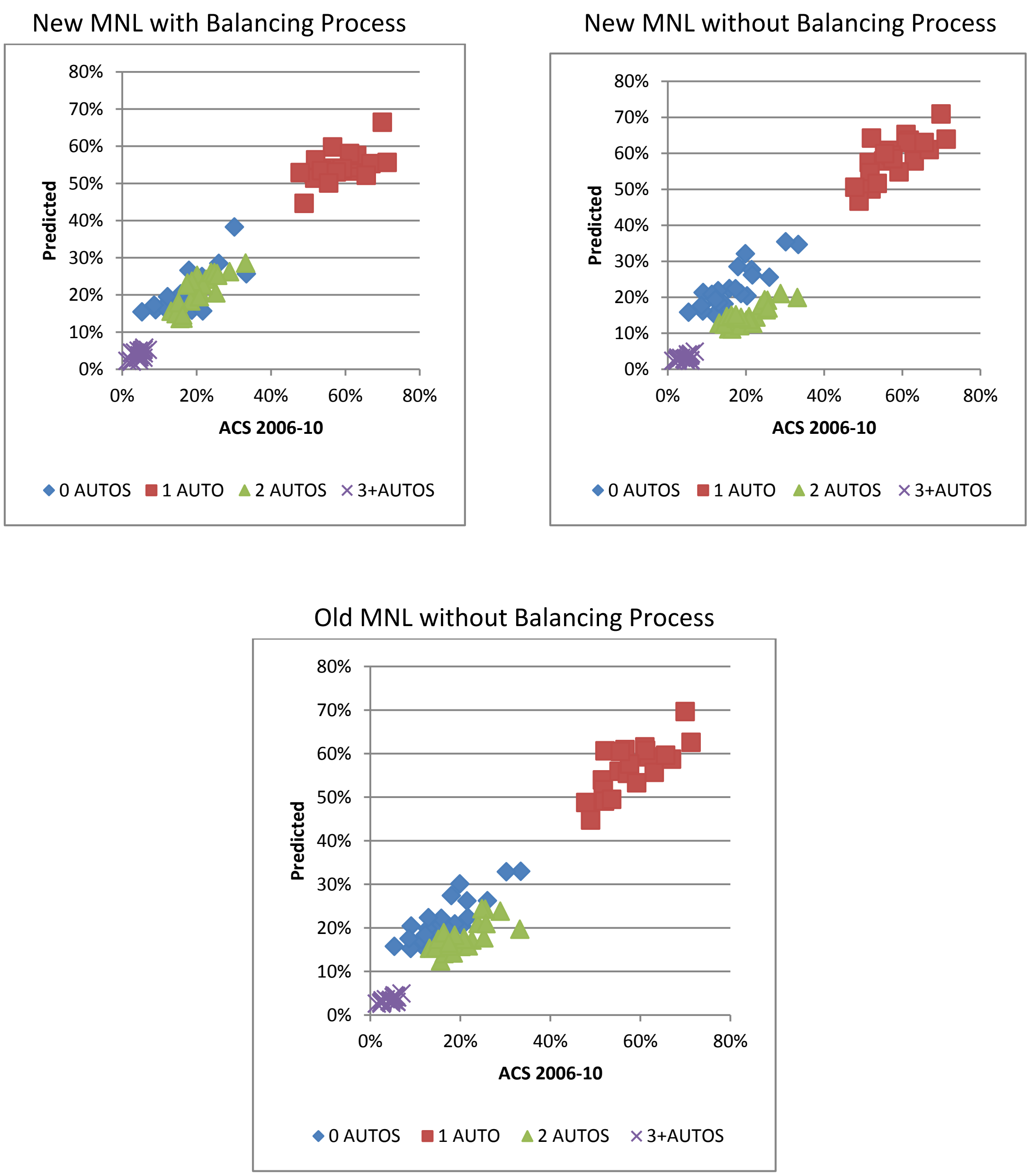


Figure 2.7 2006-2010 ACS Data PUMA Level Comparison: Income Quintile 1

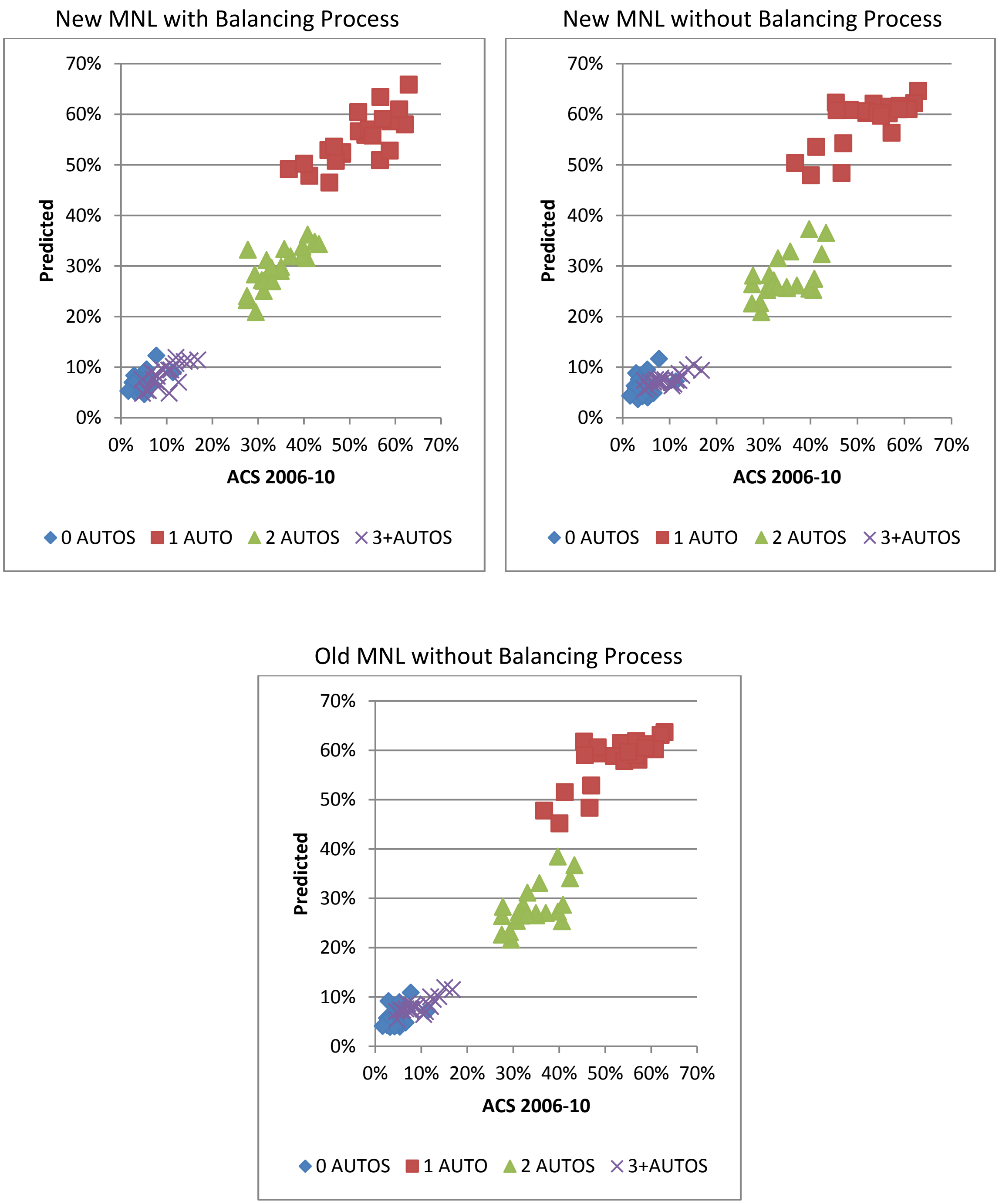


Figure 2.8 2006-2010 ACS Data PUMA Level Comparison: Income Quintile 2

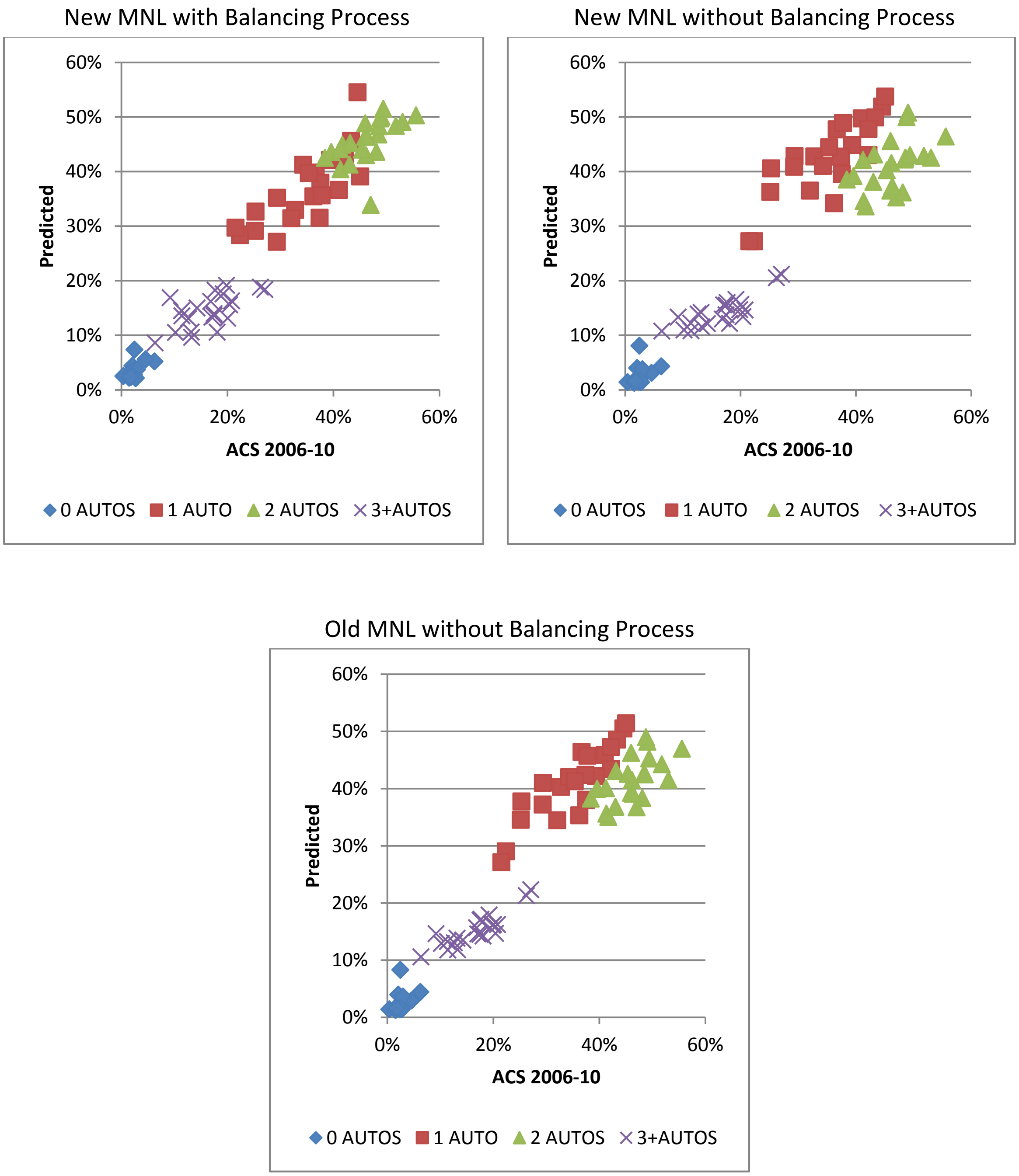


Figure 2.9 2006-2010 ACS Data PUMA Level Comparison: Income Quintile 3

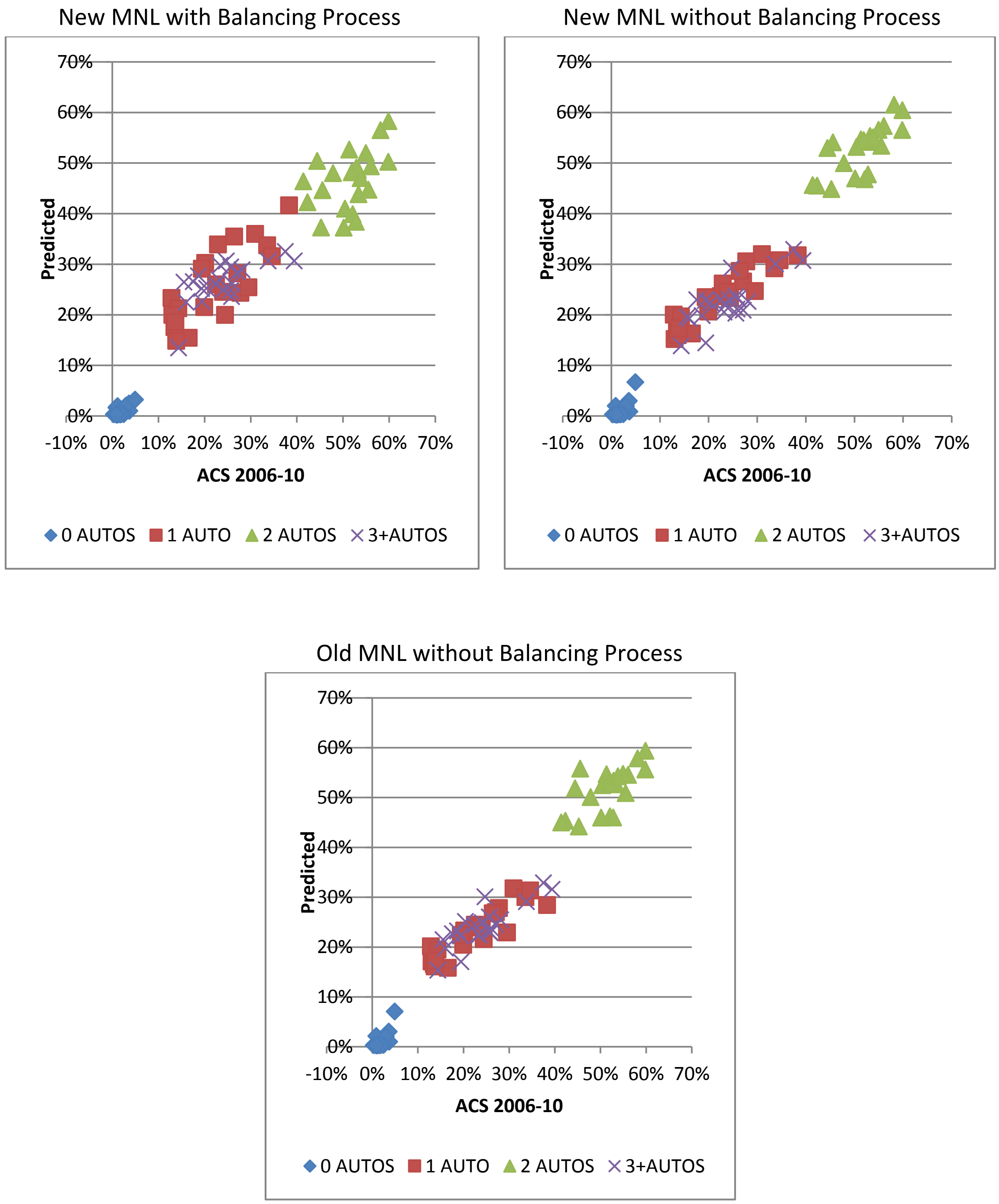


Figure 2.10 2006-2010 ACS Data PUMA Level Comparison: Income Quintile 4

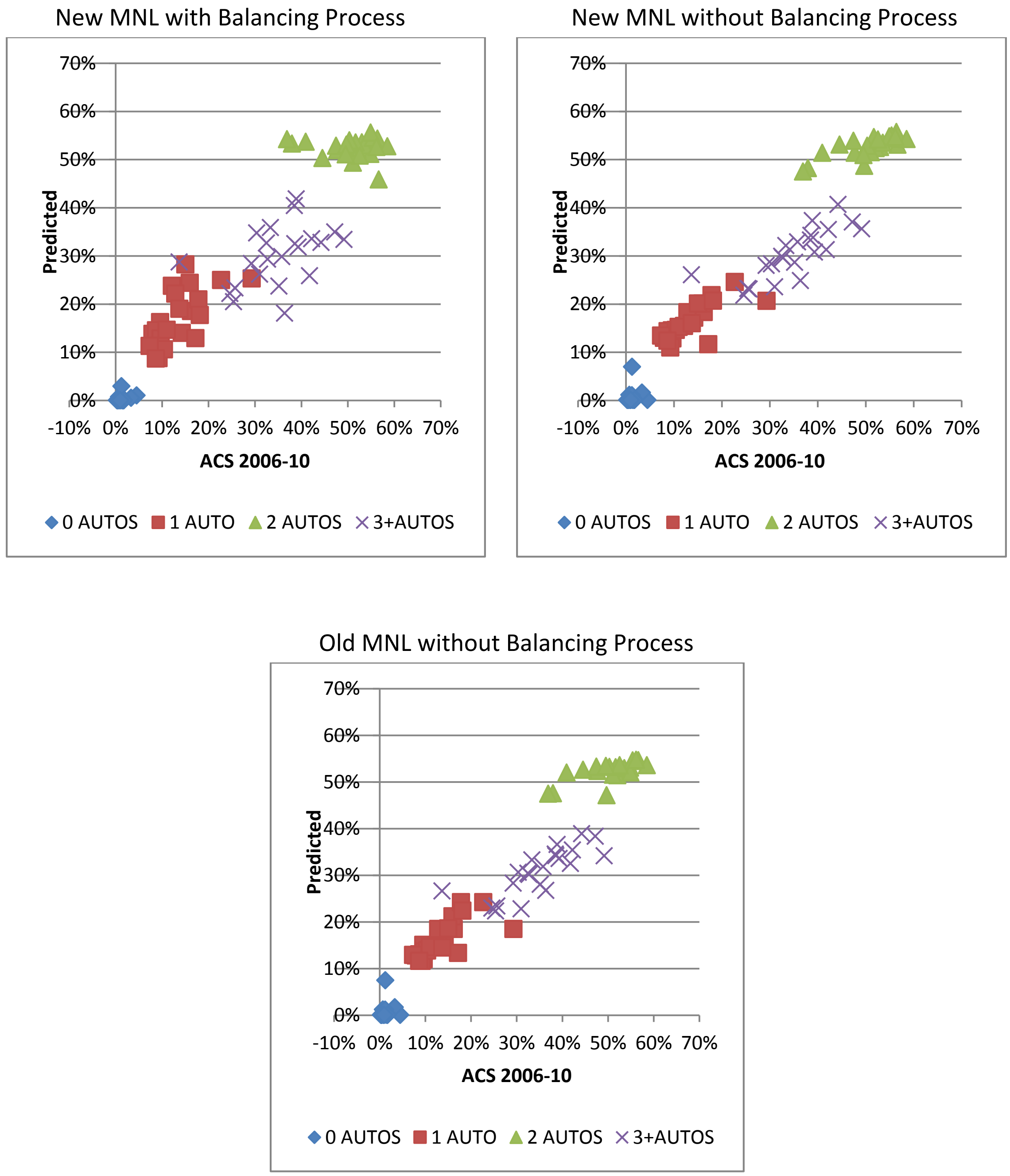


Figure 2.11 2006-2010 ACS Data PUMA Level Comparison: Income Quintile 5

2.4.3 2008 NHTS MAG Add-On Data MPA, District and PUMA Level Comparisons

The predicted household vehicle ownership shares are compared with the observed shares from NHTS MAG add-on data at different geographic levels such as MPA, District, and PUMA. Four graphs are provided at each geographic level comparing predicted shares using BP/NBP and observed shares from unweighted samples. The comparison graphs for NBP with unweighted observed sample displays almost 45 degree at all geographic levels. Taking this finding into consideration along with the observation above, model with NBP would predict household vehicle ownership shares in the households more accurately than the model with BP. This memo keeps the previous results of no balancing process for comparison with new results of no balancing process which are based on latest 2010 input data sets. The old and the new results of NBP models are shown next to each other. From the comparisons, it can be observed that the results of the old and new NBP models are very similar.

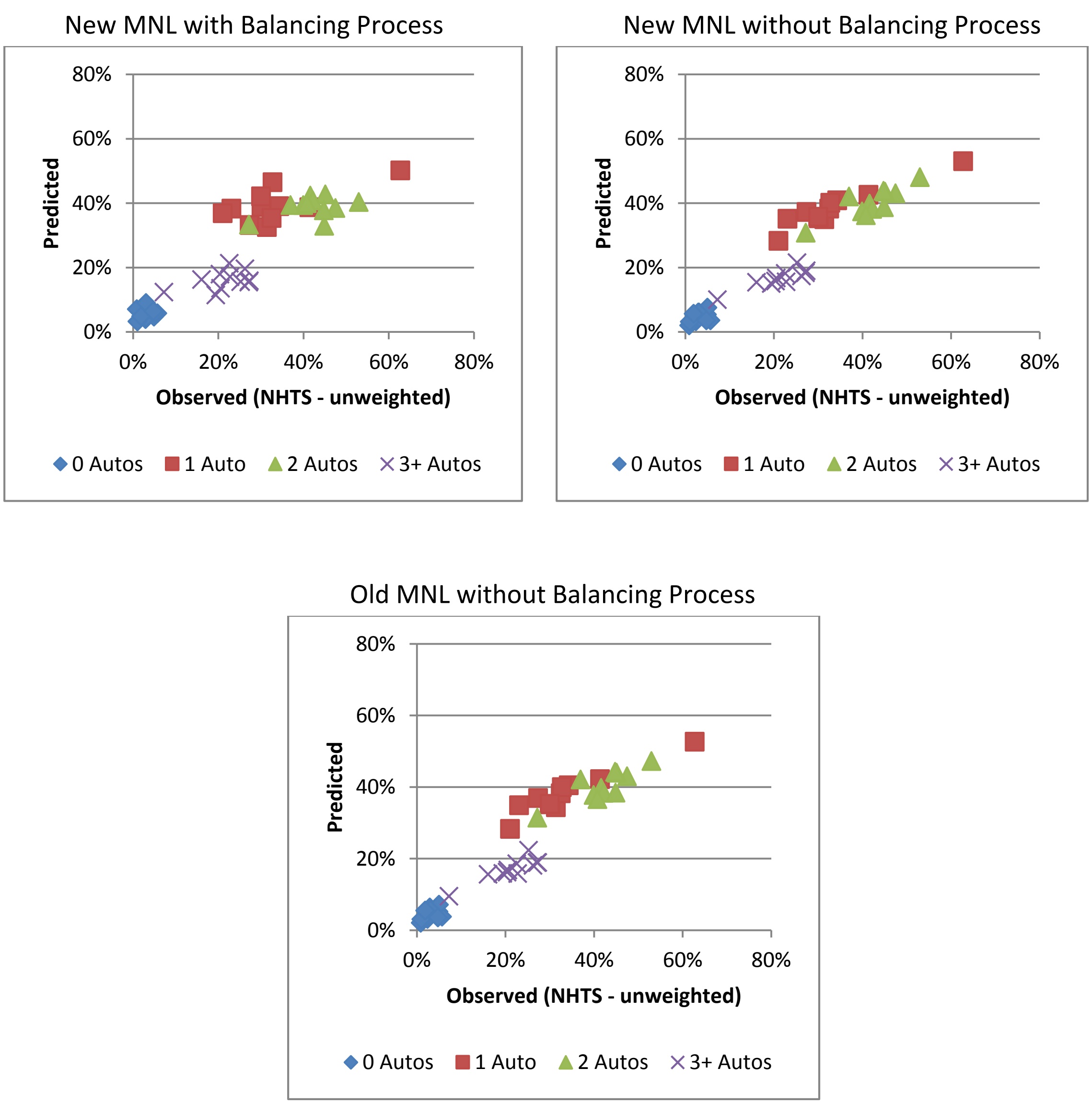


Figure 2.12 2008 NHTS MAG Add-On Data MPA Level Comparison

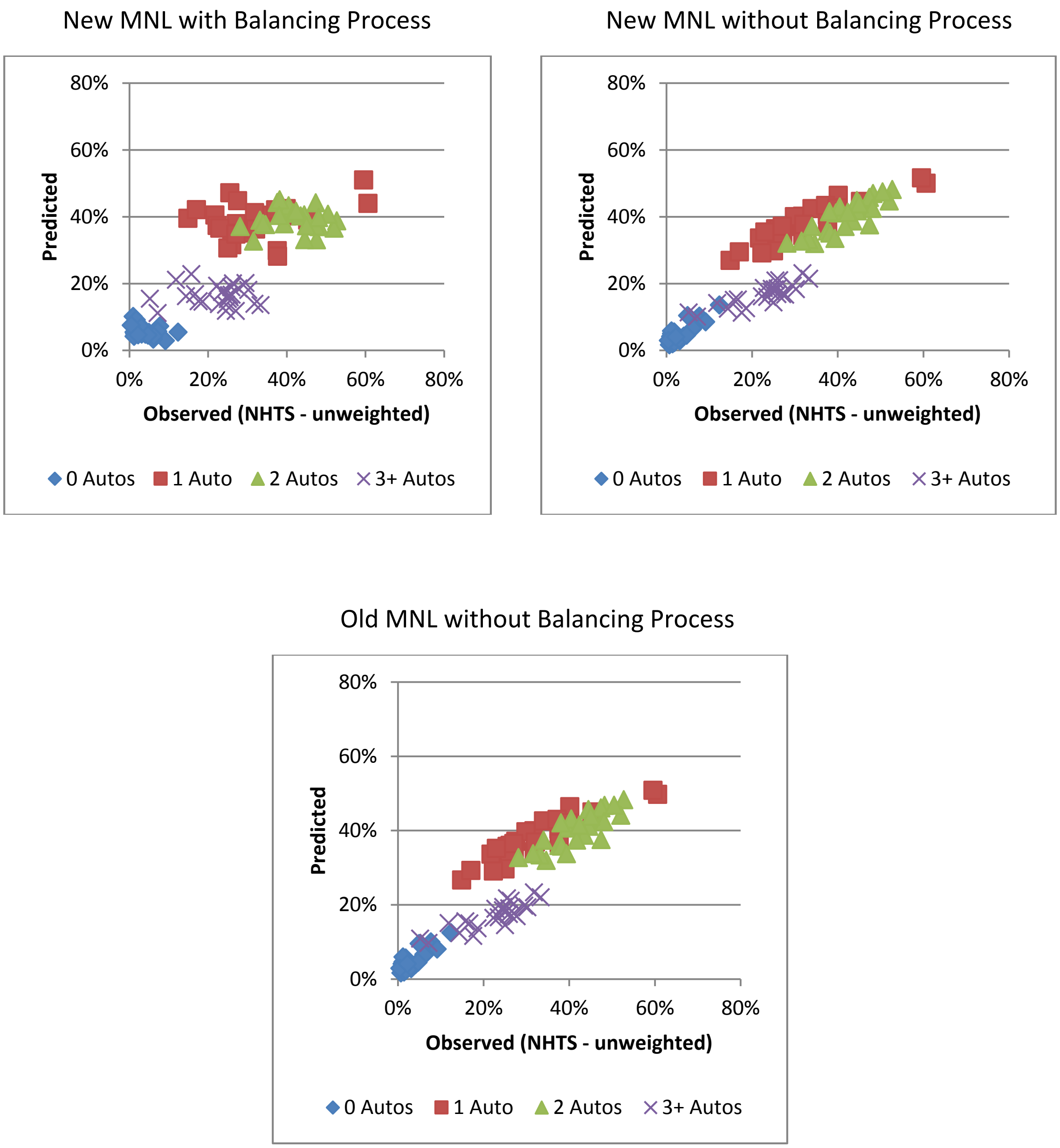


Figure 2.13 2008 NHTS MAG Add-On Data District Level Comparison

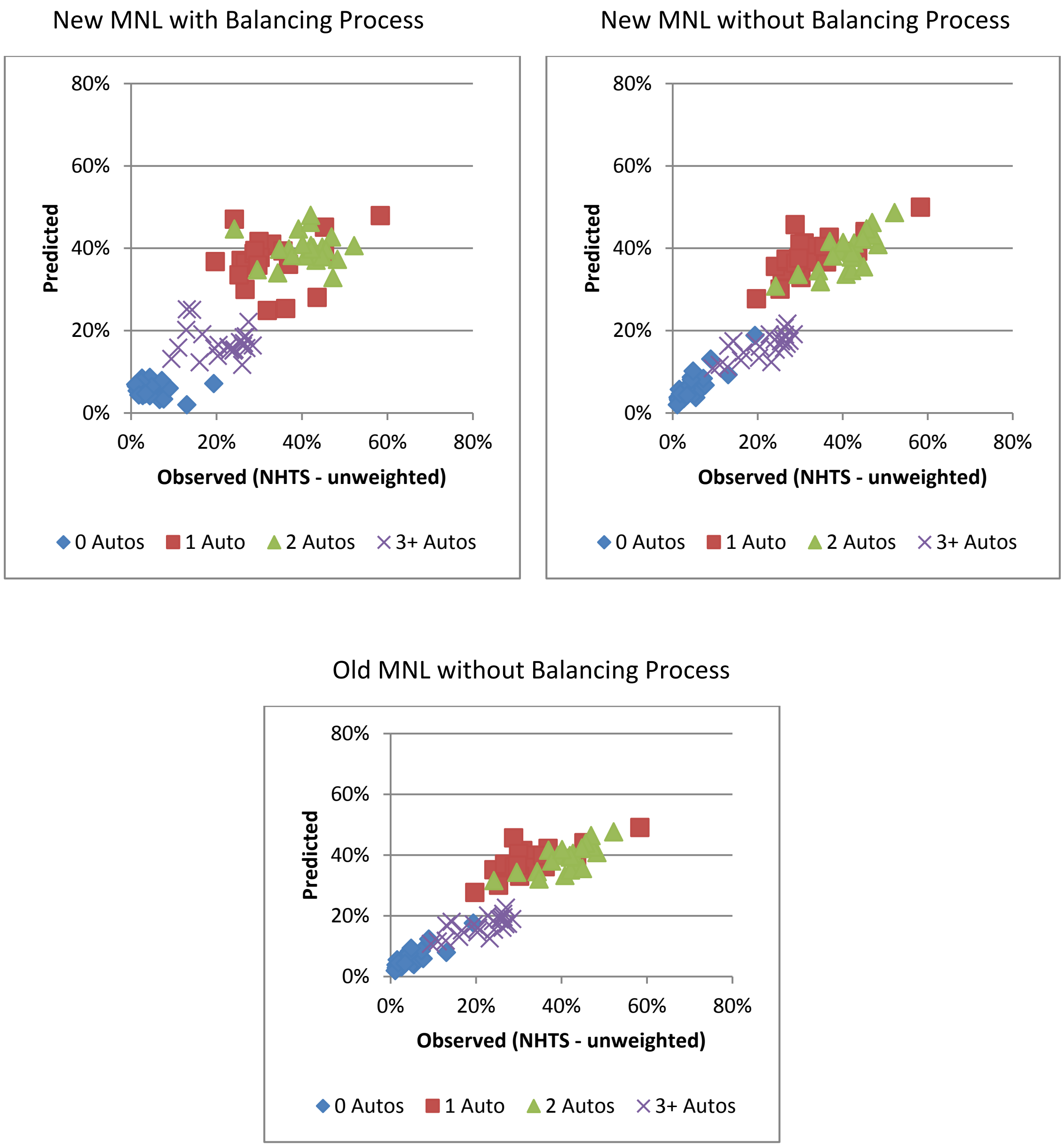


Figure 2.14 2008 NHTS MAG Add-On Data PUMA Level Comparison

2.4.4 2006-2010 ACS 5 Year Estimate Comparison

Predicted household vehicle ownership shares are compared with 2006-2010 ACS 5 year estimates stratified by number of workers in household and vehicle availability. The comparisons are very encouraging, suggesting that the estimated models are predicting household vehicle ownerships very well. Table 2.7 and Figures 2.15-2.16 show the auto sufficiency and ownership comparison with balancing process based on the 2008 socioec data. Table 2.8 and Figures 2.17-2.18 show auto sufficiency comparison without balancing process based on 2010 socioec data due to availability of the survey data.

Table 2.7 Number of Households by Auto Sufficiency with Balancing Process

	# of Households		Percent of Households		Difference	
	With Balancing	ACS	With Balancing	ACS	Absolute	Percent
0 Auto	97,829	93,566	5.8%	6.2%	-0.5%	-7.3%
Autos<Workers	84,107	74,976	5.0%	5.0%	0.0%	-0.6%
Autos>=Workers	1,511,166	1,332,286	89.3%	88.8%	0.5%	0.5%
Total	1,693,103	1,500,828	100.0%	100.0%	0.0%	0.0%

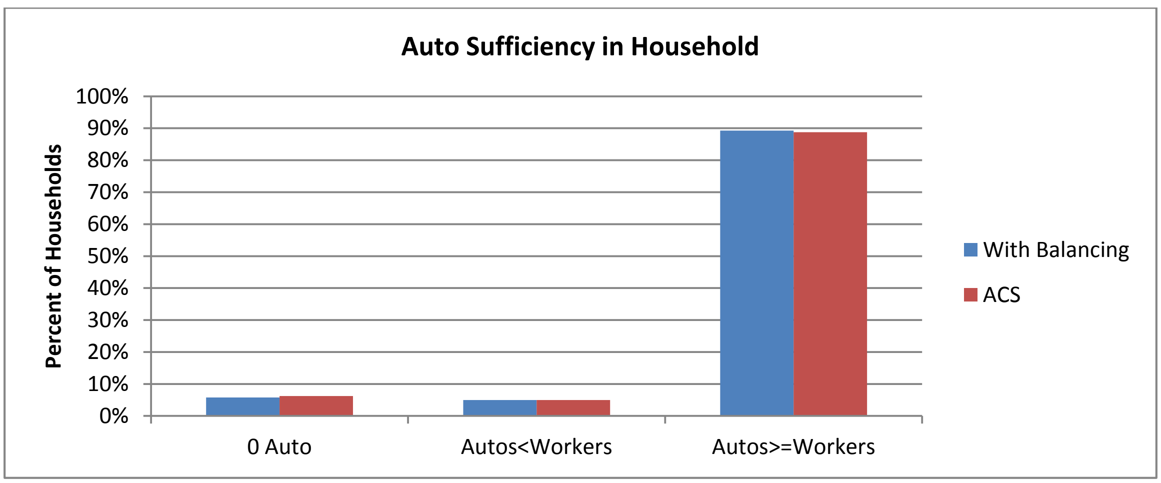


Figure 2.15 Auto Sufficiency in Household with Balancing Process

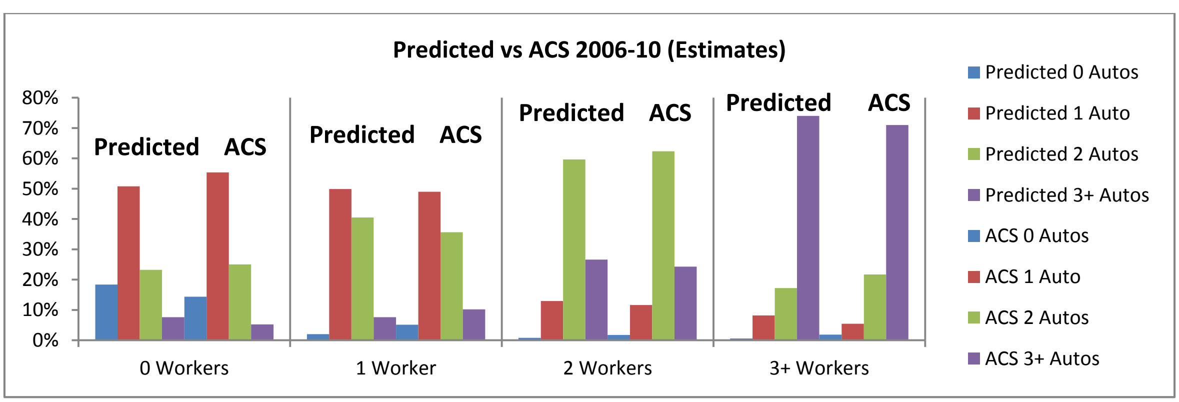


Figure 2.16 Predicted versus ACS 2006-10 Auto Ownership Comparison with Balancing Process

Table 2.8 Number of Households by Auto Sufficiency without Balancing Process

	# of Households		Percent of Households		Difference	
	No Balancing	ACS	No Balancing	ACS	Absolute	Percent
0 Auto	99,623	93,566	5.9%	6.2%	-0.4%	-6.1%
Autos<Workers	87,886	74,976	5.2%	5.0%	0.2%	3.4%
Autos>=Workers	1,513,864	1,332,286	89.0%	88.8%	0.2%	0.2%
Total	1,701,372	1,500,828	100.0%	100.0%	0.0%	0.0%

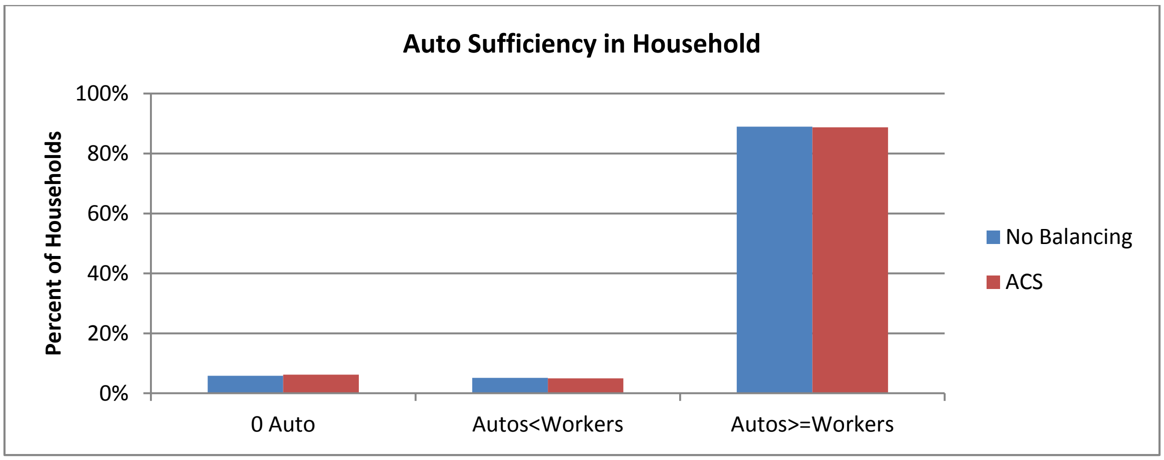


Figure 2.17 Auto Sufficiency in Household without Balancing Process

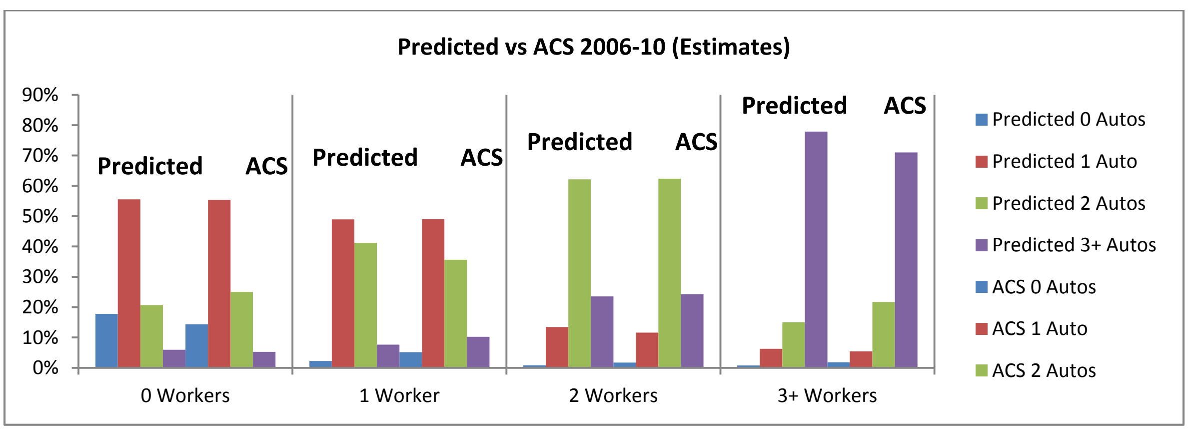


Figure 2.18 Predicted versus ACS 2006-10 Auto Ownership Comparison without Balancing Process

3. Vehicle Fleet Composition Model Development

Activity-based models provide an accurate account of travel which can then be assigned using network assignment models to quantify vehicle miles travelled (VMT) by all the households in a region. The calculated VMT in a vehicle fleet composition model is used as a determinant of greenhouse gas (GHG) emissions and fuel consumption. Emission calculations for region wide travel are usually carried out in emission modeling software such as EMFAC (2014) and MOVES (2014) that take the output of network assignment models as input. These emission modeling software have default distributions that represent the vehicle mix of a region. While these default values provide a quick and simple way to compute emissions, they are often not responsive to policy measures that might influence the vehicle fleet mix. For travel demand models to accurately predict emission footprint of a region, it is necessary to model the vehicle fleet composition at household level, which would help forecast the fleet mix (and corresponding emissions) accurately at the regional level.

Interest in modeling household vehicle fleet composition has been growing for several decades. The necessity to implement such models as an integral part of the activity based modeling framework in order to accurately quantify the emission footprint of a region has become all the more important today with increasing pollution levels, GHG emissions and global warming. While there has been tremendous amount of study in simulating activity-travel patterns of households using activity-based models, simulation of vehicle fleet and vehicle type choice at the trip/tour level has only seen light in the recent years in a handful of activity-based models. Such model systems are still in the research phase and have not yet fully made their way to be included in activity-based models in practice. This study aimed at developing a robust framework to predict the fleet mix of a household using a Multiple Discrete Continuous Extreme Value (MDCEV) model in conjunction with several other models that control and constrain the prediction of fleet mix such that it is representative of the observed fleet mix in the base year. This will impart much confidence in prediction of fleet mix made for any future year using the developed model system.

3.1 Model Framework

Most of the activity-based model systems only model auto ownership at the household level without any consideration of the types of vehicles that the household owns. Very few activity-based models that are still in the research phase are beginning to incorporate fleet composition models as a part of their modeling framework. Generalizing the household vehicle ownership has important consequences in quantifying energy and emissions footprint at the household/regional level.

To overcome this issue, a fleet composition model framework is proposed in this study to simultaneously predict the number of vehicles owned by a household, body type of each of these vehicles, their vintage and the annual mileage put on each of these vehicles by the household. The system also predicts if the household owns multiple vehicles of the same body type–age category (for example a household might own two cars, both in the age category 0-5 years). The model framework is shown in Figure 3.1. Each of the models in the fleet composition modeling framework is explained below.

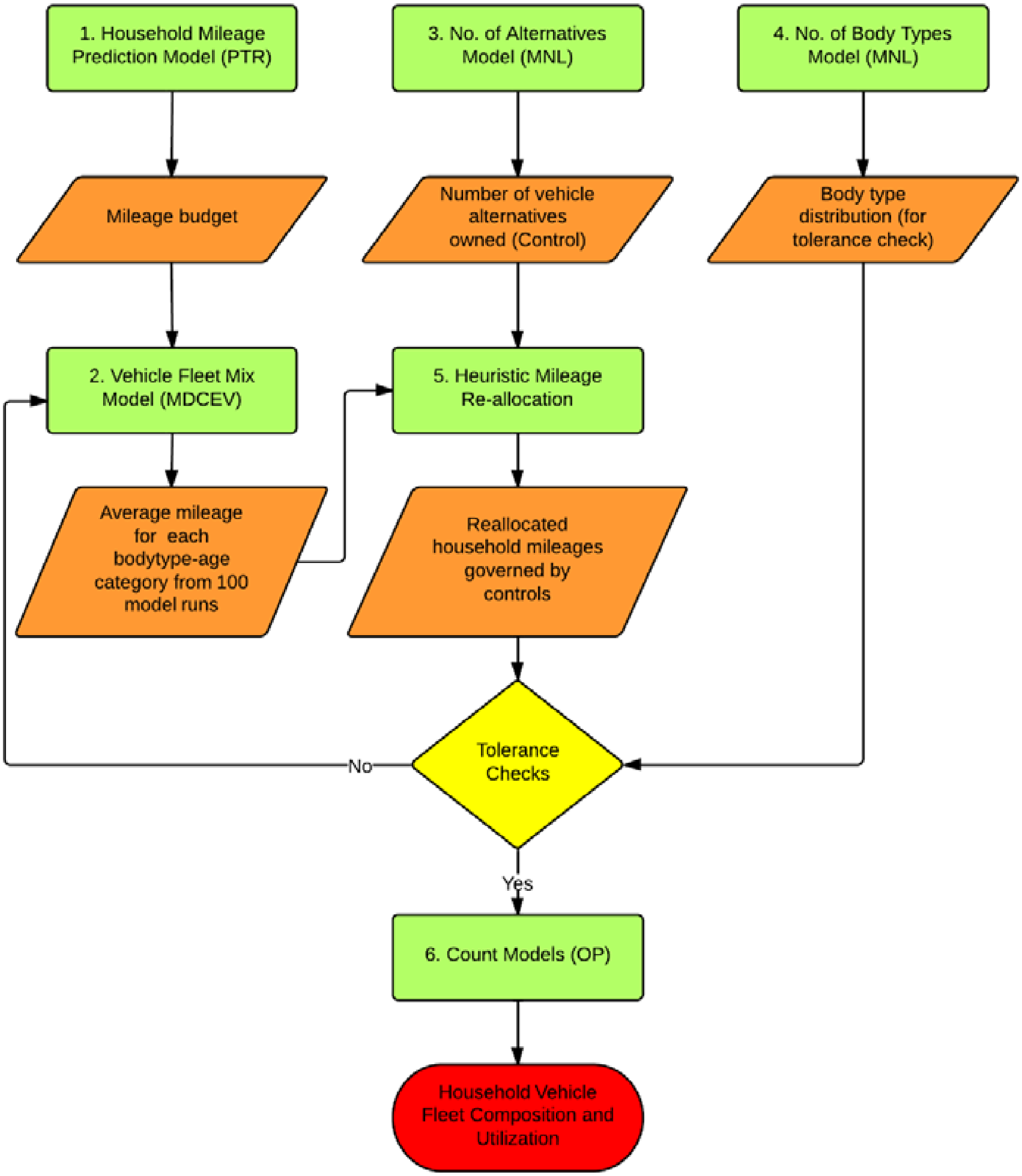


Figure 3.1 Vehicle Fleet Composition Model Framework

1. *Household Mileage Prediction Model*: The first element in the model system is a household mileage prediction model that predicts the annual motorized mileage consumption of households. This step is necessary as the subsequent component in the model system, the MDCEV model of fleet mix requires a mileage budget to allocate to the fleet owned by a household. The motorized mileage can be estimated using a simple linear regression model or some variation of it. Once the motorized mileage for each household is predicted, non-motorized mileage is computed using a preset formula (0.5 x household size x 365) as every household will inevitably have at least some amount of non-

zero mileage consumption (walking from the parking lot, jogging etc.). The combined annual mileage is given as input to the MDCEV model, which will then predict the fleet mix owned by the household and allocate the household's mileage budget to all the vehicles owned by the household.

2. *Vehicle Fleet Mix Model*: The MDCEV model of fleet mix predicts the fleet composition of the households. Fleet mix for the MDCEV model is defined as a cross classification between 4 body types (car, van, SUV and pick-up truck) and 3 vintage categories (0-5 years old, 6-11 years old and ≥ 12 years old). Motorbike is added as an alternative with no vintage categories. An additional alternative called the 'non-motorized vehicle' is added to which non-motorized mileage of a household is allocated. MDCEV model is estimated and applied in such a fashion that every household in the dataset will consume at least some non-motorized mileage. Such an alternative is termed as an 'outside good' in econometric modeling jargon. The MDCEV model produces a different output each time a simulation is run. Which of the simulations should be considered final? To answer this question, the MDCEV model is applied on the data multiple times and mileage consumptions from each simulation are stored. After 'n' simulations of the MDCEV model are completed, the average mileage consumption is computed for each alternative, which will then be re-allocated using a mileage re-allocation algorithm, as described below.

Table 3.1 Average MDCEV Model Output After 50 Iterations

Vehicle Alternative	Mileage
Car (Age > 0 & ≤ 5 Years)	5000
Car (Age > 6 & ≤ 11 Years)	8000
Car (Age ≥ 12 Years)	400
Van (Age > 0 & ≤ 5 Years)	2000
Van (Age > 6 & ≤ 11 Years)	7000
Van (Age ≥ 12 Years)	550
SUV (Age > 0 & ≤ 5 Years)	1100
SUV (Age > 6 & ≤ 11 Years)	1000
SUV (Age ≥ 12 Years)	300
Pick-up (Age > 0 & ≤ 5 Years)	200
Pick-up (Age > 6 & ≤ 11 Years)	500
Pick-up (Age ≥ 12 Years)	100
Motorbike	50
Total	***26200***

3. *Number of Vehicle Alternatives Model*: A sample output for a household from 50 simulations of the MDCEV model is shown in Table 3.1. Since each simulation gives a slightly different result, the average mileage consumption result from 50 runs shows that the household owns almost all of the vehicle categories, whereas in reality this household might own only a couple of vehicles. The heuristic mileage reallocation algorithm does the job of reallocating this mileage distribution in such a fashion that it reflects the household vehicle fleet composition. But mileage reallocation algorithm

requires information about how many categories of vehicles does the household own. A multinomial logit model of number of alternatives predicts this information and provides it as an input to the mileage reallocation algorithm. Suppose, the household owns a car 0-5 years old and a SUV 0-5 years old, and a Van 6-11 years old, the number of alternatives model predicts the number of alternatives owned by this household as three.

4. *Number of Vehicle Body-types Model*: The structure of the number of vehicle body-types model is quite similar to that of the number of alternatives model, except this model predicts the number of different vehicle body types owned by a household, which provides marginal control totals for the mileage re-allocation model. While the vehicle body type distribution for the population is known in the base year (from survey data), this distribution is unknown for future years. The MNL model of vehicle body types predicts this distribution based on the projected synthetic population characteristics. This goes in as a control distribution that should be matched by the mileage re-allocation algorithm.

5. *Heuristic Mileage Re-allocation (HMR) Algorithm*: The heuristic mileage re-allocation algorithm (HMR) takes outputs of MDCEV model and MNL model of number of alternatives as input, to re-distributes the mileage to number of alternatives owned by the household. The logic followed by the HMR algorithm is shown in Figure 3.2. The algorithm operates at the level of each household, where it reallocates the mileage using a choice occasion based approach. The output from MNL model of number of alternatives provides information regarding how many body type-age categories does the household own.

From the output of MDCEV model, cumulative mileage distribution of the household is computed. A random number is generated and based on location of the random number in the cumulative mileage distribution of the household a vehicle is selected as owned by the household. The selected alternative is removed from the dataset thereby eliminating the possibility of choosing the same alternative multiple times. This process is carried out in a loop as dictated by the number of alternatives model. At the end of the loop, the HMR algorithm would select all the alternatives owned by the household. The mileage consumed by these alternatives is scaled up proportionally to account for the annual motorized mileage consumption of the household. Once the HMR algorithm reallocates the mileages for all households according to the input provided by number of alternatives model, the predicted body type distribution of the entire population is compared against the body type distribution predicted by the MNL model of number of vehicle body types. The absolute percent difference is computed between both the distributions and checked against a pre-set tolerance limit. If the HMR algorithm passes the tolerance check, the output of HMR algorithm goes in as input to the count models. If not, the entire application is repeated after calibrating the model components as warranted.

This process is carried out repeatedly until the percent difference between the two distributions is within a set tolerance limit. The output of HMR algorithm provides the final fleet composition of every household in the dataset. The output of this algorithm would have successfully predicted the vehicle ownership of the household, body type and vintage composition of the vehicles owned. Sample output of the HMR algorithm for a household who owns 3 vehicle alternatives is shown in Table 3.2. This output goes as input to the count models.

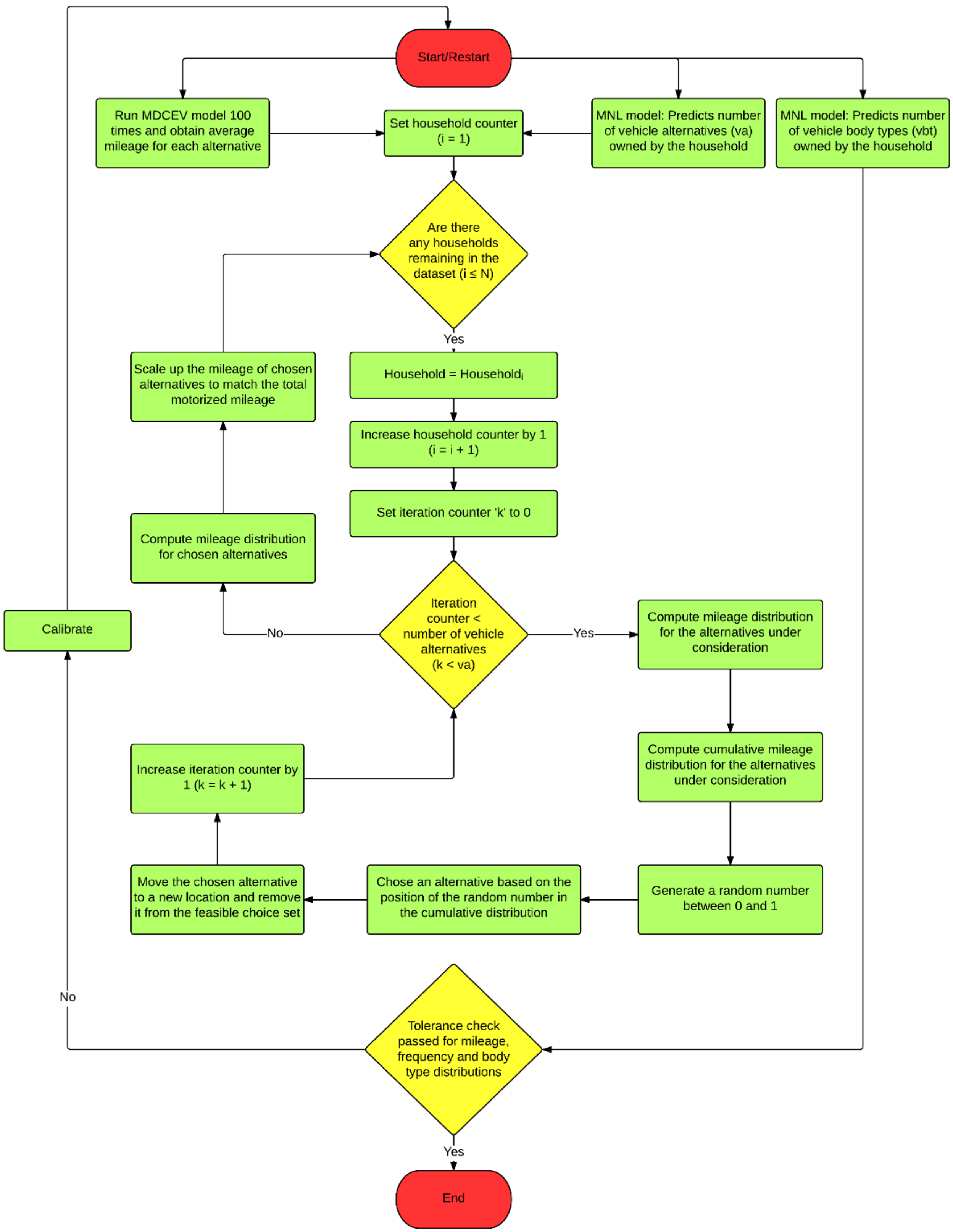


Figure 3.2 Heuristic Mileage Reallocation (HMR) Algorithm

Table 3.2 Output after HMR Algorithm

Vehicle Alternative	Mileage
Car (Age > 0 & ≤ 5 Years)	10000
Car (Age > 6 & ≤ 11 Years)	0
Car (Age ≥ 12 Years)	0
Van (Age > 0 & ≤ 5 Years)	0
Van (Age > 6 & ≤ 11 Years)	14000
Van (Age ≥ 12 Years)	0
SUV (Age > 0 & ≤ 5 Years)	2200
SUV (Age > 6 & ≤ 11 Years)	0
SUV (Age ≥ 12 Years)	0
Pick-up (Age > 0 & ≤ 5 Years)	0
Pick-up (Age > 6 & ≤ 11 Years)	0
Pick-up (Age ≥ 12 Years)	0
Motorbike	0
Total	***26200***

6. *Count Models*: Once the HMR algorithm re-distributes the mileage consumptions for all households such that they satisfy the marginal distributions provided by the body the distribution model, count models are applied for each household. The count models determine if all the mileage consumed by a household with a particular alternative belongs to one or multiple vehicles. Suppose, the output of HMR algorithm determines that a household uses a car 0-5 years old to travel 25000 miles annually, the count model determines if all of this mileage is put on just one car 0-5 years old or if the household owns multiple cars of 0-5 years of age. Ideally, a count model should be estimated for each of the 13 different vehicle categories defined for the MDCEV model, but this might make the model system vulnerable because of too many components. So, it was felt prudent to estimate one count model for each of the vehicle body types, with vintage serving as an explanatory variable in the models. If the household has non-zero mileage consumption in any of the vintages of a vehicle body type, count model of that particular body type is applied for that household.

At the end of application of the entire model system, an estimate of fleet composition of the household including body type, age and count of vehicles of each vehicle body type-age category is known along with their annual usage. Knowing the exact fleet composition is the first step toward accurate prediction of emissions. Each of the model components in the vehicle fleet composition model system are validated to test their predictive capability. Once each of the models are validated/calibrated to replicate observed distributions well, the model system is applied in its entirety to the data to see how well the model system as a whole would predict the observed fleet mix for the base year.

The MDCEV model of vehicle fleet mix is a comprehensive model that includes attributes at the household and zonal level. In addition to this, accessibility measures will be computed for each zone to conduct sensitivity analysis to test changes in fleet mix with varying zonal accessibility. The hypothesis is that increasing zonal accessibility will propel lower vehicle ownership levels and decrease pollution.

3.2 Model Estimation and Validation

This section presents the estimation and validation results of a vehicle fleet composition simulator that can be integrated into any existing activity-based microsimulation model systems. First, description of the data used for model estimation is provided. This is followed by model estimation results coupled with a sample replication result from a sequential application process of the model system. The process is continuous, in the sense that output of each component serves as input to the subsequent component in the model system. The logic followed by the fleet composition model system is discussed in detail.

3.2.1 Data for Model Estimation

Data used for estimating various components of the vehicle fleet composition model system is from the latest wave of National Household Travel Survey (NHTS) conducted in the year 2008-2009. NHTS collects data regarding socio-economic, demographic, vehicle ownership and personal travel characteristics of a random sample of households across the nation. Data collected from the survey is organized into four different files namely:

- *Household File*: Contains information regarding the household level socio-demographic characteristics such as household size, income, vehicle ownership, presence of children etc.
- *Person File*: Contains information regarding person level characteristics such as age, gender, worker status, driver status, etc. Each respondent from the household has a separate entry in this file. All the respondents in a household are grouped by the same household id.
- *Trip File*: This file has information regarding all trips made by a person in the day. Each trip made by the person gets a separate line entry grouped by the same person id. Trip level characteristics such a trip purpose, length, duration etc. are stored in this file.
- *Vehicle File*: This file has information regarding each vehicle owned by a household. Information regarding year, make, model etc. are collected.

In addition to data collected for each state at the national level, metropolitan planning organizations (MPOs) have the opportunity to purchase add-on samples for model development purposes. This effort uses the MAG add-on sample from 2008-09 NHTS for model estimation purposes. The vehicle fleet composition model system operates at the household level. Hence, the household and vehicle files are predominantly used for estimating components of the model system. A brief sketch of the household level socio-demographics of the data set is provided in Table 3.3.

From the table, it can be observed that the average number of vehicles owned by a household is about the same as average number of drivers in a household. This tells us that the data set under consideration is quite mobile and that households indeed own multiple vehicles. The intent of this effort is to explicitly identify the body type, age and annual mileage consumption of each of the vehicles owned by a household in the dataset. Majority of households in the dataset reside in urban areas and in single family dwelling units. The income distribution of the dataset is uniform, with slightly higher representation of medium income households. This lines up with the income profile of the data collected for the entire nation (National Household Travel Survey, 2009).

Table 3.3 Data Description: Household Level

Characteristic	Mean	Standard Deviation
Number of vehicles in the household	1.95	1.054
Number of persons in the household	2.43	1.333
Number of adults in the household	1.90	0.708
Number of children in the household	0.53	1.016
Number of workers in the household	0.97	0.889
Number of drivers in the household	1.83	0.771
Population density (sq miles)	4401.23	2557.65
Employment density (sq miles)	1164.92	1548.96
% of Households residing in urban area	83.80%	0.369
% Single family housing units	95.80%	0.201
% Households with income < $25,000	17.80%	0.383
% Households with income ≥ $25,000 & < $50,000	28.40%	0.451
% Households with income ≥ $50,000 & < $75,000	18.80%	0.391
% Households with income ≥ $75,000 & <$100,000	15.00%	0.357
% Households with income > $100,000	20.00%	0.4
Sample Size, N	***4,262 Households***	

For the purposes of this effort, vehicles from the NHTS data are categorized by a cross classification between four body types (car, van, sports utility vehicle (SUV), pick-up truck) and three vintages (0-5 years old, 6-11 years old, ≥ 12 years). A motorbike category is also considered (with no vintage classification) bringing the total number of motorized alternatives to thirteen. In addition to the motorized alternatives, a non-motorized vehicle alternative was considered to capture the walk/bike travel undertaken by each household in the dataset. This will allow for modeling the total annual mileage consumption for a household irrespective of the type of mode used for travel. While the NHTS data has information regarding the estimated annual mileage of each vehicle owned by a household, non-motorized mileage is not readily available. Annual non-motorized mileage is computed from the walk and bike trips reported by all individuals in a household.

The non-motorized alternative is the one that is consumed by every household in the dataset and is considered as an outside good. An outside good is an alternative that is chosen by every choice maker in the dataset in the econometric modeling perspective. In the current context, every household invariably undertakes some amount of non-motorized travel such as walking from the parking lot, walking to the bus station or jogging etc. and hence this alternative is considered as an outside good. To compute the annual non-motorized mileage of a household, (weighted) walk/bike trips reported by all of the household members are aggregated. If none of the household members reported walk/bike trips, annual non-motorized mileage for that household is computed as '0.5 (miles/person/day) x 365 (days/year) x number of persons in the household'. Previous studies have successfully incorporated this formulation in a similar context (Vyas et al., 2012). In total the model system consists of 14 alternatives (4 body types x 3 vintage categories + motorbike + non-motorized mileage).

An alternate vintage classification was also tested (0-3 years old, 4-9 years old, ≥ 10 years) and it was found that the fleet composition model system is robust to the vintage classification considered. The vintage classification (0-5 years old, 6-11 years old, ≥ 12 years) was finalized based on the observation that most car manufactures offer a five year power train warranty (My Car Stats, 2010). Also, this classification provided a healthy sample size for all the 13 motorized alternatives considered for model estimation. The vintage classification can be further disaggregated to include an alternative for each year for a vehicle body type (bringing the total number of vehicle alternatives to 50), but this level disaggregation would make the dataset sparse for model estimation and also increase the computation burden in model application process. Table 3.4 provides a description of vehicle fleet characteristics of the NHTS dataset considered for this effort.

Table 3.4 Data Description: Vehicle Level

Panel A: Vehicle Body Type					
	Car	**Van**	**SUV**	**Pick-up**	**Motor Bike**
Average Age	8.55	7.46	6.52	9.52	9.21
Average Mileage	10204.4	11317.7	11296.6	10723.0	3838.9
Number of Vehicles	3,997	635	1,537	1,376	240
Panel B: Vehicle Body Type vs. Annual Mileage					
Annual Mileage					
0 - 4,999	27.5%	18.4%	21.1%	24.9%	71.3%
5,000 - 9,999	30.6%	31.3%	28.6%	29.4%	15.8%
10,000 - 14,999	21.4%	26.9%	26.2%	22.8%	7.9%
15,000 - 19,999	11.3%	13.9%	12.8%	12.5%	2.9%
≥ 20,000	9.1%	9.4%	11.3%	10.5%	2.1%
Total	100.0%	100.0%	100.0%	100.0%	100.0%

From Panel A of the table, it can be observed that households prefer relatively newer SUVs and older pick-up trucks consistent with expectation. This finding is corroborated by the body type and age distribution shown in Figure 3.3, where it can be observed that more than half of the SUVs are in the 'newer' vehicle category, whereas pick-ups have relatively lower representation in this category.

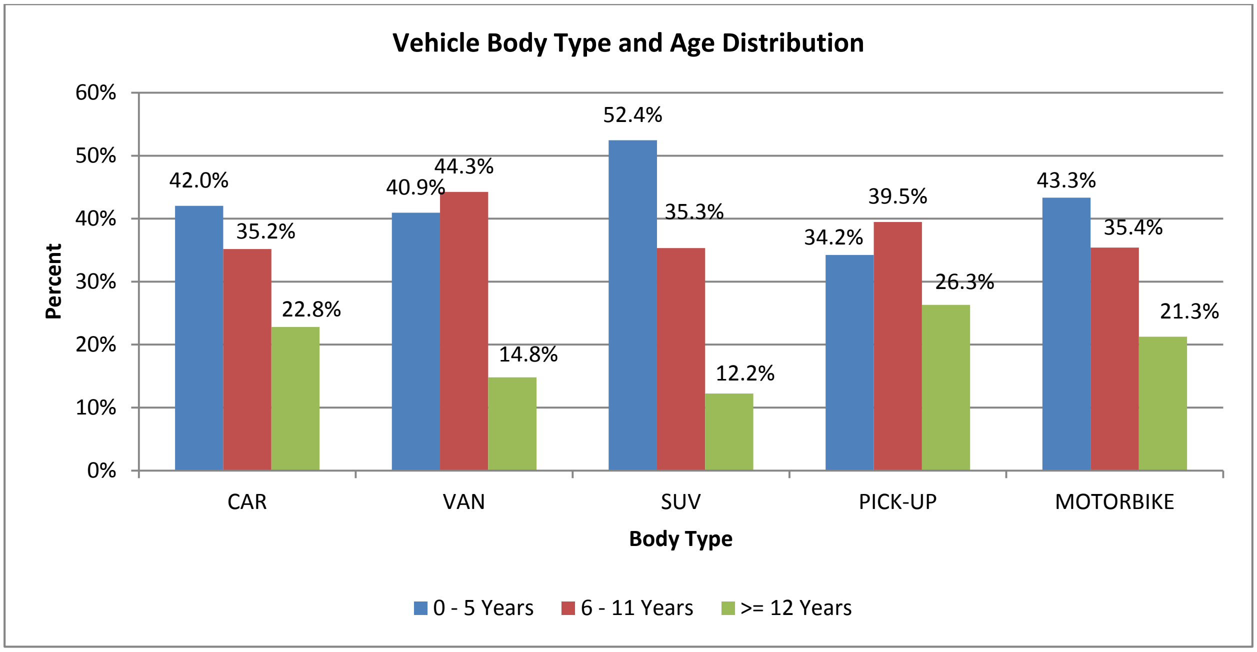


Figure 3.3 Vehicle Body Type and Age Distribution

Table 3.5 provides the distribution of vehicle body types for each household income category. It can be observed that, while lowest (< $25,000) and low income ($25,000 - $49,999) households tend to own more cars, medium and high income household tend to own a mix of vehicles. It can also be observed that with increasing household income, the ownership of SUVs gradually increases. One possible reason for this might that high income households can afford sports utility vehicles more. Another reason could be that while low income households usually own fewer vehicles and utilize them for all travel needs, households with high income might own a mix of vehicles and use them to varying degrees for specific purposes (refers to a combination of affordability and variety seeking nature of the segment). The total number of vehicles owned by individuals from each income category is shown in parenthesis in the first column of the table.

Table 3.5 Vehicle Body Type Distribution by Household Income

Household Income	Vehicle Body Type					Total
	CAR	VAN	SUV	PICK-UP	MOTORBIKE	
< $25,000 (852)	61.3%	9.5%	11.6%	16.7%	0.9%	100.0%
≥ $25,000 and < $50,000 (1,898)	54.9%	9.0%	14.6%	19.0%	2.5%	100.0%
≥ $50,000 and < $75,000 (1,547)	48.8%	7.6%	21.4%	18.3%	3.9%	100.0%
≥ $75,000 and < $100,000 (1,418)	48.0%	8.5%	20.1%	19.5%	3.9%	100.0%
≥ $100,000 (2,070)	48.2%	7.1%	26.3%	15.2%	3.3%	100.0%

3.2.2 Model Estimation and Validation Results

This section provides the estimation results of all the components in the vehicle fleet composition model system coupled with comparisons between observed and predicted patterns from a sample replication exercise. It should be identified that the process adopted here does not constitute a true validation exercise. In the traditional validation process, the data would be split (say in the proportion of 80:20) and the larger sample is used for model estimation. The estimated models are applied on the holdout sample to test the predictive capability of the model. In the current context, the number of different components included in the model system and level of disaggregation of vehicle alternatives warranted the use of entire survey sample (4,262 households/7,785 vehicles) for model estimation. The estimated models are applied to the entire survey sample to compare predicted patterns against the observed patterns in the data. In order to ensure the efficacy of the model system, the model was tested on specific market segments (different income categories, urban/rural residents) and the predictive performance of different model components was tested against observed data. A detailed sensitivity analysis exercise was carried out to predict the effect of changes in land use dynamics on vehicle fleet composition patterns.

Each of the model components was estimated and validated separately to ensure the predictive capability of the models in replicating observed vehicle ownership patterns. The model system is then applied in its entirety to the estimation dataset to test its efficacy. Estimation results as well as results of the sequential model application process are provided here.

3.2.2.1 Motorized Mileage Prediction Model

The first element in the vehicle fleet composition model system is the household mileage prediction model that predicts the annual motorized mileage consumption of households. The motorized mileage is estimated using a power transformed linear regression model. Once the motorized mileage for each household is predicted, non-motorized mileage is computed using a preset formula (0.5 x household size x 365) as every household will inevitably have at least some amount of non-zero mileage consumption. The combined annual mileage is provided as input to the MDCEV model, which will then predict the fleet mix owned by the household and allocate the mileage budget to all the vehicles owned by the household.

To fit the observed annual motorized mileage distribution, several model structures were explored and the power transformed linear regression model fit the data best. In practice, the activity-based model to which the vehicle fleet composition simulator is integrated will provide the annual mileage budget as an input. Since this is a standalone model application process, a separate mileage prediction model is estimated. Use of a power transformed linear regression model avoids the possibility of negative mileage predictions that a regular linear regression model may provide. The model structure of mileage prediction model is shown in the following equation:

$$Motorized\ Mileage^{0.3} = \beta_0 + \beta_i' x_i + \varepsilon$$

Where, β_0 is a constant, β_i' is the array of coefficients to be estimated and x_i is the array of socio-demographic characteristics included in the model. The error term 'ε' is normally distributed with mean zero and standard deviation of the dependent variable.

Estimation results of the mileage prediction model are presented in Table 3.6. Various socio-demographic characteristics, lifestyle variables and TAZ characteristics of the household's residential location were used to estimate the motorized mileage consumption patterns. Household income was observed to be a significant variable in explaining the annual motorized mileage consumption. Households in the lowest income category are likely to have low motorized mileage consumptions, while households in the highest income category are likely to have higher motorized mileage consumptions. This finding is directly related to the number of vehicles owned by respective income categories (presented in Table 3.5) where it was seen that lowest income category households own approximately 10% (852 vehicles) of the vehicles in the dataset, while highest income households own about 26% (2,070) of the vehicles. This relates to the proportional higher mileage consumption of highest income households.

Households with more number of drivers were observed to consume higher mileages, an observation consistent with expectation. Similarly, household with more number of children had higher mileage consumptions. Possible reason for this might be due to chauffeuring associated with children's activities in such households. Retired households with no children tended to have lesser mileage consumption. This observation is behaviorally intuitive as such household might not engage in a lot of activity. Households residing in TAZs that had higher proportion of affluent households tended to have higher motorized mileage consumptions. This finding couples nicely with the higher mileage consumptions for the highest income (≥ $100,000) category. Households residing in TAZs that have a lot of employment accessibility within 10 minutes of auto travel have lesser motorized mileage consumptions. These TAZs probably refer to locations in the urban core, with mixed-use development where discretionary travel can be easily undertaken by non-motorized modes (walk/bike). Residential self-selection also has a possible role to play in this finding as people residing the urban core might be more environment friendly and are willing to opt out of motorized modes of travel.

Table 3.6 Motorized Mileage Prediction Model: Estimation Results

Explanatory Variable	Coefficient	t-statistic
Constant	13.03	35.00
Number of drivers in household	2.01	12.80
Count of adult household members at least 18 years old	0.37	2.28
Household resides in rural area	0.87	5.30
Lowest income household (< $25,000)	-1.05	-6.03
Highest income household (≥ $100,000)	1.33	7.13
Number of children in the household	0.52	5.04
Zero worker household	-1.51	-8.78
Two worker household	0.83	5.52
Household size = 4 or more	-0.67	-2.48
Single family housing unit (owned)	0.57	3.14
Retired household (one/two person) with no children	-0.88	-5.15
Proportion of households in the highest income quintile	1.36	2.95
Proportion of single family housing units in the TAZ	0.69	1.99
TAZ with high regional employment accessible within 10 minutes by auto (1st Quartile)	-0.29	-2.20
R^2	0.404	

The model is applied on the estimation dataset to see how well it could replicate the observed mileage consumption patterns. Results of this comparison are presented in Figure 3.4. The model is able to replicate the observed patterns quite well. Results shown are for the calibrated model where the constant in the regression equation was slightly adjusted to better match the observed patterns. Each bar in the chart represents the percentage of households in the data set that pertain to the mileage bin under consideration.

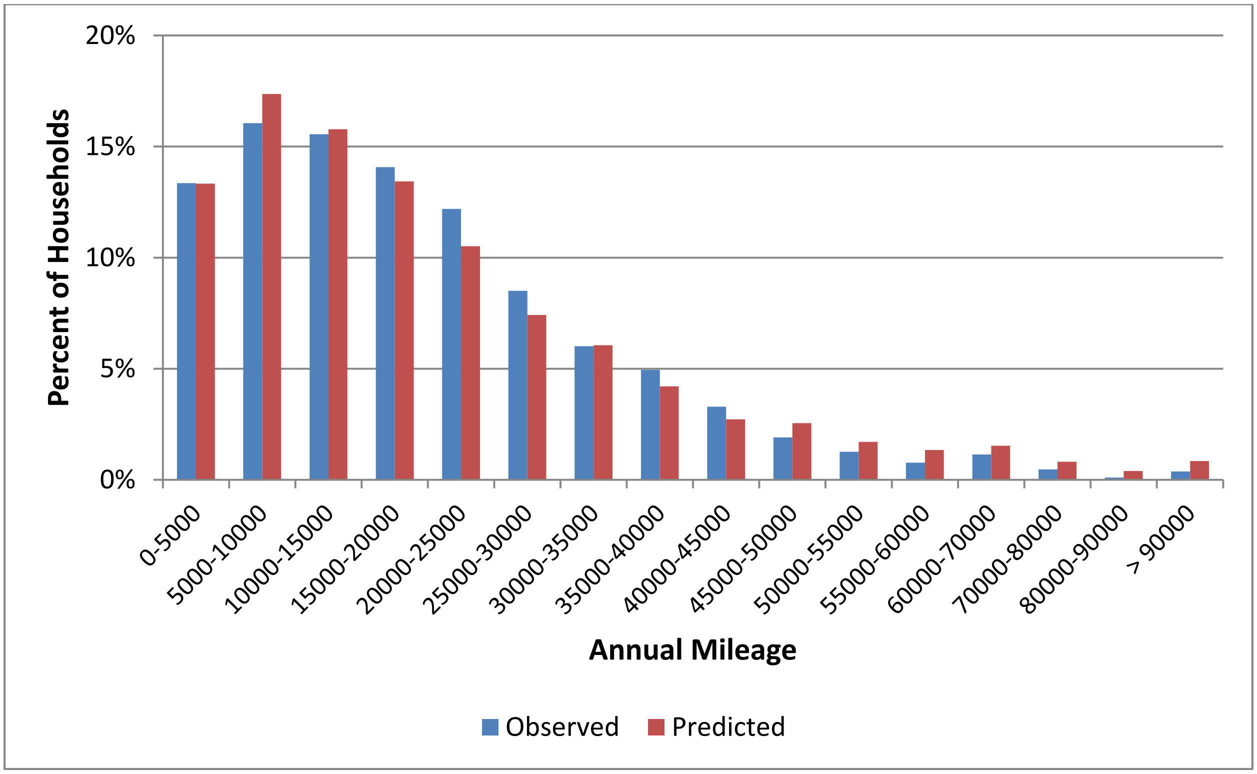


Figure 3.4 Observed versus Predicted Mileage Distributions

3.2.2.2 MDCEV Model of Vehicle Fleet Mix

The next component in the model system is the MDCEV model of vehicle fleet mix which takes the mileage predicted by the previous component as input, predicts the vehicle fleet mix owned by the household and allocates the mileage budget to different vehicles owned by the household. The fleet mix model system is an MDCEV model which is capable of simultaneously predicting the array of vehicles owned by a household. The MDCEV model is ideally suited to model the vehicle fleet mix and utilization patterns due to the multiple discrete (ownership of multiple vehicles) and continuous (mileage allocation to vehicles owned) nature of the problem. The MDCEV model was proposed by Bhat (2005; 2008) to efficiently model multiple discrete choice behavior, addressing the short shortcomings of single discrete choice models. A number of recent studies used the MDCEV model to estimate vehicle fleet mix at the household level (Bhat and Sen, 2006; Eluru et al., 2010; Vyas et al., 2012). Notable features of the MDCEV model include consideration of diminishing marginal utility with increasing consumption of an alternative and its capability to collapse to the standard MNL model structure, given every behavioral unit in the dataset chooses only one out of 'k' available alternatives.

As discussed earlier, the vehicle classification for the fleet mix model system consists of a total of 14 alternatives (4 vehicle body types x 3 vintage categories + motorbike + non-motorized alternative). In order to account for household with no vehicles at all, the MDCEV model specification with presence of an outside good is adopted in the current empirical context. An outside good is an alternative that is chosen by every household in the data set, which in this case would be the non-motorized alternative. After the mileage prediction model predicts the motorized mileage consumption of the household, non-motorized mileage is computed using a preset formula (0.5 x household size x 365) and added to the motorized mileage to determine the 'total' mileage consumption of the household. The MDCEV model takes the total mileage consumption of the household as input and distributes it to different vehicles owned (as predicted by the model) by the household. The formulation of the MDCEV model allows for selection of 'm' alternatives out of 'k' available alternatives, while definitely choosing the outside good (non-motorized alternative) for each and every household in the dataset. The functional form of the utility expression of the MDCEV model proposed by Bhat (2008) for a case with the presence of an outside good is:

$$U(x) = \frac{1}{\alpha_{out}} \psi_{out} x_{out}^{\alpha_{out}} + \sum_{k=2}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left\{ \left(\frac{x_k}{\gamma_k} + 1 \right) - 1 \right\}$$

Where baseline marginal utility for the outside good, $\psi_{out} = \exp(\varepsilon_{out})$ and baseline marginal utility for the rest of the alternatives $\psi_k = \exp(\beta' z_k + \varepsilon_k)$, is a function of various parameters that capture the observed and unobserved attributes of the alternative. z_k is a set of attributes that define and alternative 'k' and ε_k captures the effect of unobserved attributes. $U(x)$ is a quasi-concave and continuously differentiable function with respect to consumption quantity vector x ($x_k \geq 0 \; \forall \; k$). ψ_k represents the baseline marginal utility or the marginal utility at the point of zero consumption. α_k is the satiation parameter which governs the decrease in marginal utility with increasing consumption for good k. The translation parameter γ_k not only governs the level of satiation but also enables corner solutions (i.e., zero consumption of some goods).

As both γ_k and α_k are parameters that incorporate the effects of satiation, it is difficult to uniquely identify and distinguish between them. For this reason, one of the two parameters is fixed and the other parameter is free to be estimated in most empirical model estimation efforts and the best model is chosen. In the current modeling context γ-profile gave the best fit to the data. A household maximizes its utility by optimally allocating consumptions to the k available goods (vehicles), while always choosing the outside good (non-motorized alternative). Thus the constraint for the utility maximization problem is:

$$\sum_{k}^{K} t_k = T$$

The MDCEV model is estimated with 14 alternatives and the estimation results of the MDCEV model system are presented in Tables 3.7-3.10. Tables 3.7-3-8 present the significant parameters in the baseline marginal utility equation of the MDCEV model. From the model estimation result, it was found that high income households are more likely to own newer vehicles and also tend to prefer cars and SUVs over other types of vehicles. Households with children tend to own vans more than cars. It can be observed that number of children in the household has a negative impact on owing cars, meaning such households would rather prefer a vehicle that would help them attend to the child's necessities (such as a van).

Households with more number of workers tend to prefer newer cars and SUV vehicles, which is intuitive as the number of workers in a household could probably act as a proxy characteristic for affluence of the household. While the impetus for owing vans is explained best by the presence/number of children in the household, the ownership patterns for car, van and SUV body types are explained very well by the income categorization. Within the car body type, high income households prefer to own newer cars (0-5 years), low income households tend to own medium aged cars (6-11 years) while the lowest income households are more likely to own older cars (≥ 12 years). Similar patterns are observed in the van and SUV categories as well. Thus the model is able to represent the vehicle ownership patterns of different income categories, where high income households usually change their vehicle fleet more often but less affluent households do not have such flexibility. Lager households prefer to own vans as they offer the flexibility of accommodating greater number people, which facilitates joint travel in such households. Households living in rural areas are more likely to own pick-up trucks, another finding consistent with expectation.

Table 3.7 MDCEV Model Estimation Results: Cars and Vans

Vehicle Type	Explanatory Variable	Coefficient	t-statistic
Car 0-5 years old	High income household ($75,000 - $99,999)	0.16	2.16
	Number of children in the household	-0.19	-5.91
	Three or more worker household	0.17	1.38
	Proportion of households in the lowest income quintile	-1.01	-3.61
	Percent of regional employment within 10 minutes of auto accessibility from the TAZ	13.83	-3.24
Car 6-11 years old	Two worker household	-0.16	-2.21
	Low income household ($25,000 - $49,999)	0.13	1.9
Car 12 years or older	Lowest income household (< $25,000)	0.57	5.84
	Household has one/two retired adults and no children	0.2	2.59
	Proportion of households in the lowest income quintile	0.57	1.74
Van 0-5 years old	Number of children in the household	0.38	8.23
	Two worker household	-0.39	-2.53
	TAZ with high density (1st Quartile)	-0.31	-1.89
	Percent of regional employment within 30 minutes of auto accessibility from the TAZ	-1.72	-1.78
Van 6-11 years old	Number of children in the household	0.33	7.15
	TAZ with high density (1st Quartile)	-0.26	-1.87
	Low income household ($25,000 - $49,999)	0.27	1.87
Van 12 years or older	Count of HH members	0.14	1.98
	TAZ with high density (1st Quartile)	0.68	3.1
	Lowest income household (< $25,000)	0.66	2.57

Table 3.8 MDCEV Model Estimation Results: SUVs, Pick-ups and Motorbikes

Vehicle Type	Explanatory Variable	Coefficient	t-statistic
SUV 0-5 years old	Lowest income household (< $25,000)	-1.01	-5.59
	Two worker household	0.17	1.92
	Household has one/two retired adults and no children	-0.17	-1.78
	Proportion of households in the lowest income quintile	-1.48	-3.85
	Percent of regional employment within 10 minutes of auto accessibility from the TAZ	-17.83	-2.82
	TAZ with low density (3rd Quartile)	0.27	2.48
SUV 6-11 years old	Medium income household ($50,000 - $74,999)	0.26	2.41
	Household size = 4 or more	0.33	3.3
	Single family housing unit (owned)	0.86	4.37
	TAZ with high regional employment accessible within 30 minutes by auto (1st Quartile)	-0.30	-2.85
SUV 12 years or older	High income household ($75,000 - $99,999)	0.36	1.89
	Presence of children in the household	0.31	1.99
	Household in a single family housing unit	-0.74	-2.38
	TAZ with medium density (2nd Quartile)	-0.36	-2.28
Pick-up 0-5 years old	Highest income household (≥ $100,000)	0.24	2.28
	Household size = 1	-0.98	-4.64
	Household resides in rural area (from variable URBRUR)	0.24	1.95
	Proportion of single family housing units in the TAZ	0.74	2.44
Pick-up 6-11 years old	Household resides in rural area	0.15	1.3
	Household has one/two retired adults and no children	-0.35	-3.41
	High income household ($75,000 - $99,999)	0.16	1.44
	TAZ with high regional employment accessible within 10 minutes by auto (1st Quartile)	-0.23	-2.17
Pick-up 12 years or older	Proportion of households in the lowest income quintile	1.28	2.95
	Low income household ($25,000 - $49,999)	0.35	2.86
	Presence of children in the household	-0.25	-1.95
	TAZ with high regional employment accessible within 10 minutes by auto (1st Quartile)	-0.34	-2.6
Motorbike	Household resides in rural area	0.71	4.41
	Single family housing unit (owned)	0.75	2.36
	Household size = 1	-0.57	-2.11

Among TAZ characteristics, it was found that households living in TAZs with high proportion of households in the lowest income quintile are not likely to own newer cars. This finding is corroborated by another observation from the model that households in such TAZs are more likely to own older cars. Households in TAZs with high density were less likely to own vans. It is possible that spatial and social dependency effects play a role in vehicle ownership and this finding is consistent with such a notion (Paleti et al., 2013). Households in TAZs with high density are less likely to own newer vehicles. Households in such TAZs might have accessibility to alternative modes of transportation and also use walk/bike to satisfy their mobility needs, which might in turn prompt them to just keep their older vehicles in the fleet mix. This finding is nicely coupled by the observation that households in lower density TAZs tend to own newer vehicles and are more likely to have larger vehicles in the fleet mix such as SUVs.

Table 3.9 presents the model estimation results of baseline constants and translation parameters in the MDCEV model. A baseline constant provides an indication of the inherent preferences for various alternatives and the marginal utility at zero consumption. The values of the baseline constant reveals the preference for a particular type of vehicle for an 'average' user in the dataset. For cars and vans, it was found that newer vehicles have a greater baseline utility than older ones, suggesting that households would rather own newer cars, all other things being equal. In general baseline utility decreases with age of the vehicle, although this trend is not seen consistently for SUVs and pick-up trucks. For SUVs , there is lower baseline utility for middle aged SUVs suggesting that households tend to acquire newer SUVs and hold on to their SUV for a long time, which is fairly expected behavior. For pick-up trucks, there is a lower baseline utility for newer pick-up trucks, suggesting that households hold on their middle aged and older pick-up trucks more and do not see a necessity to own the 'newest' pick-up truck. This finding is consistent with general behavior, where pick-up trucks have a very slow turnover rate. All of the findings from model estimation results line up well with actual vehicle ownership patterns observed in the dataset.

Table 3.9 MDCEV Model Estimation Results: Baseline Constants and Translation Parameters

Vehicle Type	Baseline Constants		Translation Parameters	
	Coefficient	t-statistic	Coefficient	t-statistic
Non-motorized vehicle (Outside Good)	NA	NA	0	NA
Car 0-5 years old	-5.98	-83.27	23668	10.07
Car 6-11 years old	-6.51	-140.27	18621	10.37
Car 12 years or older	-7.29	-93.28	12164	9.46
Van 0-5 years old	-8.13	-61.11	29431	3.41
Van 6-11 years old	-8.43	-82.8	22248	4.21
Van 12 years or older	-10.04	-41.51	12691	3.22
SUV 0-5 years old	-6.65	-64	25172	6.7
SUV 6-11 years old	-8.32	-42.25	16717	6.71
SUV 12 years or older	-7.88	-25.94	8397	5.1
Pick-up 0-5 years old	-8.29	-30.87	20610	5.69
Pick-up 6-11 years old	-7.35	-95.28	14758	6.92
Pick-up 12 years or older	-8.06	-70.37	9542	6.7
Motorbike	-9.24	-29.1	2223	7.67

Translation parameters in the MDCEV model represent the diminishing marginal returns with increasing consumption of an alternative. A higher value for the translation parameter pertaining to a specific vehicle means that households are less satiated with the use of that vehicle and are likely to drive that vehicle alternative more. For all of the vehicle body types, the translation parameters show a consistent pattern where newer vehicles have a higher translation parameter than an older vehicle in the same body type. This finding is behaviorally intuitive as households usually tend to drive newer vehicles more than the older ones. Amongst all the alternatives, new vans have the highest translation parameter. A detailed exploration of the data revealed that households owning vans use these as multipurpose vehicles to meet the regular household travel necessities as well as the chauffeuring needs of the children. Motorbikes have the lowest translation parameter, which is expected as motorbikes are used mostly for pleasure/hobby travel but not as a primary vehicle in the household. Table 3.10 shows the goodness of fit statistics of the estimated model. The likelihood ratio of the estimated model is 645.36 which is substantially greater than the critical χ^2 value with 50 degrees of freedom at 99% level of confidence.

Table 3.10 MDCEV Model Estimation Results: Goodness of Fit Measures

Statistic	Value
Log-likelihood of final model at convergence	-77020.49
Degrees of freedom of final model	75
Log-likelihood of base model at convergence	-77343.17
Degrees of freedom of base model	25
Likelihood ratio	645.36
$\chi^2_{50,0.001}$	86.66

The estimated MDCEV model is applied on the entire data see how well the model can predict observed fleet composition patterns. Gauss codes made available by Pinjari and Bhat (2011), were translated to open source coding language 'R' to implement the MDCEV forecasting procedure. The model results were compared to match the following observed patterns:

- Average annual mileage excluding zero mileage households: Average mileage is computed as the total mileage of each alternative divided by total number of households that have non-zero mileage consumption for the alternative
- Vehicle type distribution: Frequency of vehicle ownership for each vehicle alternative is computed as the total number of households who own a particular type of vehicle, divided by the total number of households in the dataset
- Body type distribution: Vehicle body type distribution of the observed data is compared against the body type distribution derived from the output of MDCEV model. This is an important check that should be passed by the MDCEV model, in order to impart necessary confidence in the model specification to be used for predicting fleet composition for a given (future) horizon year data. The body type distribution is not a factor that is inherently modeled in the MDCEV model specification. If the model is able to accurately predict this uncontrolled distribution, it would instill required confidence in the forecasts done using this model for any future year.

In the proposed framework, MDCEV model simulation is carried out multiple times and mileage consumptions from each simulation are stored. After ‘n’ (say 100) simulations of the MDCEV model are completed, an average mileage consumption is computed for each alternative, which is then redistributed using a mileage reallocation algorithm. Since each simulation of the MDCEV model gives a slightly different result, the average mileage consumption result from ‘n’ MDCEV model runs show that a household owns almost all of the vehicle categories, whereas in reality the household might own only a subset of the vehicle categories considered by the MDCEV model of vehicle fleet mix. The heuristic mileage reallocation algorithm does the job of reallocating this mileage distribution in such a fashion that it reflects the household’s vehicle fleet composition. The mileage reallocation algorithm requires information about how many distinct categories of vehicles does the household own. MNL model of number of vehicle alternatives predicts this information and provides it as an input to the mileage reallocation algorithm. Suppose, a household owns a car 0-5 years old and a car 6-11 years old, and a van 6-11 years old, the number of alternatives model is supposed predict the number of alternatives owned by this household as three.

3.2.2.3 MNL Model of Number of Vehicle Alternatives

The purpose of this model is to provide input to the heuristic mileage reallocation algorithm regarding the number of distinct vehicle alternatives owned by a household. Since the choice phenomenon at hand is a single discrete choice case (every household owns a unique number of vehicle alternatives), an MNL model structure is opted. Ideally, this model should have a total of 14 alternatives in accordance with the number of motorized alternatives in the MDCEV model structure. It is not required to consider the non-motorized alternative for this model or any subsequent models after the MDCEV model, as this is an outside good that ‘should’ be consumed by every household and hence need not be modeled separately. Observations from the estimation dataset revealed that the maximum number of distinct alternatives that any household in the dataset own is five. So, an MNL model is estimated with six categories (0-4, ≥ 5 vehicle alternatives). The final category (≥ 5 vehicle alternatives) served as the base alternative. Model estimation results are presented in Table 3.11.

From the model results, it was observed that lowest income households are likely to own fewer vehicle alternatives, while medium and high income households tended to own multiple vehicle alternatives which is consistent with expectation. Higher income households in general have the financial flexibility to own a mix of vehicles to cater for specific purposes. Single person households are more likely to own fewer vehicles (zero/one) which is an intuitive finding. Households living in TAZs with high population density tended to own zero vehicles. This might represent the category of households, who self-select themselves into mixed urban use TAZs (environmentally proactive households). Larger households are found to own multiple vehicle alternatives, another finding consistent with expectation as such households usually sport a vehicle (such as van) for family travel in addition to a vehicle to cater for regular travel necessities. Similar behavior was found in households with children. Highest income (≥ \$100,000) households were found to own 4 vehicle alternatives, which is an intuitive observation. The likelihood ratio of the model is substantially higher than the critical χ^2 value at any reasonable level of significance.

Table 3.11 MNL Model of Number of Vehicle Alternatives

Number of Alternatives	Explanatory Variable	Coefficient	t-statistic
Zero	Constant	1.73	4.20
	Lowest income household (< $25,000)	2.37	9.70
	Low income household ($25,000 - $49,999)	0.82	3.23
	Housing unit owned (from variable HOMEOWN)	-1.65	-9.95
	Household size = 1	2.12	10.83
	Zero worker household	1.21	6.42
	Population density of the TAZ that the household resides	0.00011	4.16
One	Constant	4.27	13.72
	Lowest income household (< $25,000)	1.43	11.59
	Low income household ($25,000 - $49,999)	0.96	10.39
	Household size = 1	2.22	18.38
	Proportion of multi-family housing units in the TAZ	0.40	2.05
	Two worker household	-0.86	-6.37
Two	Constant	5.31	17.14
	Household with 2+ adults, youngest child 0-5	0.42	3.76
	Medium income household ($50,000 - $74,999)	-0.21	-2.29
	Two worker household	-0.34	-2.95
	Households in lowest income quintile (Q1)	-0.00034	-2.03
Three	Constant	1.69	3.36
	Housing unit owned (from variable HOMEOWN)	1.00	3.01
	Count of adult HH members at least 18 years old	0.48	5.40
	Three or more worker household	1.02	4.52
	Population density of the TAZ that the household resides	-0.000088	-3.49
	Presence of children in the household	0.38	3.25
	Household with 2+ adults, youngest child 16-21	0.50	2.50
Four	Constant	1.18	3.19
	Highest income household (>= $100,000)	0.72	3.00
	Household with 2+ adults, youngest child 16-21	1.63	5.47
	Household size = 4 or more	1.51	6.25
	Two worker household	-0.70	-2.59
Sample Size (Number of Households)		4,262	
ρ^2		0.469	
Adjusted ρ^2		0.468	
χ^2		2099.10	

Figure 3.5 shows the comparison of observed and predicted vehicle alternative distributions. The results shown are for the uncalibrated version of the model and it can be observed that the model replicates the observed patterns exceedingly well with no necessity for calibration.

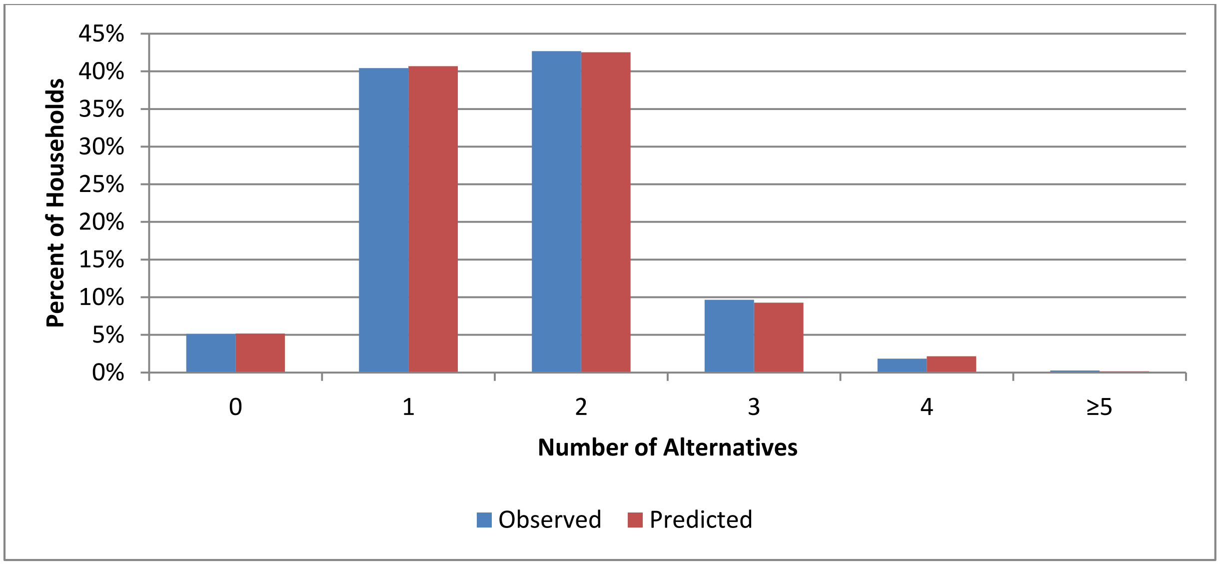


Figure 3.5 Observed versus Predicted Vehicle Alternative Distribution

In addition to comparing the overall fit of the model for the entire dataset, comparisons were made across different income categories to see how well the model could replicate the vehicle alternative distributions of different income segments. Results of this comparison are presented in Figure 3.6. Panels A-E present the comparisons by income category. Sample sizes for each income category are given on the upper right corner of the graph. The blue bars always represent the observed patterns and the orange ones show predicted distributions. From the comparison charts, it can be observed that the model estimated on aggregate data performs quite well in predicting distributions across different income categories. This signifies that the model specification is robust enough to represent the difference in vehicle ownership patterns across different income segments. A high proportion of lowest income households (< $25,000) correspond to the zero vehicle alternative category and with increase in income segment the proportion of households in this category slowly decreases with almost no households in the zero alternative category for highest income household segment. This finding is behaviorally intuitive, in the sense that as the household's income level increases, so does the financial flexibility to own more number of vehicles. It is quite heartening to see the model predict the same phenomenon.

Low income households majorly own one vehicle alternative (about 90%). Lower representation of medium through high income households in this category is supplemented their stronger presence in the higher vehicle alternative categories. For vehicle alternative categories 3 and 4, a gradually increasing representation can be observed from lowest to highest income households. This exceedingly good performance of the model to predict subtle nuances in the dataset imparts necessary confidence to use the output of this model as a governing distribution in the heuristic mileage reallocation algorithm.

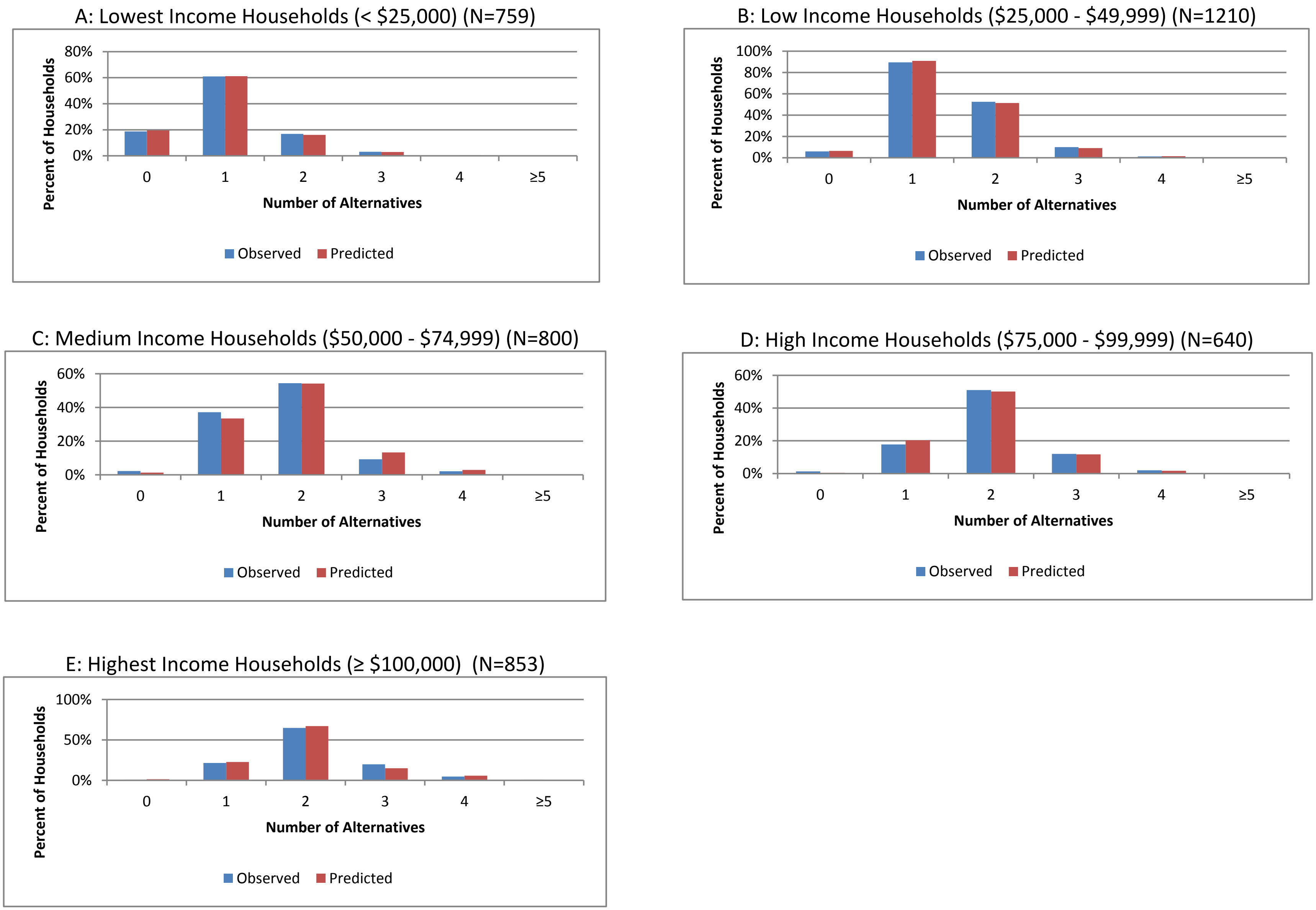


Figure 3.6 Observed versus Predicted Vehicle Alternative Distributions by Income Category

3.2.2.4 MNL Model of Number of Body Types

For every household in the dataset, the heuristic mileage reallocation algorithm takes the output of the MDCEV model and redistributes the mileage as governed by the number of vehicle alternatives model. After this process is carried out for every household in the dataset, comparisons are made across observed and predicted distributions of average annual mileage, vehicle type and body type. While the data to calibrate the model on these three grounds is readily available for base year, how can one be sure that the predicted vehicle fleet composition distributions are representation of the actual fleet composition for a future year input data? To ensure this consistency, a vehicle body type MNL model is estimated and calibrated for the base year. The model predicts this distribution based on the projected synthetic population characteristics for any horizon year. This goes in as a control distribution that should be matched by the heuristic mileage reallocation algorithm.

The structure of the MNL model of number of body types is very similar to that of the previous model, except in this case the number of distinct vehicle body types owned by a household are modeled instead of the number of distinct vehicle alternatives. Going of the example from previous section, if a household owns a car 0-5 years old and a car 6-11 years old, and a van 6-11 year old, the household is said to own a total of 2 vehicle body types (car and van). There are a total of 6 body types considered in the context of the current research effort namely car, van, SUV, pick-up, motorbike and non-motorized alternative. The MNL model should ideally include 6 alternatives, but observations from the estimation dataset revealed that the maximum number of body types owned by any household in the dataset is 5, with very few households owning 4 or more vehicle body types. So, the vehicle body type count is truncated at 4, thereby providing a total of 5 alternatives (0-3, ≥ 4 vehicle body types). The final category (≥ 4 vehicle body types) served as the base alternative. Model estimation results are presented in Table 3.12.

From the model results, it was found that lowest and low income households are most likely to own zero or one vehicle body type. This finding traces back to the question of affordability of multiple vehicle types for this segment. Also, this finding couples nicely with the result of the MNL model of number of vehicle alternatives where households in this segment appeared in the zero and one vehicle alternative category, which automatically positions them in the same category in the vehicle body type model. Single person households are found to own none or one vehicle body types at the most, which is consistent with expectation as such household do not need more than one vehicle for their travel needs in general. Larger households as well as households with more number of drivers are likely to own more vehicle body types. This translates to the variety seeking nature of different individuals in such households.

Presence of children is found to negatively influence owning a single vehicle body type. Chauffeuring needs of children require owning a bigger vehicle (such as a van) in addition to owning another vehicle body type for usual travel in such households. This finding is corroborated by the significance of same variable in the three vehicle body type category. The likelihood ratio test statistic for the model is 1658.30 which is substantially higher than the critical χ^2 value with 21 degrees of freedom and 99% level of significance. This confirms the presence of exogenous variable effects in the model specification.

Table 3.12 MNL Model of Number of Vehicle Body Types

Number of Vehicle Body Types	Explanatory Variable	Coefficient	t-statistic
Zero	Constant	0.68	2.09
	Lowest income household (< $25,000)	2.09	8.42
	Low income household ($25,000 - $49,999)	0.61	2.38
	Housing unit owned (from variable HOMEOWN)	-1.43	-8.58
	Household size = 1	2.11	10.45
	Zero worker household	1.14	6.15
	Population density of the TAZ in which the household resides	0.00011	4.08
One	Constant	3.67	19.69
	Lowest income household (< $25,000)	1.02	7.95
	Low income household ($25,000 - $49,999)	0.65	6.86
	Household size = 1	1.99	15.62
	Proportion of multi-family housing units in the TAZ	0.26	1.39
	Presence of children in the household	- 0.33	-3.20
Two	Constant	2.96	13.01
	Household resides in rural area (from variable URBRUR)	0.19	2.01
	High income household ($75,000 - $99,999)	0.35	3.34
	Highest income household (≥ $100,000)	0.23	2.37
	Household size = 4 or more	0.30	2.73
	Housing unit owned (from variable HOMEOWN)	0.98	6.47
Three	Housing unit owned (from variable HOMEOWN)	1.42	5.90
	Count of adult HH members at least 18 years old	0.55	6.40
	Three or more worker household	0.46	1.96
	Population density of the TAZ in which the household resides	-0.00008	-2.82
	Presence of children in the household	0.36	2.55

Goodness of fit	
Sample Size (Number of Households)	4,262
Likelihood ratio	1658.30
$\chi^2_{21,0.001}$	46.80

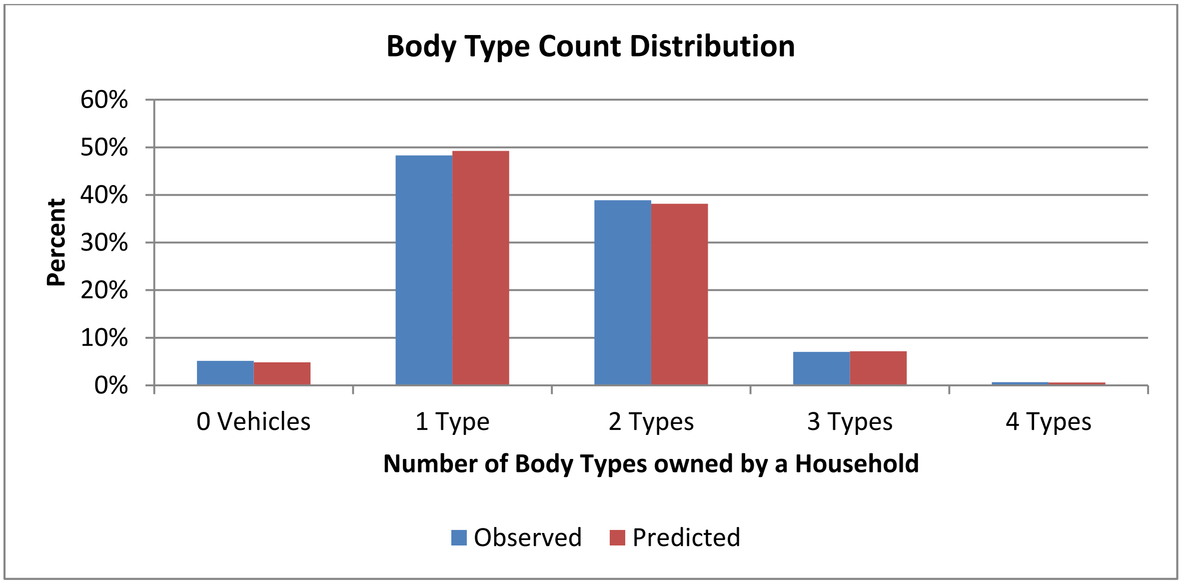


Figure 3.7 Observed versus Predicted Vehicle Body Type Distribution

Figure 3.7 presents the comparison of observed and predicted body type distribution. It can be observed that the model replicates the observed vehicle body type distributions quite well. As expected the households owning more vehicle body types (3, ≥ 4) are very few in the dataset and the model predictions line up quite well with the observed distributions. Majority of the households in the dataset are found to own one vehicle body type, which doesn't necessarily mean that these households own a single vehicle. Even households owning multiple vehicles might fall under this category, if all of their vehicles happen to be the same body type (multiple cars, vans etc.). The results presented are for uncalibrated version of the MNL model.

In addition to the comparison of aggregate distributions, the model was tested for its efficiency in replicating the body type distributions at disaggregate income level classification. Results of the comparison are presented in Figure 3.8. A more interesting observation comes from looking at the vehicle body type distribution of different income segments together (left to right in the figure). The percentage of households owning zero vehicle body types (or no vehicles at all) is higher in the lowest income category and this percentage slowly reduces as we move across the higher income segments. Similarly, the percent of households owning one vehicle body type is high in lowest and low income categories and this percentage gradually decreases as the household income increases. This observation speaks to the affordability combined with variety seeking nature of households in respective income segments. Though the observed and predicted distributions are presented side-by-side, the propagation of vehicle type distribution across different household income segments seems rather continuous because the model is able to predict the vehicle body type distributions across all the market segments quite well. This imparts much confidence to use the model prediction as control distribution that should be match by the heuristic mileage reallocation algorithm.

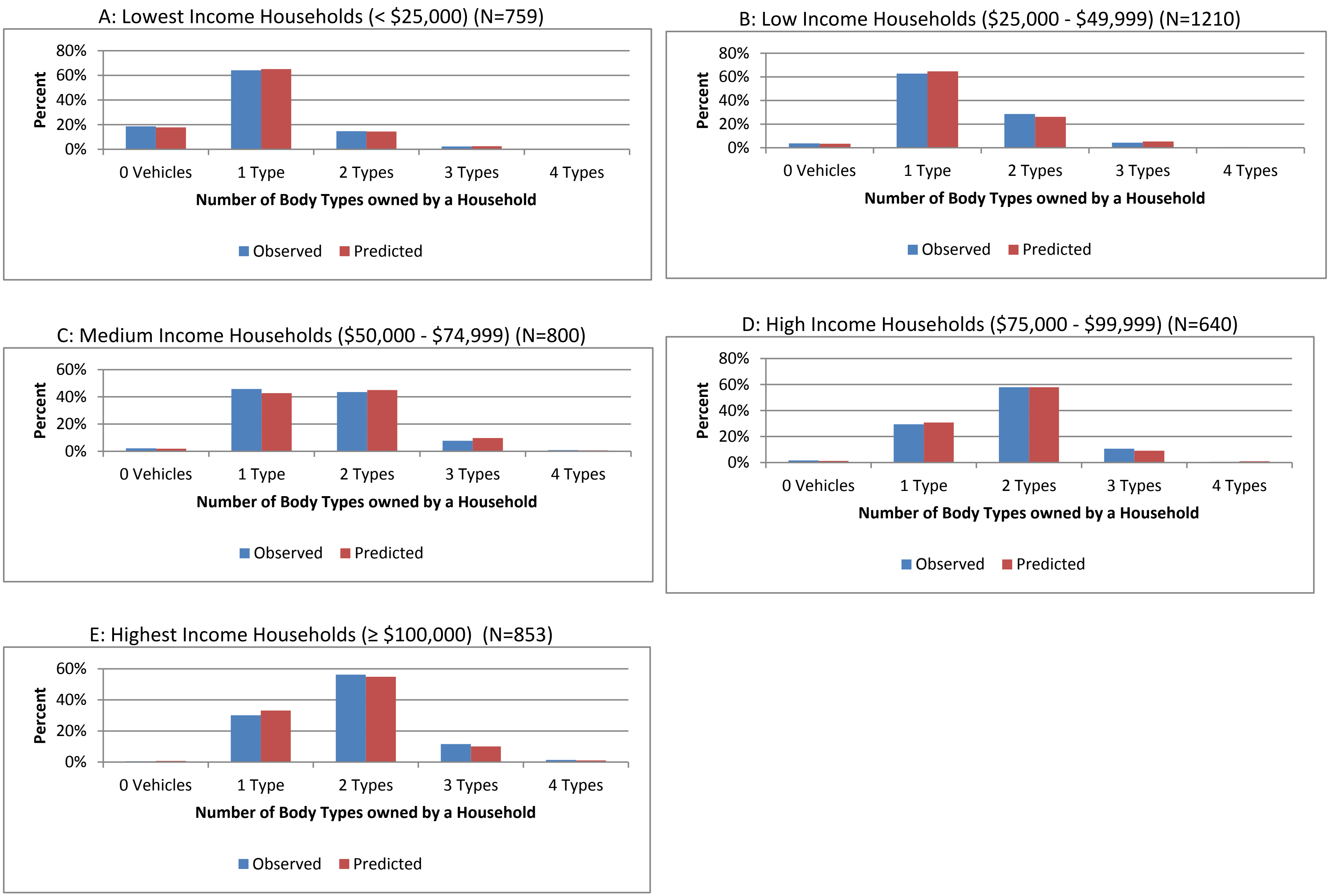


Figure 3.8 Observed versus Predicted Vehicle Body Type Distributions by Income Category

3.2.2.5 Heuristic Mileage Reallocation (HMR) Algorithm

The heuristic mileage re-allocation algorithm takes output of MDCEV model and MNL model of number of alternatives as input, to redistribute the mileage to number of alternatives owned by the household. Several alternative approaches were tested to effectively predict the fleet mix as well as body type distributions and the best approach among the once tested is presented. The mileage reallocation algorithm operates at the level of each household, where it reallocates the mileage output of the MDCEV model. The output from MNL model of number of vehicle alternatives gives information about how many distinct vehicle body-type x age categories does the household own. From the output of MDCEV model, cumulative mileage distribution of the household is computed for the household. A random number is generated and based on location of the random number in the cumulative mileage distribution of the household, a vehicle is selected as 'owned' by the household. The alternative chosen (and the corresponding mileage) is removed from the dataset thereby eliminating choice of the same alternative multiple times. This process is carried out 'k' times, where k (number of vehicle alternatives) is predicted by the number of alternatives model.

At the end of the iteration for a particular household, the HMR algorithm selects all the alternatives owned by the household. Now, the mileage consumed by these alternatives is scaled up proportionally to account for the annual motorized mileage consumption of the household. Once, the HMR algorithm reallocates the mileages for all households according to the input provided by number of alternatives model, the predicted body type distribution of the entire sample is compared against the body type distribution predicted by the MNL model of number of body types. Absolute percent difference is computed between both the distributions and this is checked against a pre-set tolerance limit selected by the user (say 5%).

If the HMR algorithm passes the tolerance check, the output of HMR algorithm goes in as input to the count models. If not, the entire application process is repeated after calibrating the model components as necessary. This process is carried out repeatedly until the percent difference between the two distributions (from MNL model of number of body types and the output of HMR algorithm) satisfies the tolerance criteria. In the context of the current modeling effort, a few coefficients in the MDCEV model specification are calibrated and asserted to match the observed vehicle fleet composition patterns better. The calibration/assertion exercise was carried out with due caution regarding any unexpected consequences such changes might bring about. The output of HMR algorithm gives us the final fleet composition of every household in the dataset. The output of this algorithm would have successfully predicted the vehicle ownership of the household, body type composition and vintage composition of the vehicles owned. As discussed before, comparisons between observed and predicted distributions are made for:

- Average annual mileage
- Vehicle fleet mix distribution
- Aggregate vehicle body type distribution

Comparison of observed and predicted fleet composition patterns from the output of the HMR algorithm are presented in Figure 3.9. Average annual mileage distribution comparisons are presented on the top axis whereas the bottom axis presents the vehicle type distributions.

It can be observed from the figure that the model replicated the observed fleet composition as well as mileage consumption patterns very well. It should be noted that an exact match between observed and predicted distributions is quite difficult to achieve and might require extensive calibration of the model system. It was felt prudent to rather capture the fleet composition patterns with slight calibration of the model system, than exactly match both of these distributions. It can be observed from the figure that within each body type, households prefer to drive newer vehicles more than that of older ones.

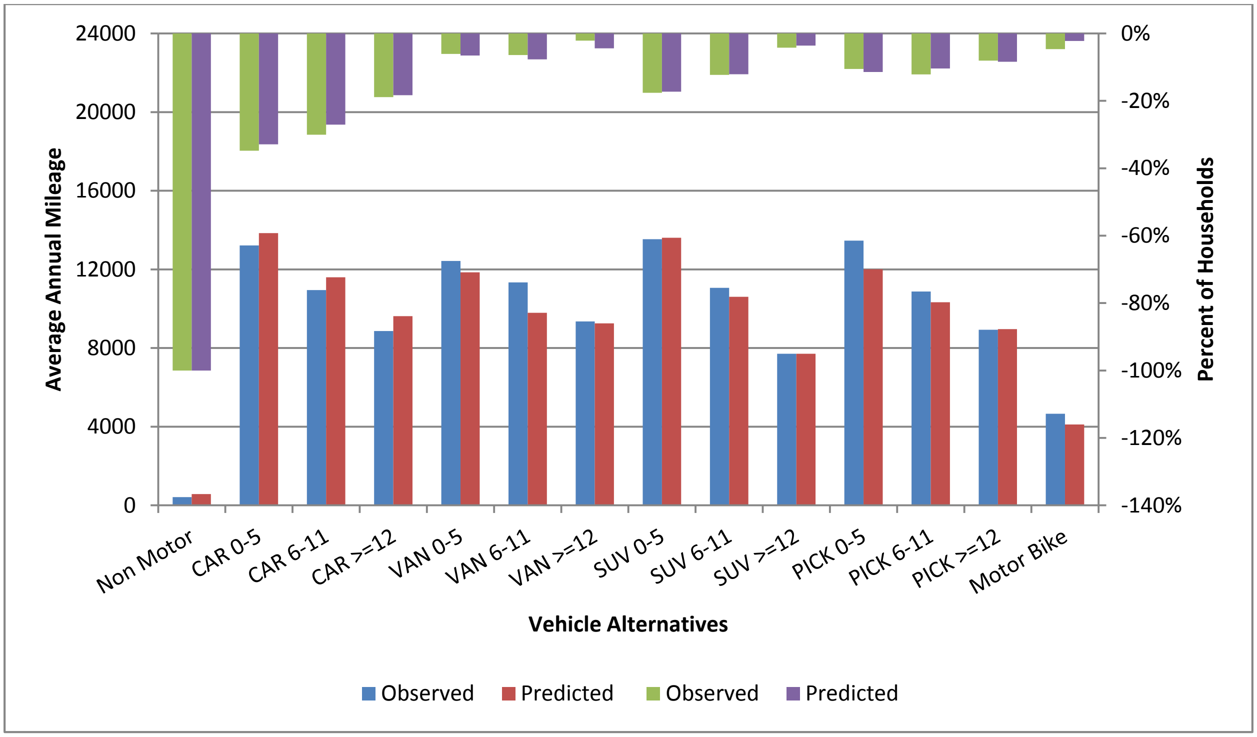


Figure 3.9 Observed versus Predicted Distributions

Another distribution that the output of the HMR algorithm is expected to match is the aggregate vehicle body type distribution. This comparison is presented in Figure 3.10. Note that the comparisons made in this figure are not against the observed vehicle body type distribution, but are against the vehicle body type distribution as predicted by the MNL model of body type count. While the base year data for this comparison is readily available for making this comparison, horizon years for which the model system would have to predict the vehicle fleet composition will not have this data readily available. Only socio-economic data will be provided for the model as input from which the vehicle fleet composition model system should predict the fleet mix. Keeping this in mind, comparisons are made against predicted data and not the observed, as this is how the model would be used for any horizon year prediction. Compared to the vehicle body type distribution without the HMR algorithm, the predicted distribution replicates the observed vehicle body type distribution quite closely. The output from HMR algorithm matches the observed patterns across all the comparisons made.

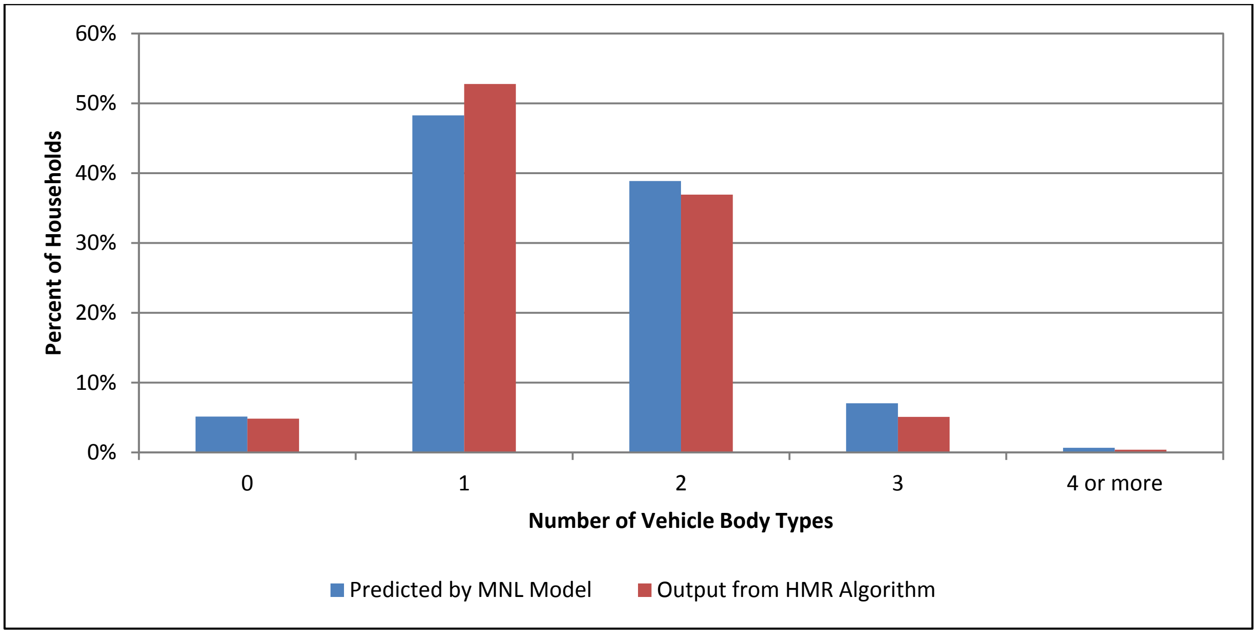


Figure 3.10 Observed versus Predicted Vehicle Body Type Distribution

In addition to comparisons at the aggregate level, the model system is tested on specific market segments to see how well it replicates the fleet mix and mileage consumptions of specific types of households. One such comparison is presented in Figures 3.11 and 3.12 for households residing in urban versus rural areas as these households might have significantly different mileage consumption as well as fleet mix patterns.

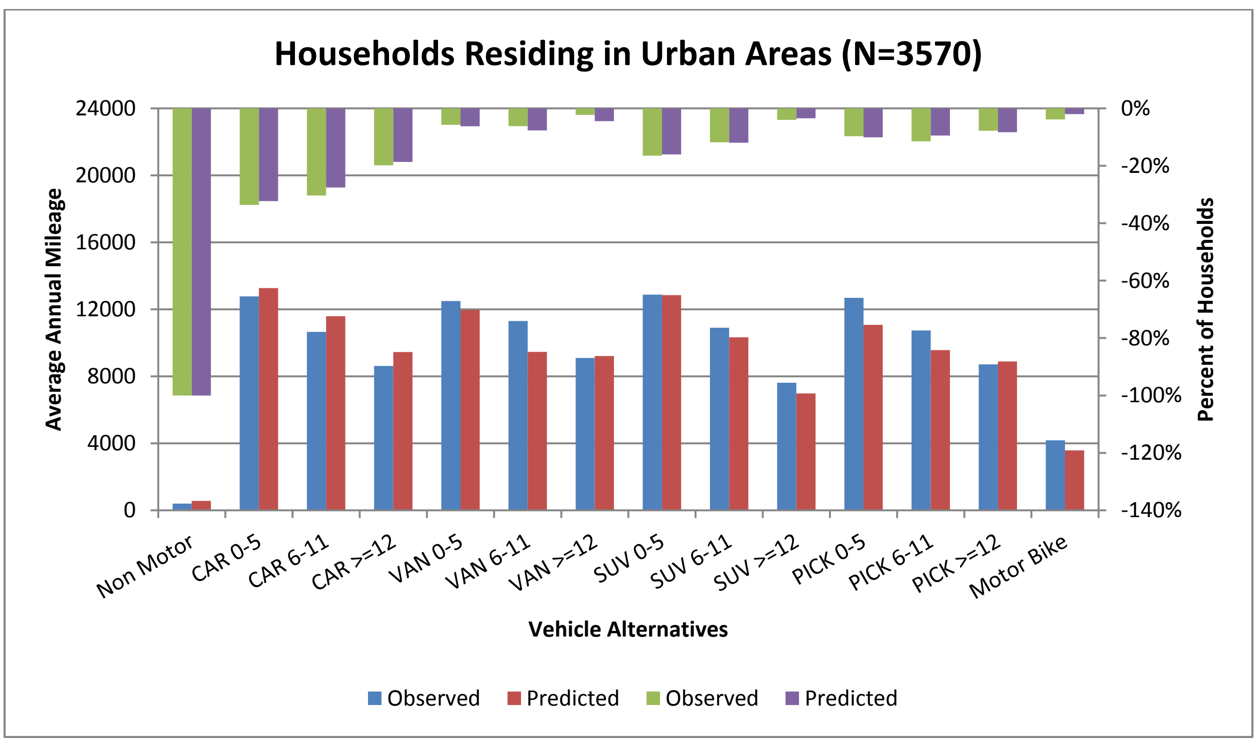


Figure 3.11 Observed versus Predicted Distributions for Urban Residents

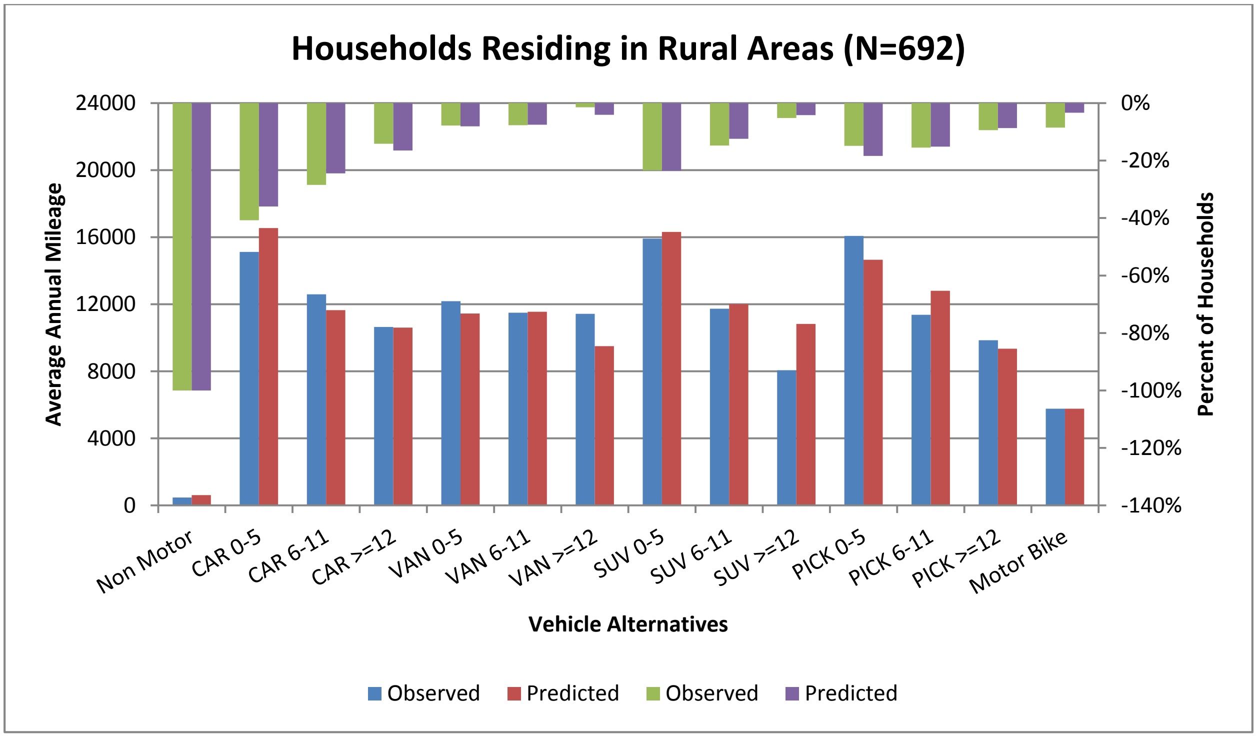


Figure 3.12 Observed versus Predicted Distributions for Rural Residents

The representation of households living in urban areas is slightly higher in the dataset (84%) than the national average of about 70%. This finding is consistent with the geography of Greater Phoenix Metropolitan Region which is predominantly urban. An important observation from the comparison chart shown in Figures 3.11 and 3.12 is that rural residents have greater average annual mileage consumptions across almost all vehicle types. This finding is behaviorally intuitive as rural residents indeed tend to have greater annual mileages as a result of travel to/from the adjacent city to engage in various activities. It is also observed that households residing in rural areas have a higher proportion of pick-ups in their vehicle fleet (across all vintage categories), than their urban counterparts. Similar comparison is shown for low and high income households in Figures 3.13 and 3.14.

The main takeaway from the comparison of vehicle fleet composition patterns of low income households is the representation of number of vehicles in the 'newest' vintage category (0-5 years old) across all vehicle body types. High income households are found to own and use newer vehicles more. This is generally expected behavior as such households usually have a faster turnover of vehicles in their fleet. The vehicle fleet composition model system is able to replicate the ownership and utilization patterns of both these market segments reasonably well. The model does not 'exactly' match the observed distributions, but in general captures the patterns in observed data quite well.

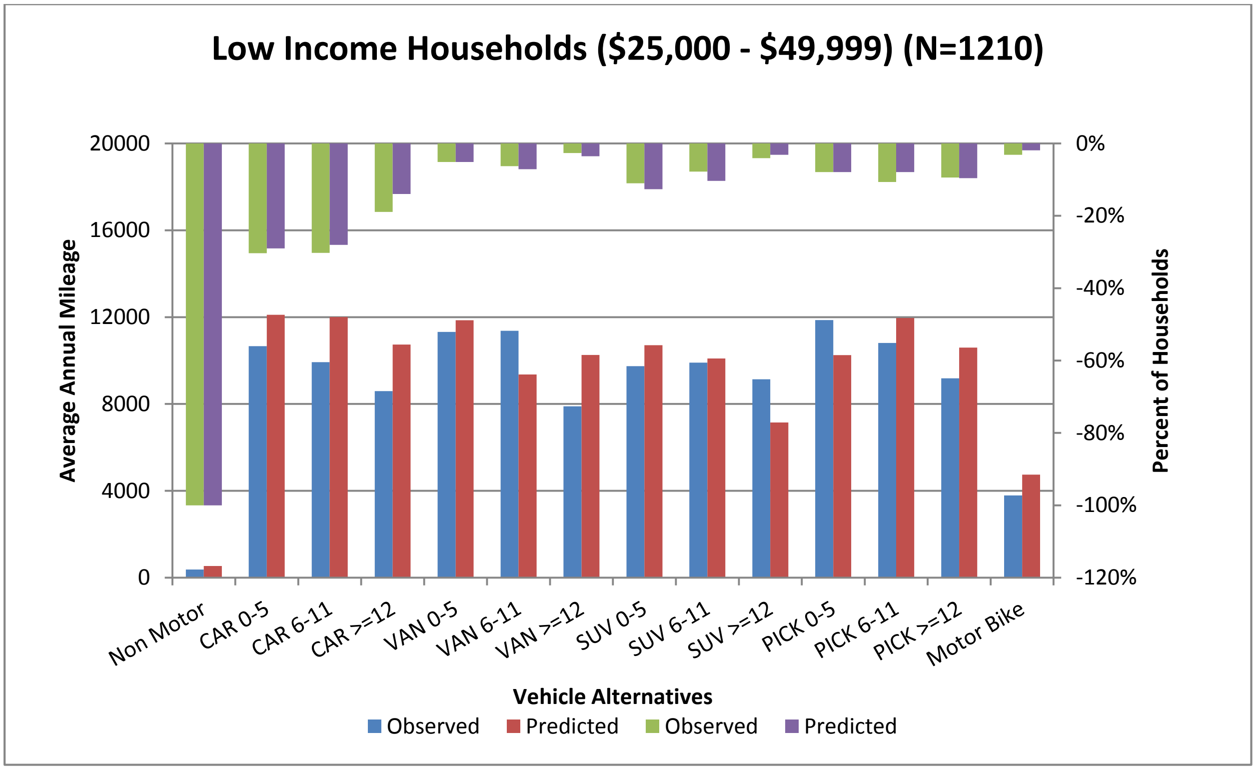


Figure 3.13 Observed versus Predicted Distributions for Low Income Households

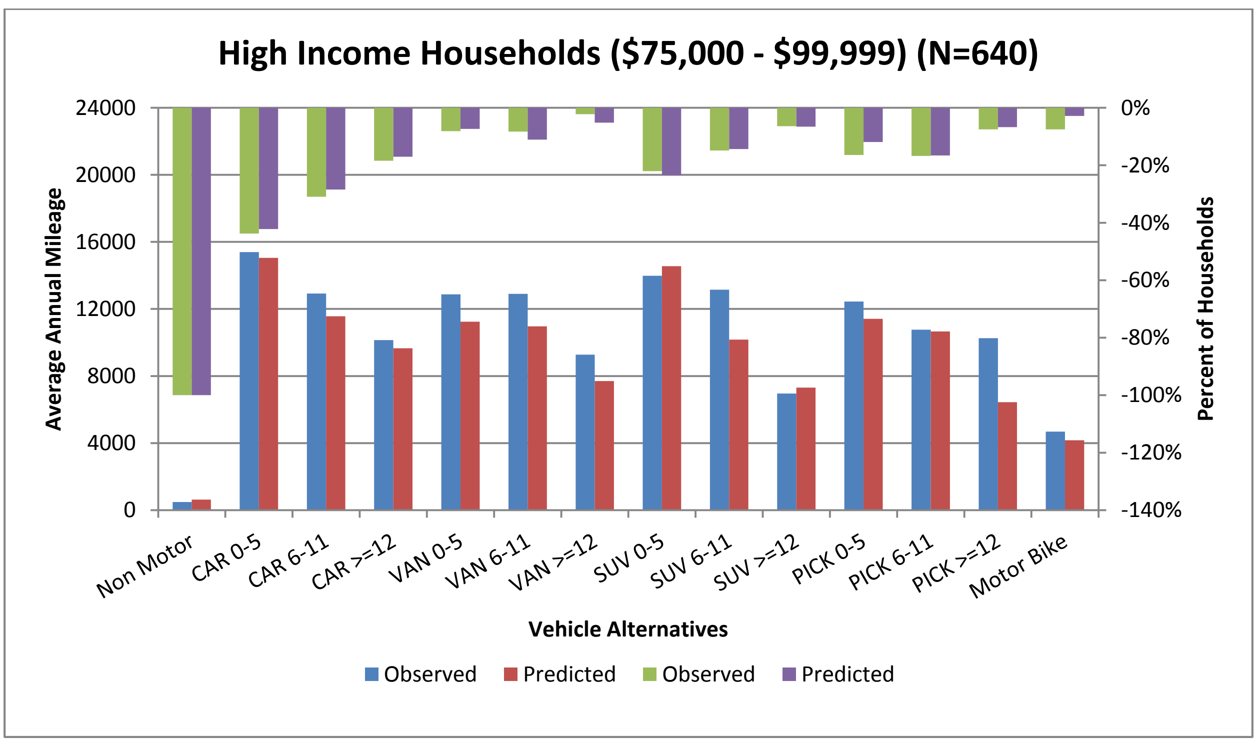


Figure 3.14 Observed versus Predicted Distributions for High Income Households

The output of HMR algorithm replicated the observed vehicle fleet composition patterns quite well. The output of HMR algorithm goes as input to the count models. The count models determine if all the mileage consumed by a household within a particular vehicle alternative belongs to one or multiple vehicles. The count models are necessary because vintage classifications are aggregated into 3 categories for ease of estimation and application of the fleet composition model system. Suppose, the output of HMR algorithm determines that a household uses a car 0-5 years old to travel 25000 miles annually, the count model determines if all of this mileage is consumed using just one car 0-5 years old or if the household owns multiple cars of 0-5years of age. Ideally, a count model should be estimated for each of the 13 different vehicle categories defined for the MDCEV model, but this will heavily increase the number of individual components in the model system while decreasing the data available to estimate each of the individual count models. So, it was felt prudent to estimate one count model for each of the body types, with vintage serving as an explanatory variable in the models. If the household has non-zero mileage consumption in any of the vintages of a particular body type, count model of that body type is applied for the household. The output of count models is a test to the efficacy of the entire model system as this is a sequential application process.

3.2.2.6 Count Models

Count models take the mileage output from HMR algorithm and determine if all of that mileage is consumed by a single vehicle or multiple vehicles in the alternative under consideration. Ideally, 13 different count models should be estimated (one for each of the body type – age classification), but this would heavily increase the number of individual components in the model system. It was felt prudent to estimate one count model for each body type with vintage categories serving as explanatory variables in the models. The effectiveness of this simplification is tested and the results were satisfactory across all the vehicle body types. Ordered probit count models are estimated for car, van, SUV and pick-up body types. Model estimation results of car count model are presented in Table 3.13.

Table 3.13 Car Count Model Estimation Results

Explanatory Variable	Coefficient	t-statistic
Constant	-2.74	-17.09
Indicator for car 0-5 years old	0.95	14.44
Indicator for car 6-11 years old	1.05	16.21
Low income household ($25,000 - $49,999)	-0.10	-1.63
Highest income household (≥ $100,000)	0.19	3.10
Three or more worker household	0.41	3.44
Count of adult household members (at least 18 years old)	0.47	9.15
Household size = 1	-0.42	-3.93
Household size = 4 or more	-0.26	-3.64
Housing unit owned (from variable HOMEOWN)	0.28	2.62
Goodness of Fit		
Log-likelihood	830.75	
$\chi^2_{9,0.001}$	27.88	

From the model estimation results, it was observed that highest income households usually include multiple cars in their fleet while low income households on the other hand are not likely to operate multiple cars. It was also observed that three or more worker households tend to own multiple cars. This finding is behaviorally consistent as such households might usually require more than one car for daily commute travel for multiple workers in the households. Households living in an owned housing unit have a greater propensity of owning multiple cars. This variable might act as a proxy for the income of the household. The likelihood ratio statistic of the model (830.75) is significantly greater than critical χ^2 value at 99% confidence level.

The car count models are applied on only those households for whom the HMR algorithm allocates at least some non-zero mileage in any of the car vintage categories. Suppose the HMR algorithm allocated a mileage of 10000 miles for car 0-5 years old category and 5000 miles for car 6-11 years old category, car count model is applied on each of these categories to identify if the households own multiple cars in the category 0-5 years old and/or car 6-11 years old. If the car count model predicts multiple cars for any of the categories, mileage for that alternative is evenly distributed among the number of predicted alternatives. In the above example, if the model predicts that the households own two cars in vintage category 0-5 years, each vehicle is assigned a mileage of 5000.

The comparison of observed and predicted car counts is presented in Figure 3.15 which shows the comparison between car count distributions of the entire dataset. The results presented are for the uncalibrated version of the count model. It can be observed that model replicates the car count distribution quite well with slight under prediction in the 2 car category. Some calibration of the model is warranted to better match the observed distributions.

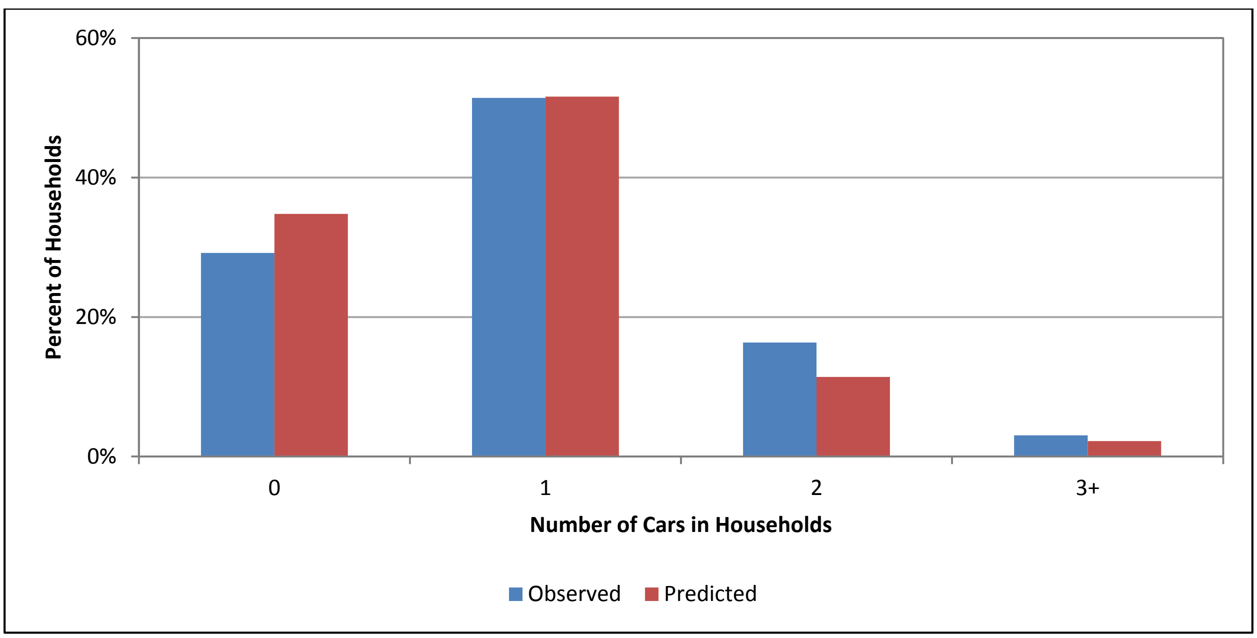


Figure 3.15 Observed versus Predicted Distributions for Car Count

Table 3.14 Van Count Model Estimation Results

Explanatory Variable	Coefficient	t-statistic
Constant	-4.62	-7.75

Indicator for Van 6-11 years old	1.95	3.92
Indicator for Van 12 years or older	2.57	4.97
Low income household ($25,000 - $49,999)	-0.59	-1.82
Annual mileage of VANs in a household	0.00006	5.46
TAZ with low density (3rd Quartile)	0.60	1.76
TAZ with high regional employment accessible within 10 minutes by auto (1st Quartile)	0.87	2.99
Employment density of the TAZ that the household resides	0.00	-1.47
Goodness of Fit		
Log-likelihood	-67.74	
$\chi^2_{7,0.001}$	103.14	

Table 3.14 presents the estimation results for van count model. From the model result, it was found that households tend to own multiple vans of older vintages than new ones. Also, the mileage consumption variable is positive and significant. If the HMR algorithm allocates a high mileage to any of the van categories the count model will be able to identify and allocate that mileage to multiple vehicles of the same category. Low income households had a lower propensity to own multiple vans (or multiple vehicles of any category for that matter), which is intuitive. Households residing in TAZs with high employment density (mixed use zones) tended not to own multiple vans.

Figure 3.16 presents the comparison between observed and predicted van count distributions.

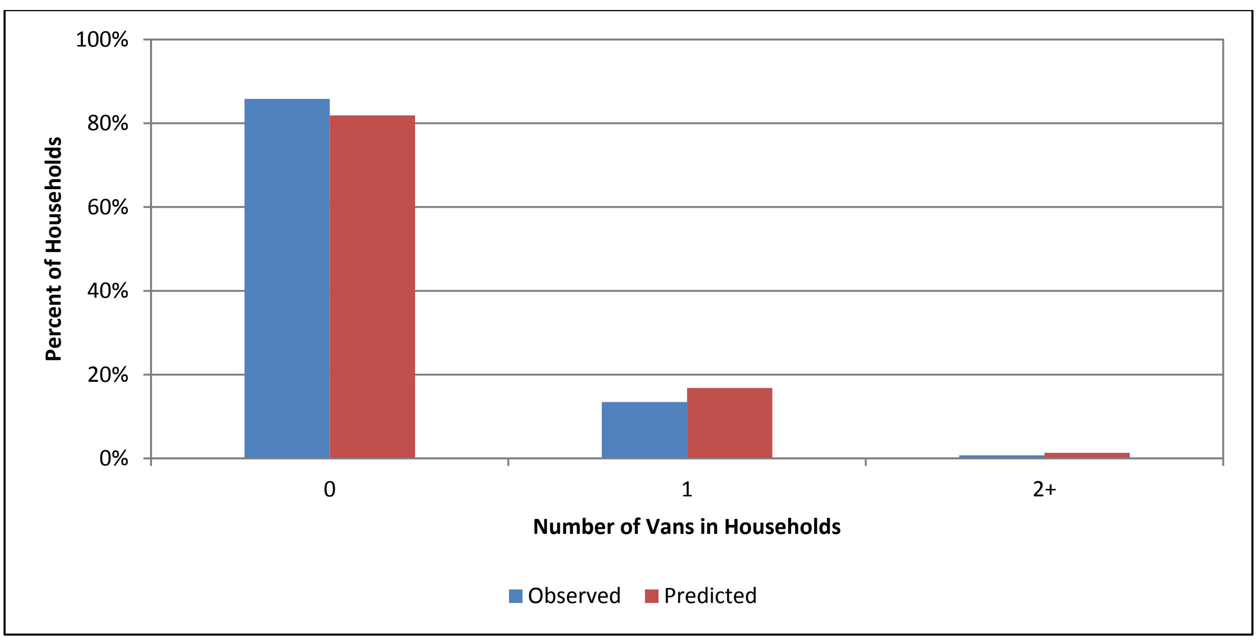


Figure 3.16 Observed versus Predicted Distributions for Van Count

Table 3.15 presents the model estimation results for SUV count model. Unlike the van count model, SUV count model has a positive and significant coefficient for the newer vintage category (0-5 years), which specifies that household who own and drive SUVs, like them rather new than old. Income categories used in the SUV count model show very intuitive findings. From model results, low income households in general have a lower probability of owning multiple SUVs, whereas high income households on the other hand have a greater propensity to multiple vehicles of this type. Usage of annual mileage variable in the model ensures distribution of high mileage predictions for any alternative to multiple vehicles in this category.

Table 3.15 SUV Count Model Estimation Results

Explanatory Variable	Coefficient	t-statistic
Constant	-3.64	-14.06
Indicator for SUV 0-5 years old	0.90	6.16
Indicator for SUV 12 years or older	1.50	8.61
Count of adult household members (at least 18 years old)	0.46	5.23
Low income household (\$25,000 - \$49,999)	-0.53	-2.84
Highest income household (≥ \$100,000)	0.40	3.57
Annual mileage of SUV trips	0.00004	9.92
Three or more worker household	-0.46	-1.93
Percent of regional employment within 30 minutes of transit accessibility from the TAZ	-34.16	-2.07
Goodness of Fit		
Log-likelihood	-359.60	
$\chi^2_{8,0.001}$	309.57	

An interesting observation with respect to the SUV count model is the magnitude of the coefficient on annual mileage. For SUV count model the value of this coefficient is 0.00004, where as in the van and pick-up count models, the value same coefficient is greater (0.00006). This finding translates to the fact that households who own multiple SUVs drive them for relatively lower annual mileage than that of vans and pick-up truck. This finding is behaviorally consistent in that SUVs are generally used for leisure travel.

Figure 3.17 presents comparison of observed and predicted distributions for SUV count model. Panel A depicts the comparison of SUV count distributions for the entire dataset and Panel B presents the results by income level. The model predicts presence of multiple SUVs across different income categories. As the category of household income increases, it can be observed that the presence SUVs in the household (1, ≥ 2) slowly increases and the model is able to predict this pattern quite well.

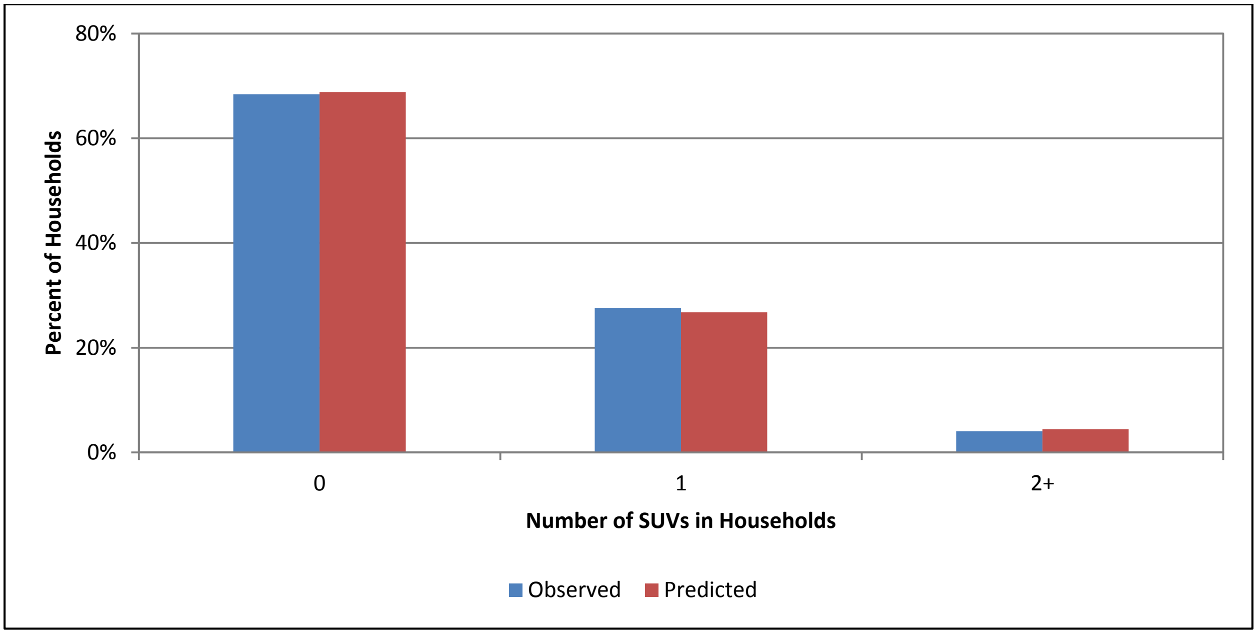


Figure 3.17 Observed versus Predicted Distributions for SUV Count

Model estimation results for pick-up count models are presented in Table 3.16. Annual mileage variable is positive and significant in the model specification which avoids allocation of greater mileages to a single pick-up truck category. Households residing in TAZs with low density are found to own multiple pick-up trucks. These might refer to households residing in rural areas.

Table 3.16 Pick-up Count Model Estimation Results

Explanatory Variable	Coefficient	t-statistic
Constant	-3.95	-8.40
Indicator for pickup truck 0-5 years old	0.87	5.16
Indicator for pickup truck 6-11 years old	1.02	6.21
Annual mileage of pickup truck in the household	0.00005	9.88
Highest income household (≥ $100,000)	-0.29	-2.08
Three or more worker household	0.46	2.32
Housing unit owned (from variable HOMEOWN)	0.81	1.96
Employment density per 10 square miles	0.01	1.58
TAZ with low density (3rd Quartile)	0.30	1.88
TAZ with high regional employment accessible within 30 minutes by auto (1st Quartile)	0.27	2.09
Goodness of Fit		
Log-likelihood	-301.91	
$\chi^2_{9,0.001}$	240.62	

Highest income households (≥ $100,000) had a lower propensity to own multiple pick-up trucks. Figure 3.18 shows the comparison of observed and predicted distributions of pick-up counts. Percentage of households who own multiple pick-trucks are quite low in the dataset (only 3 % households own ≥ 2 pick-up trucks). The model is able to replicate observed patterns quite well across different income categories as well as the overall distribution. On the whole, results of the fleet composition model system are quite encouraging. Some calibration of the model system is warranted to better match the observed patterns. Information at such disaggregate level regarding the fleet composition patterns of households in a region help in accurate emission predictions.

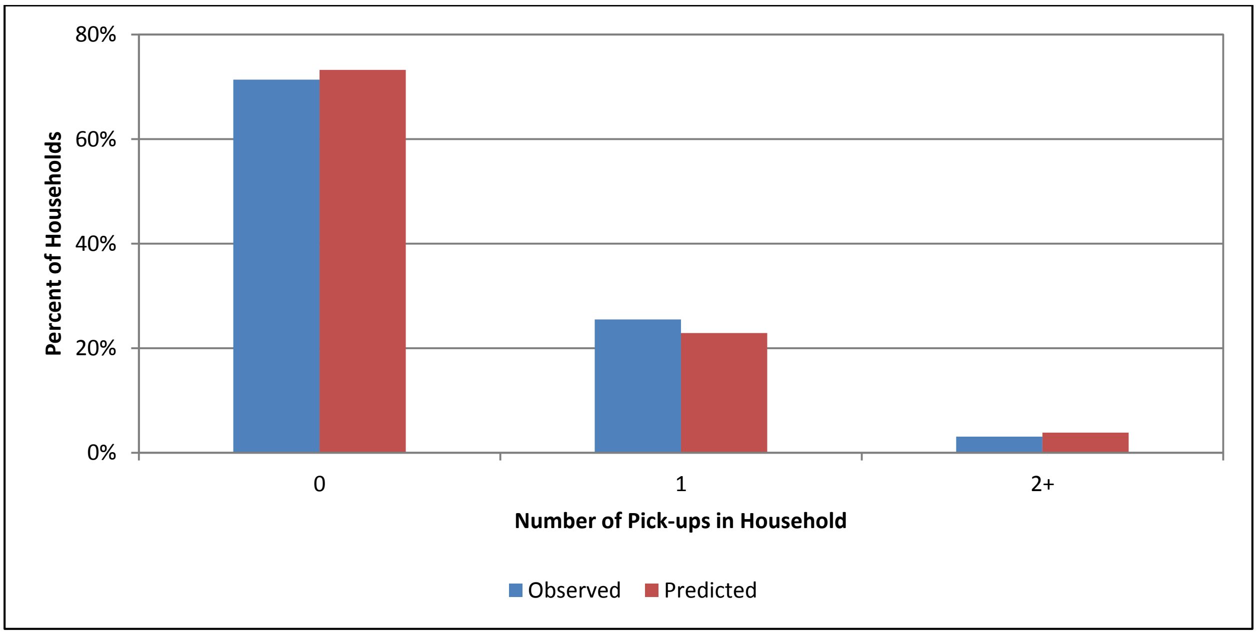


Figure 3.18 Observed versus Predicted Distributions for Pick-up Count

3.3 Sensitivity Analysis

Once the model was found to replicate the observed vehicle ownership patterns satisfactorily, a sensitivity analysis exercise was carried out to examine the ability of the model system to respond in a meaningful way to changes in input test conditions. First, a baseline scenario was established by applying the model system on the entire sample (4,262) households. Five scenarios were created by varying the percent of regional employment accessible from a households TAZ location by auto mode. The regional employment accessible was increased incrementally by 10%, 20%, 30%, 50% and 100% from the baseline. To build these scenarios, auto skims were used to select the percent regional employment accessible within a set travel time (10 minutes and 30 minutes) in the baseline and this employment was increased by respective percentages for each of the scenarios. This translates to increasing the accessibility of a household's TAZ location by enhancing the percentage of regional employment accessible from every TAZ by auto.

It was observed in the modeling exercise that accessibility has a negative impact on vehicle ownership patterns i.e., households living in denser developments usually tended to own fewer vehicles. The reason for this behavior is twofold. First, households who are more environmentally proactive and already own fewer vehicles might self-select themselves into dense mixed-use urban locations. This phenomenon is called as residential self-selection and plays in an important role in auto ownership as well as travel demand in general. This topic has been the focus of many earlier studies (Cao et al., 2006; Bhat and Guo, 2007; Pinjari et al., 2008a; Bhat et al., 2013) and is not dealt with in the current research effort. The second reason is the fundamental causality between built environment and travel-behavior, which explains why dense urban development tend to be more walk/bike/transit friendly than sparse suburban neighborhoods (Frank and Pivo, 1994; Cervero and Seskin, 1995; Cervero and Kockelman, 1997; Ewing, 2008). Results of the sensitivity analysis test are presented in Figure 3.19. The figure depicts the changes in vehicle ownership patterns with varying accessibility measures. The results are aggregated by vehicle body type for easier understanding.

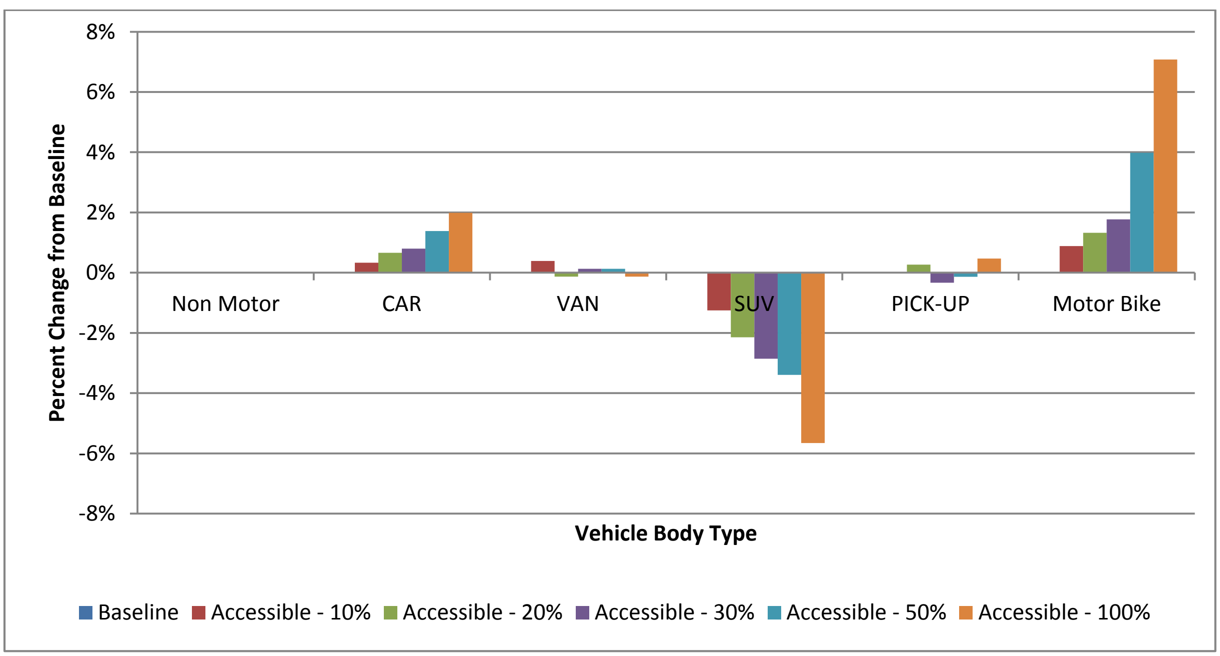


Figure 3.19 Change in Vehicle Ownership Patterns in response to Accessibility Enhancement

In general, the model system provides intuitive predictions for changes in vehicle ownership and utilization patterns in response to increasing accessibility measures. It was found that increasing accessibility positively influences ownership of smaller vehicles (such as cars) while the percent of households owning larger vehicle types (such as SUVs) gradually decreases. The percentage of households owning vans drops as well, but the changes are not as large in magnitude as that of SUVs, suggesting that households are more inclined to hold onto multipurpose vehicles in their fleet than the ones that are used mostly for luxury travel. Percent of households owning pick-up trucks remains largely unchanged.

The reason for this might be the type of TAZs in which households owning pick-up trucks usually reside in. It was observed from the results of the vehicle fleet mix model that households residing in rural localities usually tend to own pick-up trucks more. Such TAZs do not have a lot of regional employment accessible within 10 minutes of auto travel to begin with. And for this reason, even doubling the percent of regional employment does not have a significant effect on the ownership of this specific vehicle body type.

The changes in the vehicle ownership patterns with increasing accessibility are largely consistent with the notion that with increasing accessibility, households need to drive smaller distances to fulfill their daily travel needs (going to a grocery store, a movie etc.). Similar studies have found that larger vehicles are preferred for long distance travel (Konduri et al., 2011) and are less preferred in light of increased accessibility. The percent of households owning motorbikes heavily increases with increasing accessibility. While this is an intuitive finding in the sense that such vehicles are more convenient to make short trips, the magnitude of this change should be interpreted with caution. The changes are amplified owing to a lower level of motorbike ownership in the baseline conditions (only 190 households owned a motorbike). Even an increase of this category by 6% from the baseline only translates to 12 more motorbikes.

The average annual mileage patterns in the dataset with varying accessibility levels is shown in Figure 3.20 that zero mileage households are not included. The average annual mileage is computed by summing up the mileage attributed to a vehicle type and dividing it by the total number of vehicles in the fleet (for each scenario). As expected, the average annual mileage value for all motorized body types decreases as accessibility increases. Car mileage gradually decreases with increasing accessibility. This is an interesting finding in that, though the market share of this body type has seen an increase with increasing accessibility, there is an associated decrease in the usage of cars. This translates to the convenience of owning smaller vehicles and driving them to lower degrees less with increasing accessibility.

Vans and SUV categories also show a decreasing trend in annual mileage patterns though the pattern is not as consistent as in the case of cars. One plausible reason for this might be that as the number of vehicles in the fleet drops, even a slight change in the denominator in the calculation of average annual mileage might contribute to a more modest increase/decrease in per vehicle mileage. A closer look at both the market share and average mileage graphs for the van and SUV categories reveals that decrease in market share for these categories is more pronounced than the decrease in average annual mileage. It is possible that while fewer households own these vehicles in light of increased accessibility, the households who own such vehicles continue to drive these vehicles (on a per vehicle basis) for about the same number of miles. This points to the households who own and use larger vehicle type for long distance travel; such usage is therefore not impacted by changes in local accessibility. The non-motorized vehicle on the other hand shows a consistent increase in average annual mileage with increasing accessibility. This change is readily explained by the fact the decreased mileage consumption of the motorized alternatives translates to a corresponding increase in non-motorized mileage. Again, the percent increase in non-motorized mileage seems amplified due the smaller magnitude of non-motorized mileage in the baseline scenario. The increase in non-motorized mileage in the extreme accessibility scenario is about 25 miles per household per year (6% increase). While this may not seem all that significant, it should be noted that this is complemented by a corresponding decrease in motorized mileage. For the dataset under consideration this comes out to about 100,000 lesser vehicle miles driven per year. An increase in accessibility is in general associated with increased levels of walking and bicycling (Ewing and Cervero, 2001; Krizek, 2003) synonymous to the travel characteristics of mixed use urban neighborhoods.

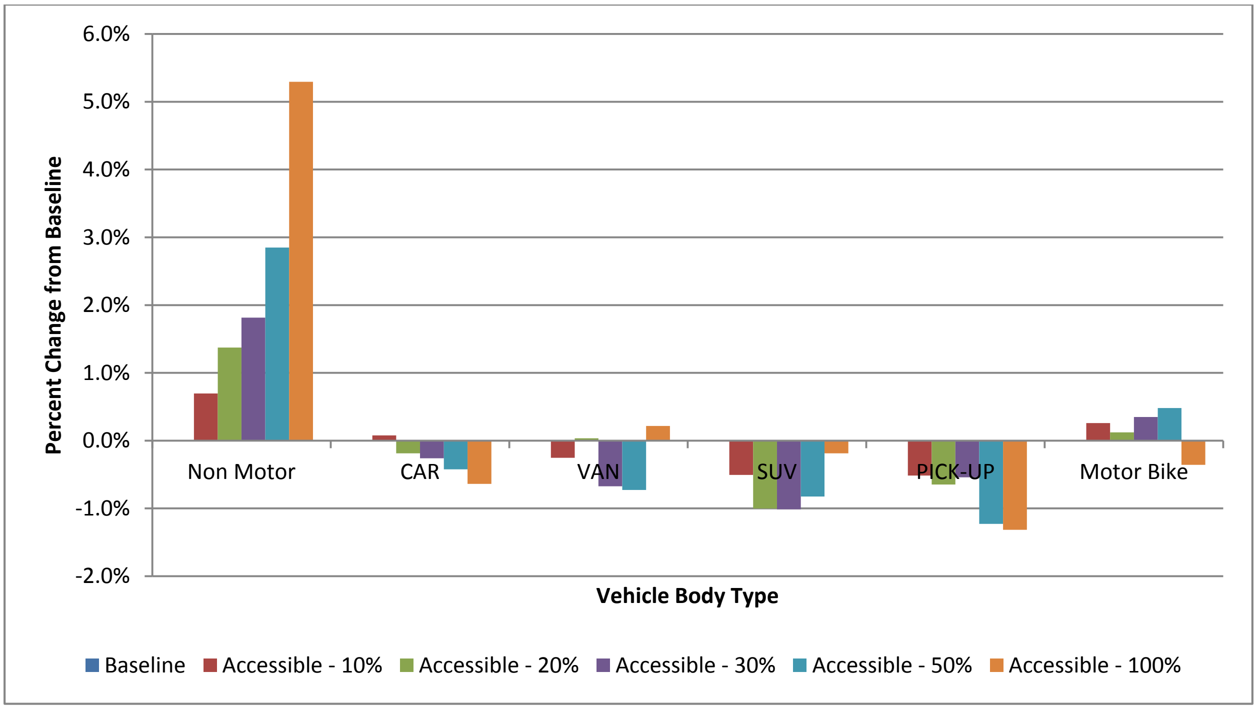


Figure 3.20 Change in Annual Mileage Consumption Patterns in response to Accessibility Enhancement

In addition to replicating the observed fleet composition patterns quite well in the base year, the model system is found to respond in a behaviorally intuitive way to changes in inputs provided. This reinforces the confidence in using the developed model system as a plug-in fleet composition module to any of the existing activity-based models in practice. For ease of integration, the model system is completely coded on open source coding platform 'R'.

4. Conclusions

The household vehicle ownership model for four-step travel demand model has been updated based on the National Household Travel Survey (NHTS) 2008 add-on dataset for the MAG modeling area. The updated model results were quite well matched to the Public Use Microdata Sample (PUMS) dataset based on ACS 2006-2010.

The vehicle fleet composition model shows promise in depicting the snapshot of vehicle fleet composition in the observed data along multiple dimensions. The model system developed is parsimonious in the sense that a number of model components are kept to a minimum, yet quite effective in predicting vehicle ownership patterns accurately. The model system takes any horizon year data comprising of the socio-demographic characteristics of the households as well as built environment characteristics of household's residential location and predicts the vehicle owned by a household classified by body type, age and count. The model system is tested for its sensitivity to changing land use characteristics and it provided logically intuitive results.

5. References

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

Bhat, C. R. (2005). A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. Transportation Research Part B: Methodological, 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: Methodological, 42(3), 274-303

Bhat, C. R., and Guo, J. Y. (2007). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological*, *41*(5), 506-526

Bhat, C. R., Paleti, R., Pendyala, R. M., Lorenzini, K., and Konduri, K. C. (2013). Accommodating immigration status and self-selection effects in a joint model of household auto ownership and residential location choice. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2382*, Transportation Research Board of the National Academies, Washington, D.C., 142-150

Cao, X., Handy, S. L., and Mokhtarian, P. L. (2006). The influences of the built environment and residential self-selection on pedestrian behavior: evidence from Austin, TX. *Transportation*, *33*(1), 1-20

Cervero, R., and Kockelman, K. (1997). Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199-219.

Cervero, R., and Seskin, S. (1995). An evaluation of the relationships between transit and urban form. *Transit Cooperative Research Program*, *Transportation Research Board*, National Research Council, Washington, D.C.

Eluru, N., Bhat, C. R., Pendyala, R. M., and Konduri, K. C. (2010). A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. Transportation, 37(4), 603-626

EMFAC. *California Department of Transportation*. Retrieved April 10, 2014 from http://www.dot.ca.gov/hq/env /air/pages/emfac.htm.

Ewing, R. H. (2008). Characteristics, causes, and effects of sprawl: a literature review. In *Urban Ecology* (519-535). Springer, US.

Ewing, R. H., and Cervero, R. (2001). Travel and the built environment: a synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1780*, Transportation Research Board of the National Academies, Washington, D.C., 87-114.

Frank, L. D., and Pivo, G. (1994). Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit, and walking. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1466*, Transportation Research Board of the National Academies, Washington, D.C., 44-52.

Krizek, K. J. (2003). Neighborhood services, trip purpose, and tour-based travel. *Transportation*, *30*(4), 387-410

Konduri, K. C., Ye, X., Sana, B., and Pendyala, R. M. (2011). Joint model of vehicle type choice and tour length. *Transportation Research Record: Journal of the Transportation Research Board*, *No*. *2255*, Transportation Research Board of the National Academies, Washington, D.C, 28-37

MOVES, *Motor Vehicle Emission Simulator*. Retrieved April 10, 2014 from http://www.epa.gov/otaq/models/moves/.

My Car Stats. *Manufacturer Car Warranty Expiration Chart*. (October, 2010). Retrieved August 18, 2014 from http://www.mycarstats.com/content/car_warrantylimits.asp.

Pinjari, A. R., Eluru, N., Bhat, C. R., Pendyala, R. M., and Spissu, E. (2008a). Joint model of choice of residential neighborhood and bicycle ownership: accounting for self-selection and unobserved heterogeneity. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2082*, Transportation Research Board of the National Academies, Washington, D.C., 17-26

Paleti, R., Bhat, C. R., Pendyala, R. M., and Goulias, K. G. (2013). Modeling of Household Vehicle Type Choice Accommodating Spatial Dependence Effects. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2343*, Transportation Research Board of the National Academies, Washington, D.C., 86-94.

Pinjari, A.R., and C.R. Bhat, "Computationally Efficient Forecasting Procedures for Kuhn-Tucker Consumer Demand Model Systems: Application to Residential Energy Consumption Analysis," Technical paper, Department of Civil and Environmental Engineering, University of South Florida, July 2009

Pinjari, A. R., and Bhat, C. R. (2011). An efficient forecasting procedure for kuhn-tucker consumer demand model systems: application to residential energy consumption analysis. Technical Paper, Department of Civil & Environmental Engineering, University of South Florida.

Vyas, G., Paleti, R., Bhat, C. R., Goulias, K. G., Pendyala, R. M., Hu, H. H., Adler, T. J., and Bahreinian, A. (2012). Joint vehicle holdings, by type and vintage, and primary driver assignment model with application for California. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2302*, Transportation Research Board of the National Academies, Washington, D.C., 74-83.