UNIVERSITY OF MARYLAND

Modeling car ownership in the US

Cinzia Cirillo

Discrete Choice Model Workshop
EPFL, June 20th, 2014

The Present

Car Ownership in US and NL

Country	N cars x Hhld	0 car	1 car	2+ cars	Mileage
NL	1.07	~ 30%	~ 50 %	~ 20 %	~10.000
US	1.92	8.7 %	32.2 %	59.1 %	~ 20.000
Wash MA	1.87	7.8 %	26.7 %	65.5 %	21.922

More numbers on the comparison across US and Europe

- The average U.S. vehicle travels 42% more miles than the average car in Germany.
- The average U.S. vehicle consumes 112.5% more fuel than its German counterpart (2,040 gallons vs. 960), and even 21.4% more than a Canadian car (1,680 gallons).

Introduction

- The American households are highly dependent on private vehicles – in 2009, the average vehicle ownership per household is 1.92, and there are only about 9% of the households who do not have a car.
- In the U.S., transportation contributes approximately 27 percent of total greenhouse gas emissions. 71 percent of the oil consumption directs to fuels used in transportation, in which 40 percent is used to fill up gasoline tanks in our personal vehicles. The use of private vehicles has strong relationship with traffic congestion, energy consumption and our environment.
- Therefore, it is very crucial to understand the people's behavior on the wheels, particularly, how many vehicles they own, the types of the vehicles and how many miles they travel.
- In fact, households make those decisions simultaneously. As transportation modelers, we'd better to estimate the decisions in one system, in stead of separately, in order to best understand their travel behavior hence provide better reference for the policy makers.
- However, in the literature there are only a few studies that investigated the three choices jointly.

Literature Review

- Discrete-continuous models derived from conditional indirect utility function
  - The models estimate the choice probabilities and the demand equations sequentially, not simultaneously .
  - The estimates are consistent but not as efficient as full information maximum likelihood, because the unobserved component of utility and the error in the demand equation generally contain some common unobserved factors.
- Multiple Discrete Continuous Extreme Value (MDCEV) model
  - Does not include vehicle holding decision.
  - Requires fine classification of vehicles as one type of vehicle cannot be chosen twice by the household.
  - The assumption of fixed total mileage budget for every household implies that it is not possible to predict changes in the total number of miles in response to policy changes.
  - There is only a single error term underlying both discrete and continuous choices.
- Bayesian Multiple Ordered Probit and Tobit (BMOPT) Model
  - The computation becomes intensive for a large number of vehicle categories, as the number of equations to be estimated increases proportionally with the number of vehicle types.
  - Ordered mechanism may not perform as well as unordered mechanism in modeling car ownership decisions.

Research Objectives

- Develop a mathematical framework to model the household choices on vehicle ownership, the types and annual mileage traveled; in particular, the model should be able to
  - simultaneously estimate discrete (vehicle holding and types) and continuous (vehicle usage) decision variables;
  - take into account a large number of alternatives in both the vehicle holding and the vehicle type choices;
  - have no budget on the mileage traveled;
  - capture the correlations of the unobserved factors between the discrete and continuous parts;
  - have flexible specifications; and
  - be sensitive to policy analysis.

Research Objectives (Con't)

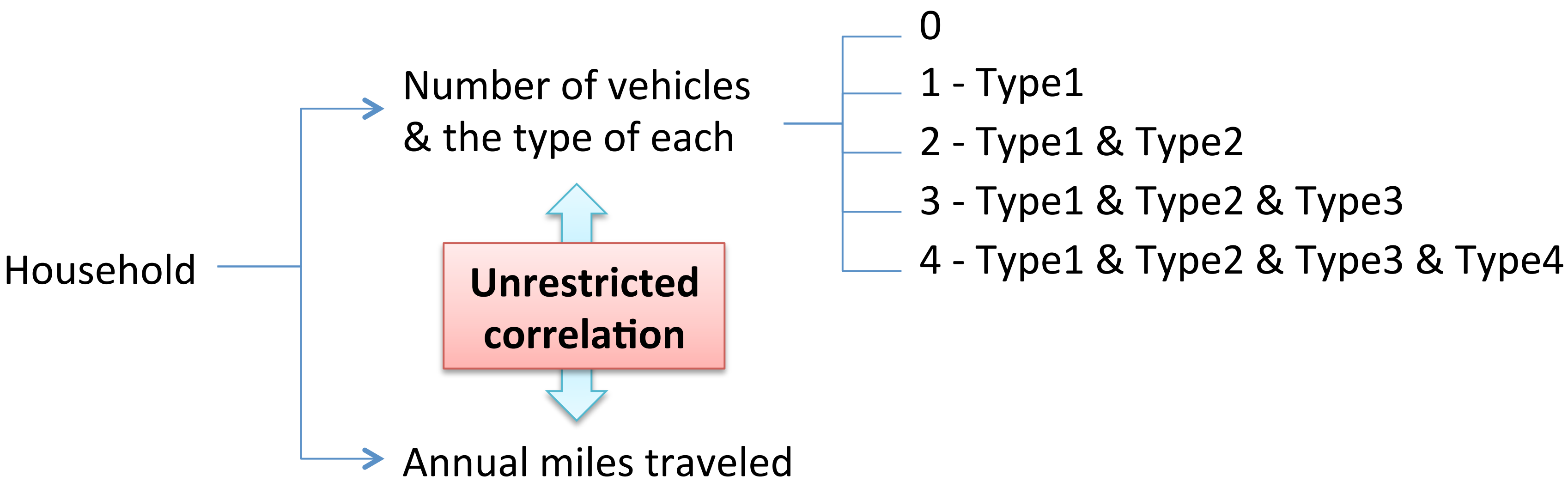
- Examine and compare the performance of ordered and unordered structures in discrete-continuous models.

- Apply the framework and develop the national models of vehicle ownership and use.

- Investigate the effects of improved public transportation services on household vehicle ownership and use

Framework of the Integrated Discrete Continuous Model

- The discrete choices:
  - Number of vehicles in the household
  - The type choice of each vehicle in the household
- The continuous choice:
  - Annual miles traveled of the household


Household

Number of vehicles & the type of each

Unrestricted correlation

Annual miles traveled

0

1 - Type1

2 - Type1 & Type2

3 - Type1 & Type2 & Type3

4 - Type1 & Type2 & Type3 & Type4

Unordered Discrete-Continuous Model

- The household is assumed to be rational and makes the *vehicle holding and type choices* that maximize its utility.

$$U_0 = \epsilon_0$$
$$U_1 = V_1 + \lambda V_{t_1|1} + \epsilon_1$$
$$U_2 = V_2 + \lambda V_{t_2|2} + \epsilon_2$$
$$\dots$$
$$U_k = V_k + \lambda V_{t_k|k} + \epsilon_k$$

$$P_{t_k|k} = \frac{exp(V_{t'_k|k})}{\sum_{t_k} exp(V_{t_k|k})}$$

$$J_k = ln \sum_{t_k} exp(V_{t_k|k})$$

$$\longrightarrow$$

$$U_0 = \epsilon_0$$
$$U_1 = X_1^T \beta_1 + J_1 \lambda \boxed{+ \epsilon_1}$$
$$U_2 = X_2^T \beta_2 + J_2 \lambda \boxed{+ \epsilon_2}$$
$$\dots$$
$$U_k = X_k^T \beta_k + J_k \lambda \boxed{+ \epsilon_k}$$

- The continuous choice *annual miles traveled* is in a linear form:

$$Y_{reg} = X_{reg}^T \beta_{reg} \boxed{+ \epsilon_{reg}}$$

Unordered Discrete-Continuous Model (Con't)

- The integrated discrete-continuous model:

$$(\tilde{\epsilon}_{1y}, \tilde{\epsilon}_{2y}, \ldots, \tilde{\epsilon}_{ky}, \epsilon_{reg}) \sim MN(0, \Sigma_{k+1})$$

$$P(Y, Y_{reg}) = \boxed{P(Y_{reg})}\boxed{P(Y|Y_{reg})}$$

$$P(Y_{reg}) = \phi(y_{reg}|X_{reg}^T \beta_{reg}, \sigma^2)$$

$$P(Y|Y_{reg}) = \int_{R^k} I(\tilde{V}_{jy} + \tilde{\epsilon}_{jy} < 0 \quad \forall j \neq y)\varphi(\tilde{\epsilon}_y)d\tilde{\epsilon}_y$$

$$\tilde{\epsilon}_y \sim \mathcal{N}(0 + \frac{\boldsymbol{\Sigma_{disc,reg}}}{\boldsymbol{\Sigma_{disc}}}(err - 0), \sigma_{reg}^2 - \frac{\boldsymbol{\Sigma_{reg,disc}\Sigma_{disc,reg}}}{\boldsymbol{\Sigma_{disc}}})$$

- Estimation methods:
  - Monte Carlo Simulation
  - Numerical Computation (Genz, 1992)

Ordered Discrete-Continuous Model

- The ordered response structure uses latent variables to represent the vehicle ownership propensity of the household.

$$y_d = X_d^T \beta_d + \epsilon_d$$

- The number of vehicles holding by the household (Y) is determined by the value of latent variable y_d, specifically:

$$\begin{array}{lll} Y = 0 & \text{if} & y_d < \alpha_1 \\ Y = 1 & \text{if} & \alpha_1 < y_d < \alpha_2 \\ Y = 2 & \text{if} & \alpha_2 < y_d < \alpha_3 \\ \dots & & \\ Y = k-1 & \text{if} & \alpha_{k-1} < y_d < \alpha_k \\ Y = k & \text{if} & \alpha_k < y_d \end{array}$$

- The error terms follow a bivariate normal distribution:

$$(\epsilon_d, \epsilon_r) \sim BN(0, \Sigma)$$

$$P(Y_d, Y_r) = P(Y_r)\boxed{P(Y_d|Y_r)}$$

Model Comparisons

- Objectives:
  - Compare the unordered and ordered discrete continuous models
  - Compare two estimation methods for the unordered discrete continuous model
- Data sources:
  - 2009 National Household Travel Survey (NHTS) data – 1420 observations in the Washington D.C. Metropolitan area
  - Vehicle characteristics
- Choice set:
  - Vehicle holding: 0, 1, 2, 3 and 4 car(s)
  - Vehicle type: 120 alternatives for the type choice of each vehicle (12 classes x 10 vintages)
  - Vehicle usage: annual miles traveled

Model Comparisons (Con't)

	Model 1		Model 2		Model 3		Model 4	
Variable	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Dependent variable: Number of cars								
logsum		0.012	**0.514**					
constant								
1 car								
2 cars								
3 cars								
4+ cars								
income					**0.086**	0.006		
1 car	**-0.051**	0.011	**-0.101**	0.020			**0.151**	0.010
2 cars	**0.056**	0.006	**0.692**	0.072			**0.395**	0.029
3 cars	**0.105**	0.010	**0.745**	0.077			**0.429**	0.028
4+ cars	**0.111**	0.012	**0.693**	0.074			**0.327**	0.026
num. of drivers					**0.608**	0.057		
1 car	**-0.010**	0.007	-0.304	0.236			-0.048	0.027
2 cars	**3.223**	0.079	**9.236**	0.214			**2.111**	0.215
3 cars	**4.041**	0.102	**10.167**	0.190			**1.742**	0.086
4+ cars	**4.432**	0.092	**10.120**	0.165			**3.314**	0.142
gender (female)					**-0.235**	0.063		
1 car	**-0.129**	0.551	-0.063	0.249			-0.054	0.068
2 cars	**-0.874**	0.054	**-3.434**	0.213			**-0.732**	0.245
3 cars	**-0.928**	0.073	**-3.605**	0.211			**-0.854**	0.281
4+ cars	**-0.885**	0.059	**-3.667**	0.194			**-2.208**	0.360
urban size					**-0.032**	0.013		
1 car	**0.077**	0.035	-0.109	0.058			-0.013	0.028
2 cars	**-0.120**	0.074	-0.270	0.178			0.103	0.277
3 cars	**-0.199**	0.093	-0.354	0.186			**-0.038**	0.018
4+ cars	**-0.201**	0.084	**-0.406**	0.183			**-0.368**	0.06[illegible]

unordered discrete-continuou
model **with** simulation

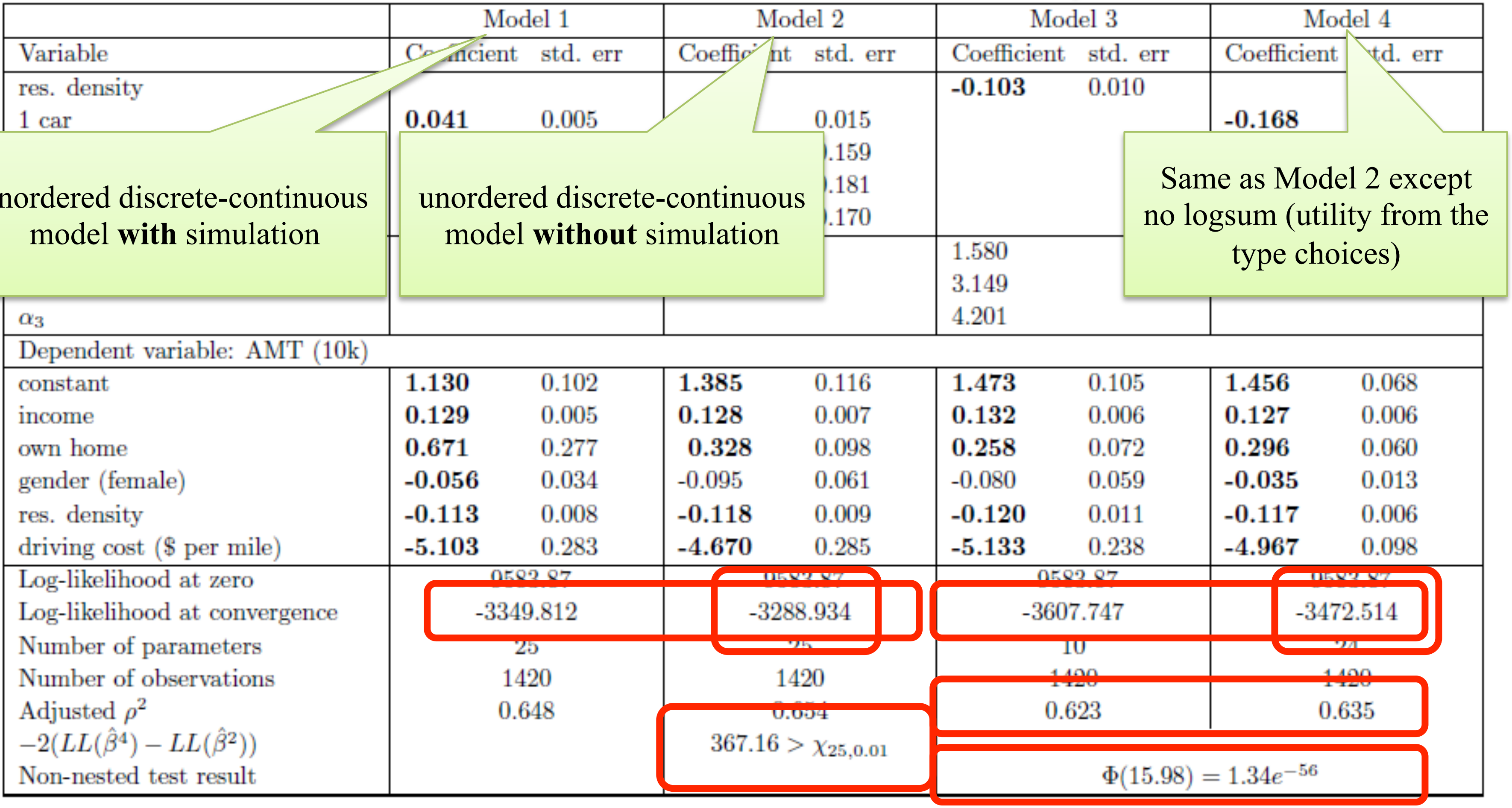
unordered discrete-continu
model **without** simulatio

Ordered discrete-
continuous model

s Model 2 except
m (utility from the
pe choices)

Model Comparisons (Con

Ordered discrete-continuous model



unordered discrete-continuous model **with** simulation

unordered discrete-continuous model **without** simulation

Same as Model 2 except no logsum (utility from the type choices)

	Model 1		Model 2		Model 3		Model 4	
Variable	Coefficient	std. err	Coefficient	std. err	Coefficient	std. err	Coefficient	std. err
res. density					**-0.103**	0.010		
1 car	**0.041**	0.005	[illegible]	0.015			**-0.168**	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			[illegible]	[illegible]
[illegible]					1.580			
[illegible]					3.149			
α_3					4.201			
Dependent variable: AMT (10k)								
constant	**1.130**	0.102	**1.385**	0.116	**1.473**	0.105	**1.456**	0.068
income	**0.129**	0.005	**0.128**	0.007	**0.132**	0.006	**0.127**	0.006
own home	**0.671**	0.277	**0.328**	0.098	**0.258**	0.072	**0.296**	0.060
gender (female)	**-0.056**	0.034	-0.095	0.061	-0.080	0.059	**-0.035**	0.013
res. density	**-0.113**	0.008	**-0.118**	0.009	**-0.120**	0.011	**-0.117**	0.006
driving cost ($ per mile)	**-5.103**	0.283	**-4.670**	0.285	**-5.133**	0.238	**-4.967**	0.098
Log-likelihood at zero	[illegible]		[illegible]		[illegible]		[illegible]	
Log-likelihood at convergence	-3349.812		-3288.934		-3607.747		-3472.514	
Number of parameters	25		25		10		24	
Number of observations	1420		1420		1420		1420	
Adjusted ρ^2	0.648		0.654		0.623		0.635	
$-2(LL(\hat{\beta}^4) - LL(\hat{\beta}^2))$			$367.16 > \chi_{25,0.01}$					
Non-nested test result					$\Phi(15.98) = 1.34e^{-56}$			

**Note: Model 1 is the unordered discrete-continuous model with simulation; Model 2 is the unordered discrete-continuous model with numerical computation; Model 3 is the ordered discrete-continuous model; Model 4 is the same as Model 2 except excluding the "logsum" variable, which make it comparable to Model 3.*

Model Estimations (Con't)

$$\hat{\Sigma}_1 = \begin{array}{ccccc|l} \text{1 car} & \text{2 cars} & \text{3 cars} & \text{4 cars} & \text{Mileage} & \\ 2.00 & -10.35 & -10.23 & -10.57 & -0.73 & \text{1 car} \\ -10.35 & 58.32 & 61.45 & 61.61 & 4.48 & \text{2 cars} \\ -10.23 & 61.45 & 68.59 & 67.08 & 5.23 & \text{3 cars} \\ -10.57 & 61.61 & 67.08 & 66.33 & 5.02 & \text{4 cars} \\ \mathbf{-0.73} & \mathbf{4.48} & \mathbf{5.23} & \mathbf{5.02} & 1.25 & \text{Mileage} \end{array}$$

$$\hat{\Sigma}_2 = \begin{array}{ccccc|l} \text{1 car} & \text{2 cars} & \text{3 cars} & \text{4 cars} & \text{Mileage} & \\ 2.00 & -10.34 & -10.24 & -10.57 & -0.73 & \text{1 car} \\ -10.34 & 58.26 & 61.44 & 61.57 & 4.46 & \text{2 cars} \\ -10.24 & 61.44 & 68.64 & 67.11 & 5.21 & \text{3 cars} \\ -10.57 & 61.57 & 67.11 & 66.34 & 5.00 & \text{4 cars} \\ \mathbf{-0.73} & \mathbf{4.46} & \mathbf{5.21} & \mathbf{5.00} & 1.25 & \text{Mileage} \end{array}$$

$$\hat{\Sigma}_3 = \begin{array}{cc|l} \text{\#cars} & \text{mileage} & \\ 1.00 & 0.50 & \text{\#cars} \\ \boxed{\mathbf{0.50}} & 1.56 & \text{mileage} \end{array}$$

$$\hat{\Sigma}_4 = \begin{array}{ccccc|l} \text{1 car} & \text{2 cars} & \text{3 cars} & \text{4 cars} & \text{Mileage} & \\ 2.00 & 4.73 & 7.92 & 7.73 & 1.58 & \text{1 car} \\ 4.73 & 17.70 & 26.12 & 30.45 & 3.68 & \text{2 cars} \\ 7.92 & 26.12 & 83.11 & 57.66 & 6.15 & \text{3 cars} \\ 7.73 & 30.45 & 57.66 & 57.94 & 5.96 & \text{4 cars} \\ \mathbf{1.58} & \mathbf{3.68} & \mathbf{6.15} & \mathbf{5.96} & 1.26 & \text{Mileage} \end{array}$$

Model Applications (Con't)

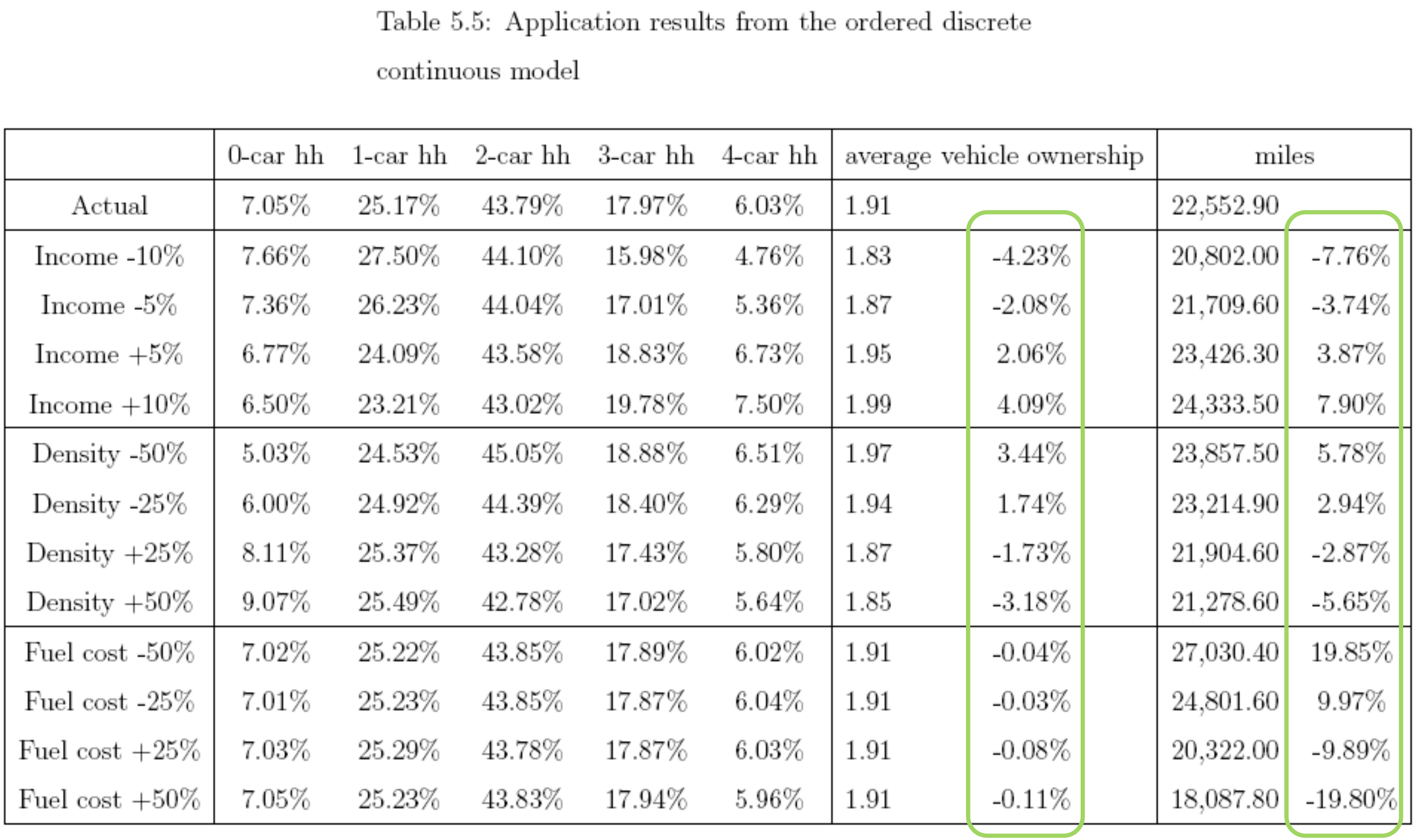


Table 5.5: Application results from the ordered discrete continuous model

	0-car hh	1-car hh	2-car hh	3-car hh	4-car hh	average vehicle ownership		miles	
Actual	7.05%	25.17%	43.79%	17.97%	6.03%	1.91		22,552.90	
Income -10%	7.66%	27.50%	44.10%	15.98%	4.76%	1.83	-4.23%	20,802.00	-7.76%
Income -5%	7.36%	26.23%	44.04%	17.01%	5.36%	1.87	-2.08%	21,709.60	-3.74%
Income +5%	6.77%	24.09%	43.58%	18.83%	6.73%	1.95	2.06%	23,426.30	3.87%
Income +10%	6.50%	23.21%	43.02%	19.78%	7.50%	1.99	4.09%	24,333.50	7.90%
Density -50%	5.03%	24.53%	45.05%	18.88%	6.51%	1.97	3.44%	23,857.50	5.78%
Density -25%	6.00%	24.92%	44.39%	18.40%	6.29%	1.94	1.74%	23,214.90	2.94%
Density +25%	8.11%	25.37%	43.28%	17.43%	5.80%	1.87	-1.73%	21,904.60	-2.87%
Density +50%	9.07%	25.49%	42.78%	17.02%	5.64%	1.85	-3.18%	21,278.60	-5.65%
Fuel cost -50%	7.02%	25.22%	43.85%	17.89%	6.02%	1.91	-0.04%	27,030.40	19.85%
Fuel cost -25%	7.01%	25.23%	43.85%	17.87%	6.04%	1.91	-0.03%	24,801.60	9.97%
Fuel cost +25%	7.03%	25.29%	43.78%	17.87%	6.03%	1.91	-0.08%	20,322.00	-9.89%
Fuel cost +50%	7.05%	25.23%	43.83%	17.94%	5.96%	1.91	-0.11%	18,087.80	-19.80%

Findings from the Model Comparisons

- The advantage of the ordered structure over the unordered is that it offers a closed mathematical form for the choice probabilities and does not require simulations for the estimation.
- However, the unordered discrete-continuous models always performs better in terms of goodness of fit statistics and forecasting capabilities when compared to ordered discrete-continuous models.
- In terms of the unordered discrete-continuous models, the estimation based on numerical computation provides less running time and better model goodness of fit than the estimation with Monte-Carlo simulation.
- This analysis confirms that the unordered structure is better suited for vehicle holding and use decisions in the context of joint discrete-continuous decisions.

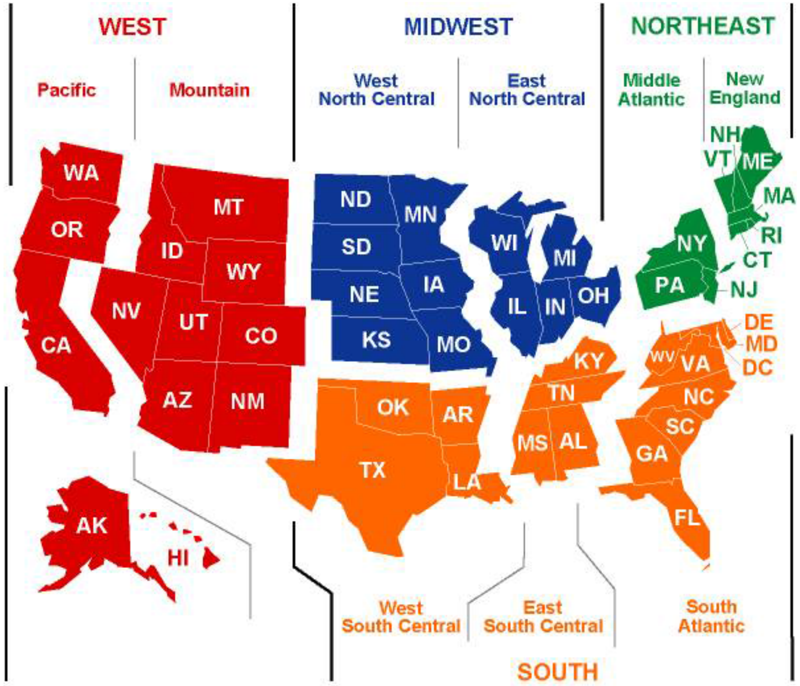
National Models of Vehicle Ownership and Use

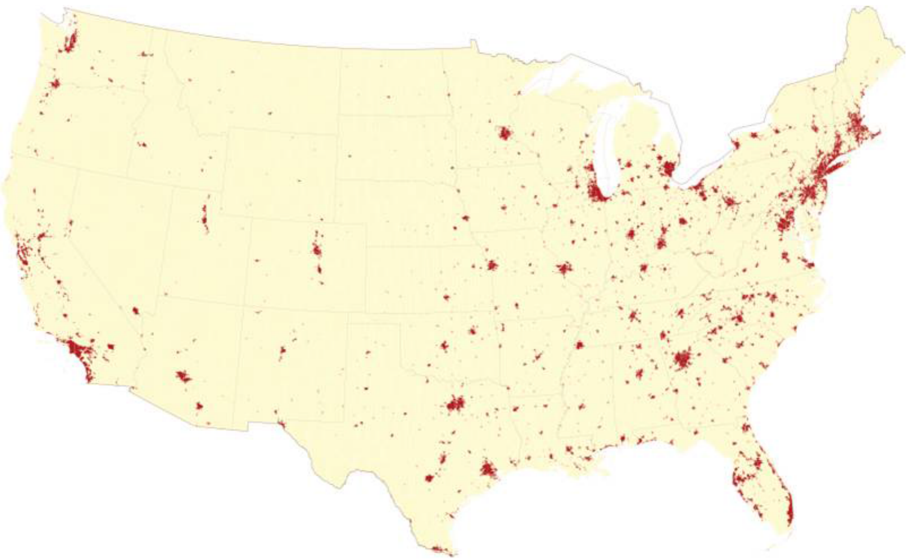
- This section develops a series of vehicle ownership and usage models for the entire United States, which is motivated by the lack of national vehicle ownership models in the literature, and the needs to determine vehicle/ driving demand in small areas with limited data availability.

- Hu et al. (2007) combined the 2001 NHTS data and 2000 census data to provide estimates of regional or local travel, including vehicle trips (VT), vehicle miles of travel (VMT), person trips (PT), and person miles of travel (PMT) by trip purpose and a number of demographics.

National Models of Vehicle Ownership and Use

- The models are estimated for four Census Regions (Northeast, Midwest, South and West;) and 3 area types (urbanized area, urban clusters and rural) with 2009 NHTS data.
  - Household income, household size, number of worker, has children, own home, residential density, driving cost





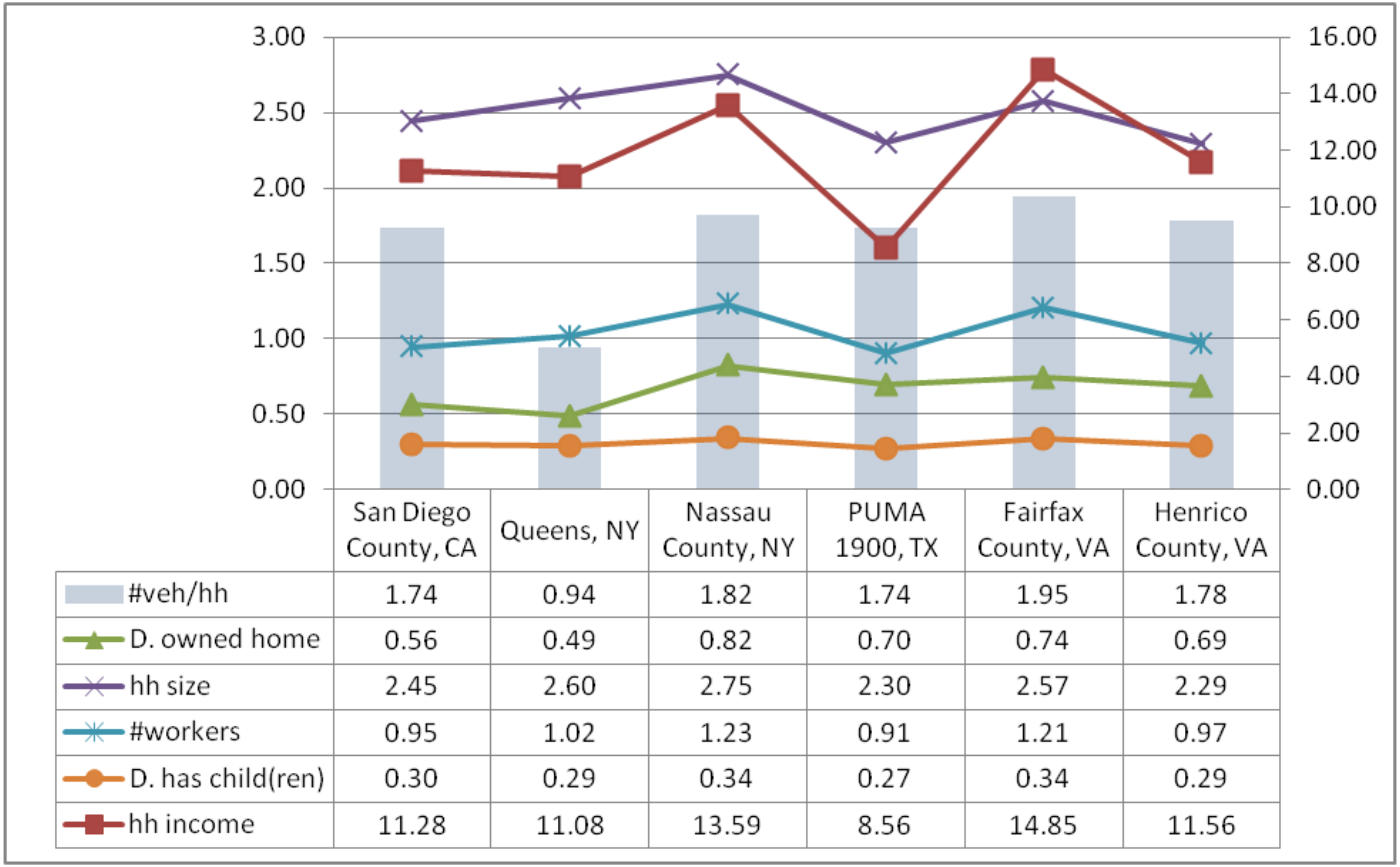
Application with ACS Data for Local Counties/Areas

- Then the models are applied to small areas using 2009 American Community Survey (ACS) Public Use Microdata Sample (PUMS) files.
  - San Diego County, CA – West, Urban
  - Queens, NY – Northeast, Urban
  - Nassau County, NY – Northeast, Urban
  - PUMA 1900, TX – South, Rural
    - Hill County, TX
    - Navarro County, TX
    - Limestone County, TX
    - Freestone County, TX
    - Navarro County, TX
  - Fairfax County, VA – South, Urban
  - Henrico Country, VA – South, Urban

Sample sizes of ACS and NHTS data

- San Diego County, CA
  - ACS : 11653 obs.
  - NHTS: 3712 obs.
- Queens, NY
  - ACS : 6985 obs.
  - NHTS: 251 obs.
- Nassau County, NY
  - ACS : 4875 obs.
  - NHTS: 265 obs.
- PUMA 1900, TX
  - ACS : 894 obs.
  - NHTS: 93 obs.
- Fairfax, VA
  - ACS : 4033 obs.
  - NHTS: 205 obs.
- Henrico, VA
  - ACS : 1274 obs.
  - NHTS: 379 obs.

Basic statistics



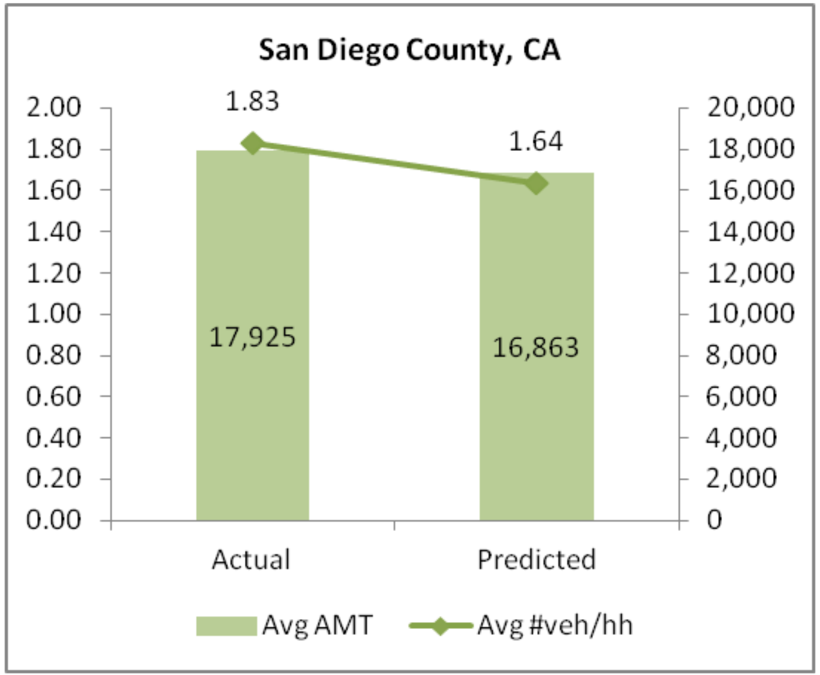
	San Diego County, CA	Queens, NY	Nassau County, NY	PUMA 1900, TX	Fairfax County, VA	Henrico County, VA
▨ #veh/hh	1.74	0.94	1.82	1.74	1.95	1.78
▲ D. owned home	0.56	0.49	0.82	0.70	0.74	0.69
× hh size	2.45	2.60	2.75	2.30	2.57	2.29
✳ #workers	0.95	1.02	1.23	0.91	1.21	0.97
● D. has child(ren)	0.30	0.29	0.34	0.27	0.34	0.29
■ hh income	11.28	11.08	13.59	8.56	14.85	11.56

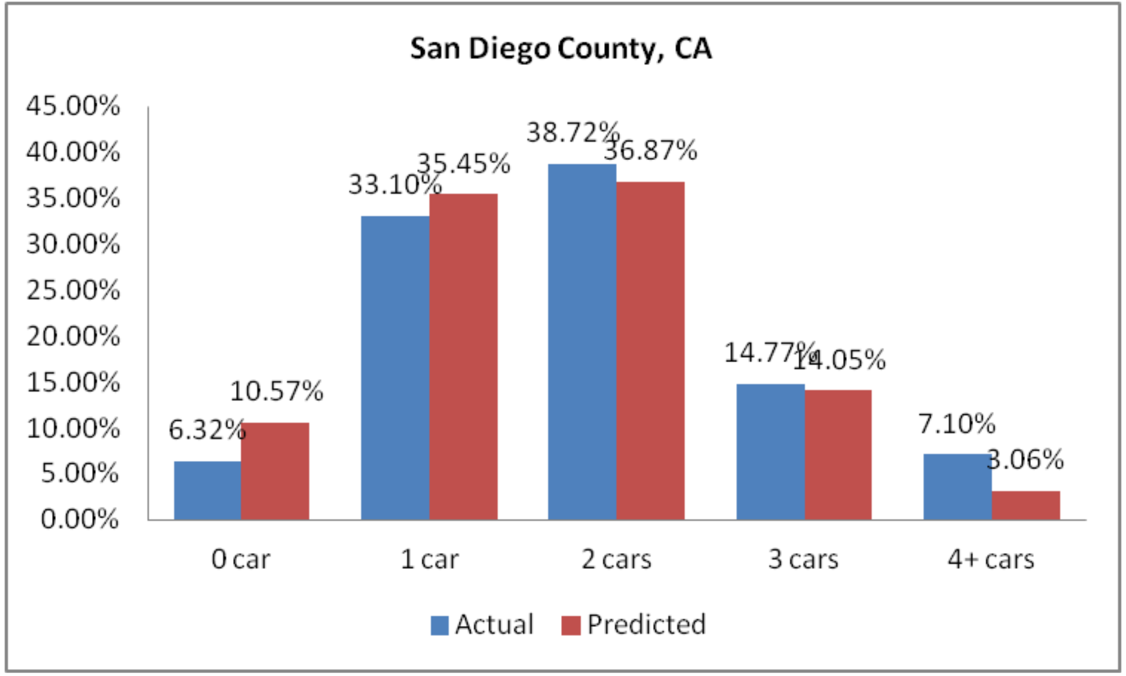
Population density (from U.S. Census)

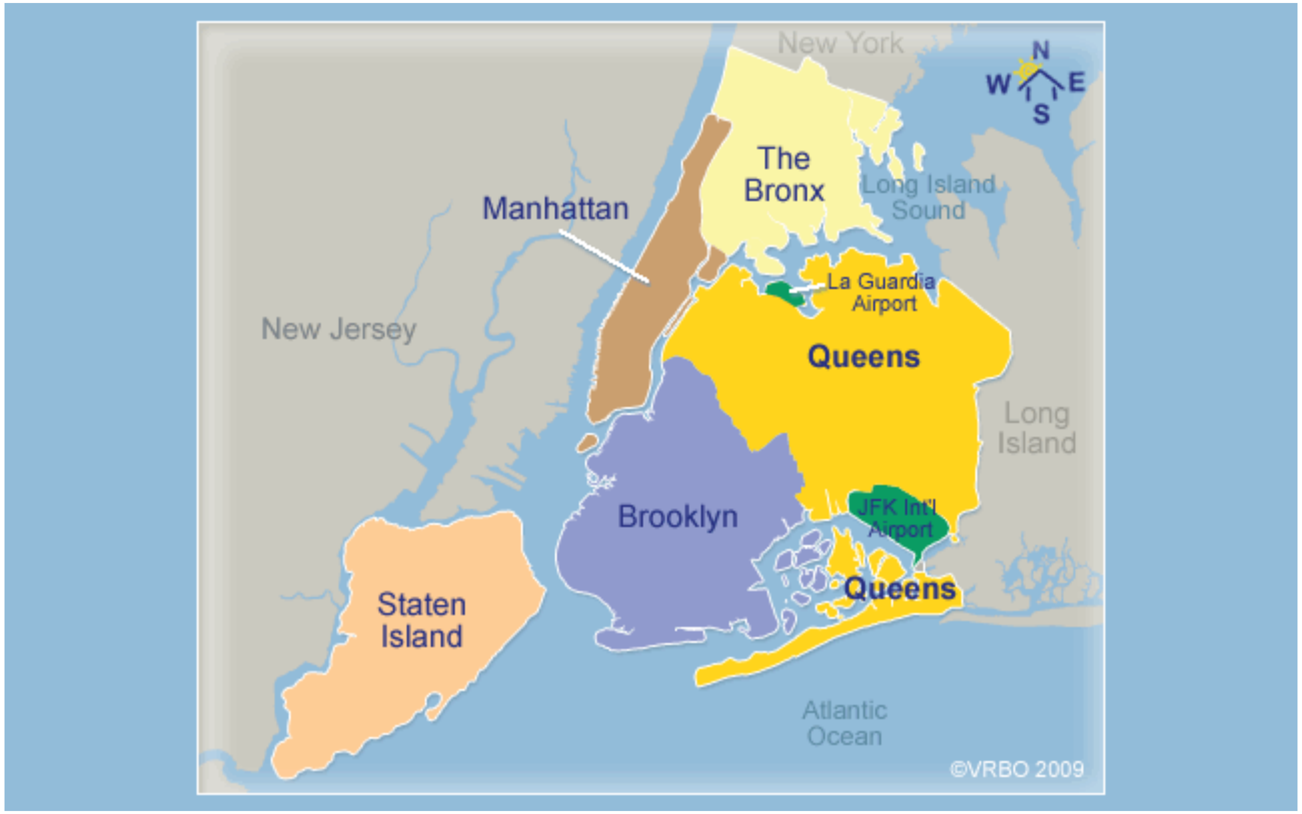
- San Diego County, CA
  - Population density: 680/sq mi
- Queens, NY
  - Population density: 21,116/sq mi
- Nassau County, NY
  - Population density: 4,669/sq mi
- PUMA 1900, TX
  - Population density: 18-57/sq mi
- Fairfax, VA
  - Population density: 2,738.5/sq mi
- Henrico, VA
  - Population density: 1,323/sq mi

- San Diego County, CA
  - Total area: 4,525.52 sq mi
  - Total population: 3,095,313 (2010 Census)
  - Population density: 680/sq mi

Prediction results for San Diego County, CA

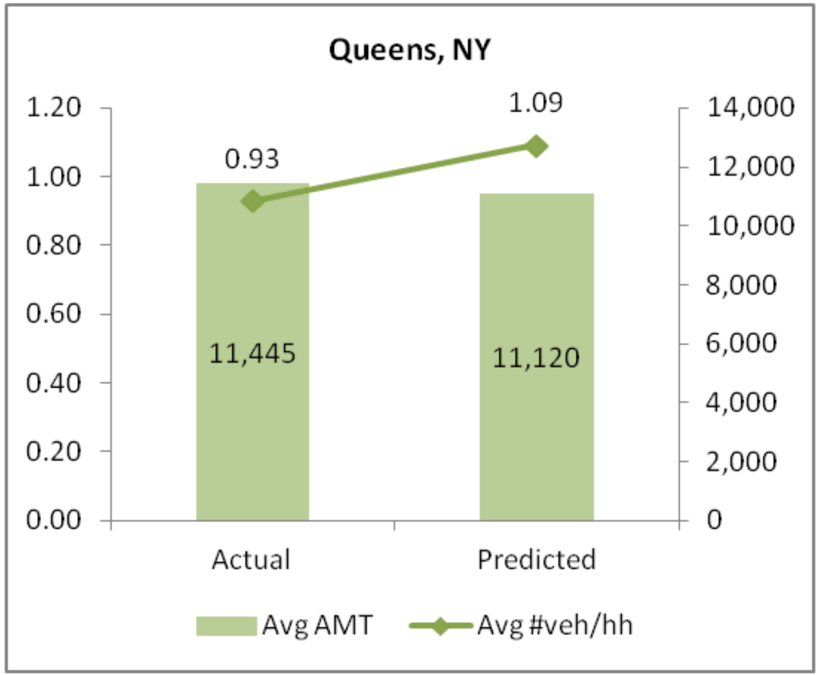


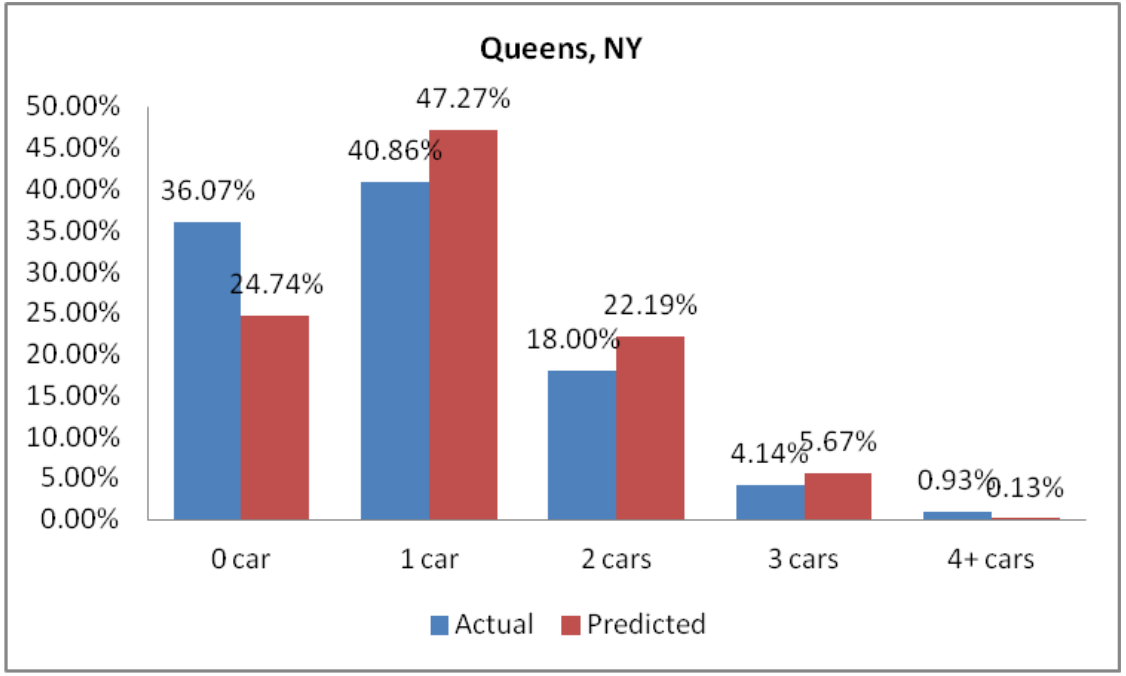




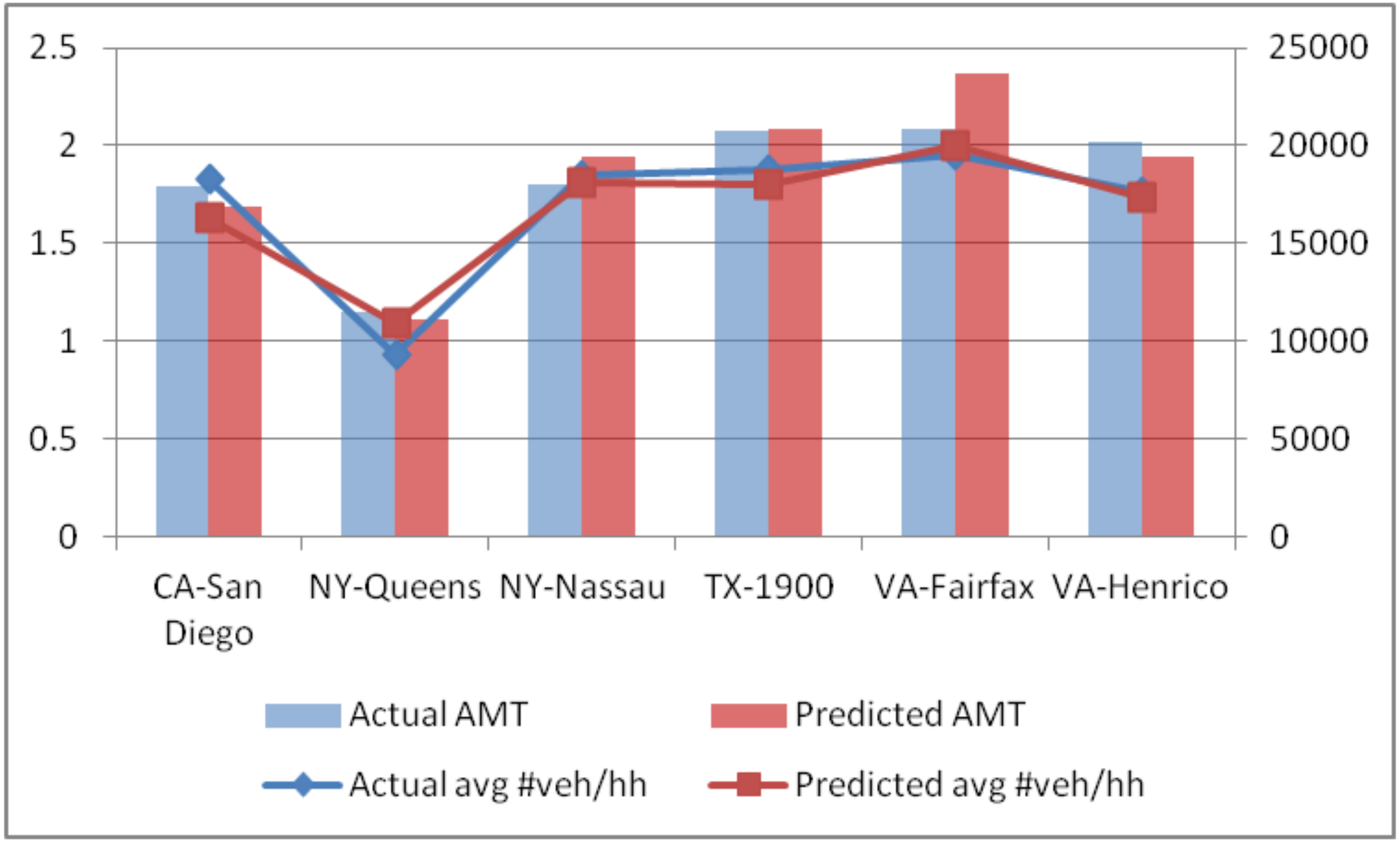
- Queens, NY
  - Total area: 178.28 sq mi
  - Total population: 2,272,771 (2010 Census)
  - Population density: 21,116/sq mi

Prediction results for Queens, NY





Summary of the Application Results



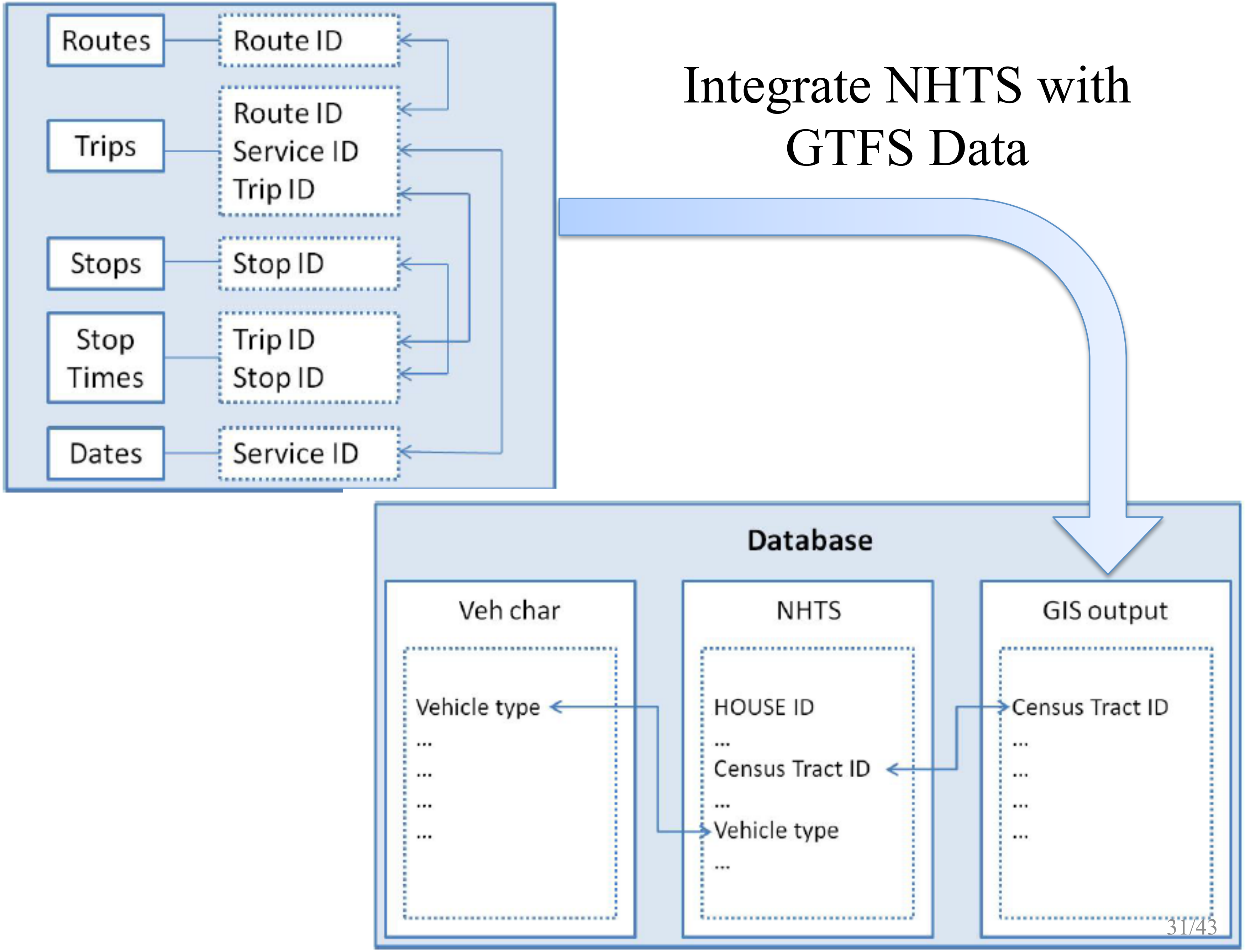
Findings from the National Models

- The system of models are estimated using 2009 NHTS data for each combination of four regions (Northeast, Midwest, South and West) and three area types (urban, suburban and rural).
- The system of models is applied to six randomly selected counties/areas using the 2009 ACS PUMS data.
- The results from the model applications demonstrate the ability of the national models in providing accurate estimates for various city/area types.
- The national models are valuable both for the national level (to evaluate federal policies) and small areas (that lack of local household travel survey data).
- The results also demonstrate that the integrated discrete-continuous framework has good prediction capabilities in modeling household vehicle ownership decisions.

Measuring Transit Service Impacts on Vehicle Ownership and Usage

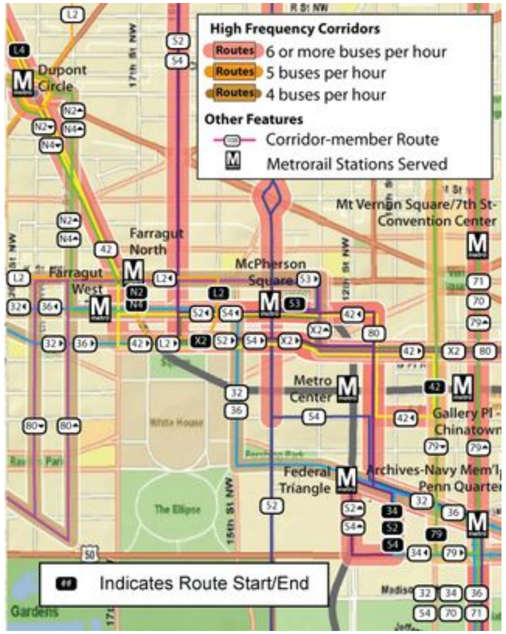
- Recent studies provide evidence that good public transportation might encourage people to reduce vehicle ownership and use. However, very few studies use advanced quantitative methods to investigate the relationship between public transit service and vehicle ownership and use.
- Study area: Washington D.C. Metropolitan area
- Data sources:
  - 2009 National Household Travel Survey (NHTS) data with geographic reference (U.S. Census Tract level)
  - General Transit Feed Specification (GTFS) data was obtained from the Washington Metropolitan Area Transit Authority (WMATA).
  - U.S. Census TIGER/Line shapefiles

Integrate NHTS with GTFS Data



Data Geo-Processing and Data Integration (Census Tract level)

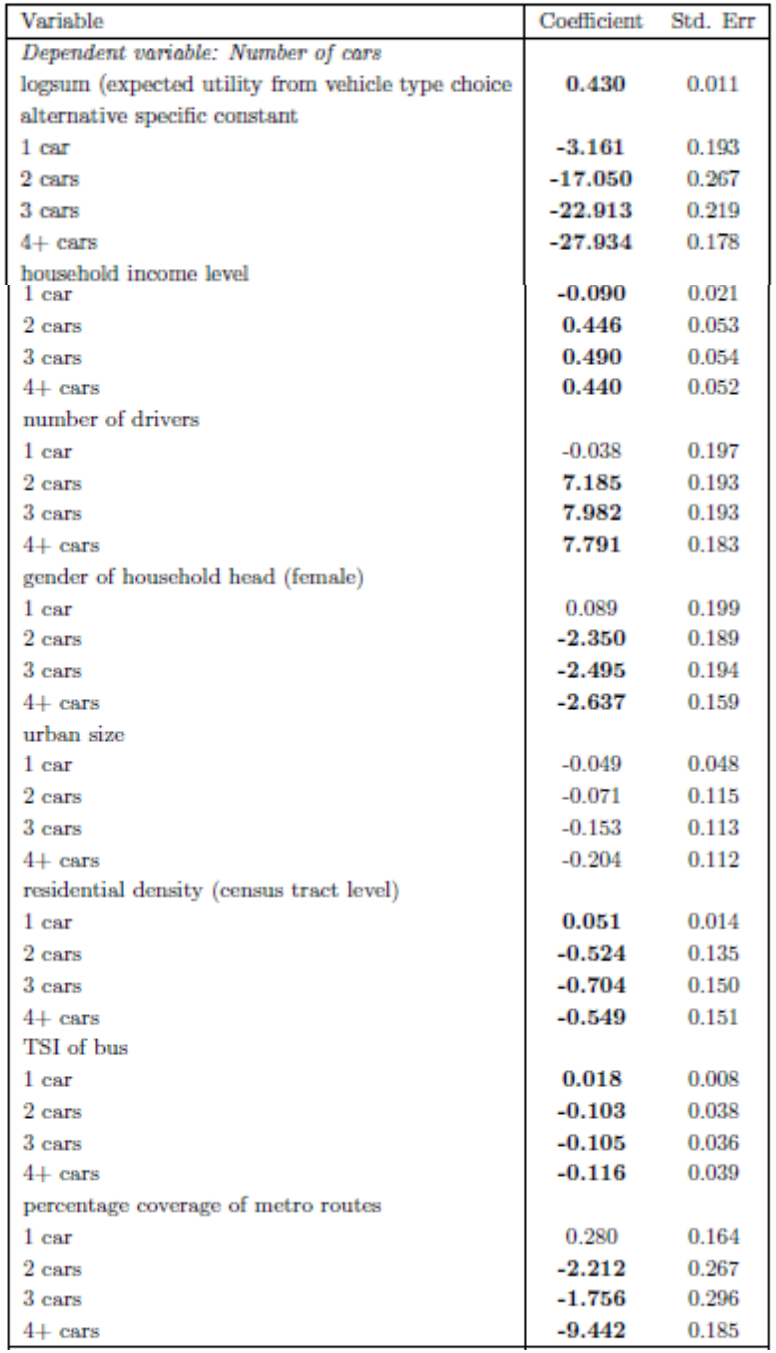
- Spatial measurements of transit service
  (main reference: Transit Capacity and Quality of Service Manual)
  - percentage of bus stops coverage,
  - percentage of metro routes coverage,
  - total length of bus routes,
  - total length of metro routes, and
  - total number of bus stops.



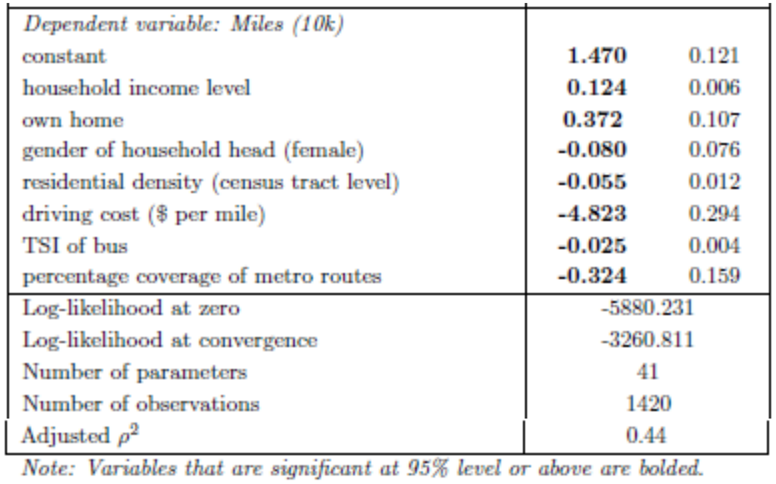
- Temporal measurements of bus service
  - the average duration
  - the average headway
- Transit service index (TSI) [Keller, 2012]

$$TSI = \frac{percent\ service\ coverage\ area}{average\ service\ headway} \times service\ duration$$

Estimation results



Variable	Coefficient	Std. Err
Dependent variable: Number of cars		
logsum (expected utility from vehicle type choice	**0.430**	0.011
alternative specific constant		
1 car	**-3.161**	0.193
2 cars	**-17.050**	0.267
3 cars	**-22.913**	0.219
4+ cars	**-27.934**	0.178
household income level		
1 car	**-0.090**	0.021
2 cars	**0.446**	0.053
3 cars	**0.490**	0.054
4+ cars	**0.440**	0.052
number of drivers		
1 car	-0.038	0.197
2 cars	**7.185**	0.193
3 cars	**7.982**	0.193
4+ cars	**7.791**	0.183
gender of household head (female)		
1 car	0.089	0.199
2 cars	**-2.350**	0.189
3 cars	**-2.495**	0.194
4+ cars	**-2.637**	0.159
urban size		
1 car	-0.049	0.048
2 cars	-0.071	0.115
3 cars	-0.153	0.113
4+ cars	-0.204	0.112
residential density (census tract level)		
1 car	**0.051**	0.014
2 cars	**-0.524**	0.135
3 cars	**-0.704**	0.150
4+ cars	**-0.549**	0.151
TSI of bus		
1 car	**0.018**	0.008
2 cars	**-0.103**	0.038
3 cars	**-0.105**	0.036
4+ cars	**-0.116**	0.039
percentage coverage of metro routes		
1 car	0.280	0.164
2 cars	**-2.212**	0.267
3 cars	**-1.756**	0.296
4+ cars	**-9.442**	0.185



Dependent variable: Miles (10k)		
constant	**1.470**	0.121
household income level	**0.124**	0.006
own home	**0.372**	0.107
gender of household head (female)	**-0.080**	0.076
residential density (census tract level)	**-0.055**	0.012
driving cost ($ per mile)	**-4.823**	0.294
TSI of bus	**-0.025**	0.004
percentage coverage of metro routes	**-0.324**	0.159
Log-likelihood at zero	-5880.231	
Log-likelihood at convergence	-3260.811	
Number of parameters	41	
Number of observations	1420	
Adjusted ρ^2	0.44	

Note: Variables that are significant at 95% level or above are bolded.

$$\hat{\Sigma} = \begin{pmatrix} 2.00 & -7.51 & -7.43 & -7.54 & -0.68 \\ -7.51 & 29.22 & 30.68 & 30.41 & 2.92 \\ -7.43 & 30.68 & 35.04 & 33.67 & 3.55 \\ -7.54 & 30.41 & 33.67 & 32.72 & 3.34 \\ \mathbf{-0.68} & \mathbf{2.92} & \mathbf{3.55} & \mathbf{3.34} & 1.23 \end{pmatrix}$$

Application results

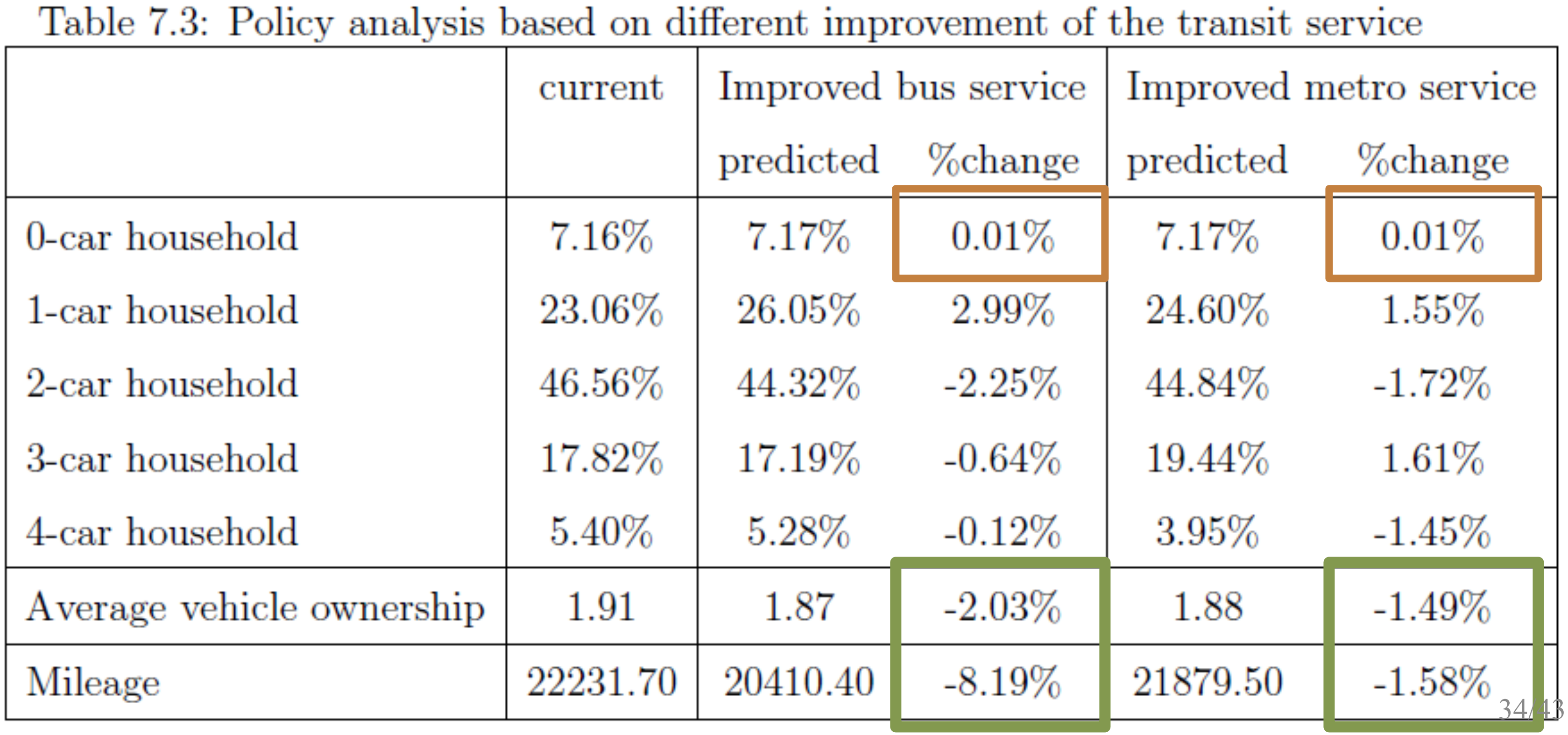


Table 7.3: Policy analysis based on different improvement of the transit service

	current	Improved bus service		Improved metro service	
		predicted	%change	predicted	%change
0-car household	7.16%	7.17%	0.01%	7.17%	0.01%
1-car household	23.06%	26.05%	2.99%	24.60%	1.55%
2-car household	46.56%	44.32%	-2.25%	44.84%	-1.72%
3-car household	17.82%	17.19%	-0.64%	19.44%	1.61%
4-car household	5.40%	5.28%	-0.12%	3.95%	-1.45%
Average vehicle ownership	1.91	1.87	-2.03%	1.88	-1.49%
Mileage	22231.70	20410.40	-8.19%	21879.50	-1.58%

- improved bus services: in this hypothetical scenario every census tract zone has at least 50% bus stop coverage, 15-minute average headway and 6 peak hours duration (6:30AM - 9:30AM and 3:30PM - 6:30PM).
- improved bus services: In the improved metrorail service scenario, the core area of Washington Metropolitan area (urban size greater than 1 million) has at least 50 percent metro route coverage.

Findings from the Transit Impacts Analysis

- This section estimates a discrete-continuous model for the Washington D.C. Metropolitan Area and analyzes the impact of improved bus and metro services on household ownership and use decisions in that area.
- The 2009 National Household Travel Survey data and the General Transit Feed Specification data are integrated, and then both spatial and temporal measurements of transit services are created on the Census Tract level.
- The results show that improved transit is a significant factor in household vehicle ownership choices and that the proposed methods are able to effectively predict changes in vehicle ownership and usage with respect to the transit improvements.

Conclusions

- An integrated discrete continuous choice model is developed to simultaneously estimate the household choices on vehicle ownership (discrete), the types (discrete) and annual mileage traveled (continuous).
  - The model is able to include a large number of alternatives in both the vehicle holding and the vehicle type choices.
  - The model allows unrestricted correlations of the unobserved factors between the discrete and continuous parts.
  - The model accommodates flexible specifications.
  - There is no budget constraint in the mileage traveled.
  - The model can be applied for policy analysis.
  - The model can generate reasonable estimates of the coefficients.
  - The covariance matrix well explains the correlations between the unobserved factors from the utilities of the discrete choices and the demand function of the continuous choice.
  - The non-simulation approach provides a better model fit.
  - The performance of the model would be improved if the information about vehicle type choice is included.

Conclusions (Con't)

- A comparison of unordered and ordered structures in discrete-continuous framework is conducted with operational data. The results show that the unordered discrete continuous model is more appropriate than the ordered discrete continuous model in estimating household vehicle ownership and usage decisions.
- A system of national models on household vehicle ownership choices is developed with National Household Travel Survey data and American Community Survey data. Applications for six randomly selected areas demonstrate that the models are able to produce accurate estimates.
- The model is further applied using geographic data to study the impacts of improved transit service on household vehicle ownership choices in the Washington D.C. metropolitan area. Results show that transit service variables are significant factors in household vehicle ownership choices and that the proposed methods are able to effectively predict changes in vehicle ownership and usage due to transit service improvements.

The Future

Background

- Discrete choice models are commonly used in transportation planning and modeling, but their theoretical basis and applications have been mainly developed in a static context.
- With the continuous and rapid changes in modern societies (i.e. introduction of advanced technologies, aggressive marketing strategies and innovative policies) it is more and more recognized by researchers in various disciplines that choice situations take place in a dynamic environment and that strong interdependencies exist among decisions made at different points in time.

Dynamics models in economics

- Dynamic discrete choice models have been firstly developed in economics and related fields.
- In dynamic discrete choice structural models, agents are forward looking and maximize expected inter-temporal payoffs.
- The consumers get to know the rapidly evolving nature of product attributes within a given period of time and different products are supposed to be available on the market.
- As a result, a consumer can either decide to buy the product or to postpone the purchase at each time period. This dynamic choice behavior has been treated in a series of different research studies.

Review of economics literature

- John Rust (1987) --- bus engine replacement, single agent, two options, one purchase, homogenous attributes of the products, infinite-horizon. Nested Fixed Point method to estimate.
- Oleg Melnikov (2000) --- printer machine demand one purchase, differentiated durable products, homogenous consumers.
- Szabolcs LŐrincz (2005) --- computer servers demand, persistency effects, choice between using the original product and upgrading its format (operating systems). Dynamic nested logit model.
- Juan Esteban Carranza (2006) --- digital camera demand, heterogeneity over consumers' preferences and dynamics of quality.
- Gowrisankaran and Rysman (2007) --- digital camcorder, repeat purchases, heterogeneous consumers and differentiated products.

Model formulation

Dynamic, regenerative, optimal stopping problem

Consumer *i* state at time *t*

$$S_{it} = \{0,1\} \quad \begin{cases} 0 \text{ if } i \text{ is in the market;} \\ 1 \text{ otherwise.} \end{cases}$$

In each time period consumer *i* in status $S_{it} = 0$ has two options:

(a) to buy one of the products $j \in \Im_t$ or

(b) to postpone

If (a) the consumer *i* obtains a terminal payoff u_{ijt}

If (b) is chosen the consumer obtains a one period payoff c_{it} .

One period pay off

$$c(x_{it}, q_{it}; \theta_i, \alpha_i)$$

x_{it}, a vector of attributes for *i* at *t*, e.g. gender, education, professional status, income.

q_{it}, a vector of characteristics of current vehicle owned by *i*, e.g. age, mileage, purchase price, etc.

θ_i, α_i, are parameters for x_{it} and q_{it}.

Terminal payoff

$$u_{ijt} = u\left(x_{it}, d_j, y_{jt}, \theta_i, \gamma_i, \lambda_i, \varepsilon_{ijt}\right)$$

x_{it} **is a vector of static individual attributes** (e.g. age, income, education) and is the θ_i related parameter;

d_j **is a vector of static product attributes** (e.g. vehicle size) and γ_i is the related parameter;

y_{jt} **is a vector of dynamic attributes** (e.g. energy cost per mile, purchase cost, environment incentives), λ_i is the related parameter;

ε_{ijt} **is a random utility** component (i.i.d. GEV)

$$u_{jt} = \delta_{jt} + \varepsilon_{jt}$$

δ_{jt} is the mean utility.

Each time period, the consumer decides to buy or postpone

$$D\left(v_{it}, c_{it}\right) = \max\left\{v_{it}, c_{it} + \beta E\left[D\left(v_{i,t+1}\right)\right]\right\}$$

where: $v_t = \max_{j \in \Im_t} u_{jt}$

Hypothesis:

c_{it} is the payoff when postponing

β is time period when consumer decides to buy (set 1)

$E_t[\cdot] = E[\cdot \mid I_t]$ expected utility

(Based on Bellman equation):

$$D(u_{i1t}, \ldots u_{iJt}, c_{it}) = \max_{\tau}\left[\sum_{k=t}^{\tau-1} \beta^{k-t} c_{it} + \beta^{\tau-t} E_t \max_{j \in J} u_{ij\tau}\right]$$

where:

τ is time period when consumer decides to buy

Industry evolution

The evolution of the industry is represented by a so called **random walk**; dynamic variable y_{jt} is supposed to follow a normal diffusion process, specified as a random walk with drift η_j

$$\begin{aligned} y_{j,t+1} &= \mu(y_{jt}) + L(y_{jt})\upsilon_{j,t+1} \\ &= \psi_j y_{jt} + \eta_j + L(y_{jt})\upsilon_{j,t+1} \end{aligned}$$

υ_{jt} *(j=1,…,J, t = 1,…,T)* are i.i.d. multivariate standard normal random vectors.

L is the Cholesky factor of the variance-covariance matrix

$$L(y_{jt})L(y_{jt})^T = \Sigma(y_{jt})$$

Utility formulation

The previous equation becomes:

$$D\left(v_{it},c_{it}\right)=\max\left\{v_{it},c_{it}+\beta E[D\left(v_{i,t+1}(y_{t+1},c_{i,t+1})\right)\mid y_t]\right\}$$

This is standard optimal stopping problem. The stopping set is given when:

$$T(y_{jt})=\left\{v_{it}\mid v_{it}\geq c_{it}+\beta E\left[D\left(\cdot\right)\mid y_t\right]\right\}$$

Reservation utility

$$W\left(y_t\right)=c_{it}+\beta E[D\left(v_{i,t+1}(y_{t+1},c_{i,t+1})\right)\mid y_t]$$

Here,

$$y_t=(y_{1t},\ldots y_{Jt})$$

Demand structure

Probability of postponing until next period:

$$\pi_{i0t}\left(y_t\right) = P\left[v_{it} \leq W\left(y_t\right)\right] = P\left[postpone \mid s_{it} = 0, y_t\right] = F_\gamma\left(W\left(y_t\right), y_t\right)$$

Product adoption rate:

$$h\left(y_t\right) = P\left[buy \mid s_{it} = 0, y_t\right] = 1 - \pi_{ot}\left(y_t\right)$$

$$\pi_{ijt}\left(y_{jt}\right) = P\left[U_{ijt} \geq U_{ikt} \quad \forall\ k \neq j \ \ u_{ijt} \geq W\left(y_{it}\right)\right]$$

$$= h(r_t)\frac{\exp(\delta_{jt})G_j(e^{\delta_{j1}},\ldots,e^{\delta_{jt}})}{G(e^{\delta_{j1}},\ldots,e^{\delta_{jt}})}$$

Estimation methodology

The parameters estimation can therefore be formulated as a traditional maximum likelihood problem:

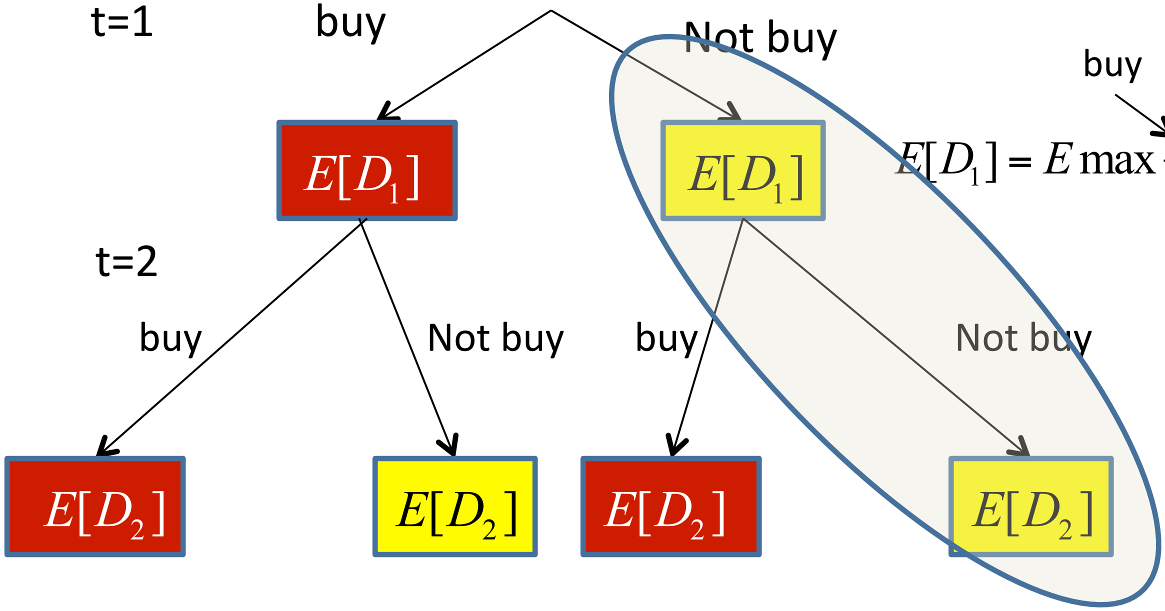
$$\max_{h} LL(h) = \sum_{i=1}^{M} \sum_{t=1}^{T} \ln P_{it} [\text{decision} \mid s_{it=0}].$$

Decisions include: buy a car of type *j*, not buy a car.

$$P_{it} = \{\pi_{i0t}, \pi_{ijt}\}$$

Scenario tree

At t=0 $W\left(y_0\right) = c_{i0} + \beta E[D_1]$



t=1

buy

Not buy

$E[D_1]$ $E[D_1]$

$$E[D_1] = E\max\{\overset{\text{buy}}{v_{i1}}, \overset{\text{Not buy}}{c_{i1} + E[D_2]}\}$$

t=2

buy

Not buy

buy

Not buy

$E[D_2]$ $E[D_2]$ $E[D_2]$ $E[D_2]$

$$E[D_2] = E\max\{v_{i2}, c_{i2} + E[D_3]\}$$

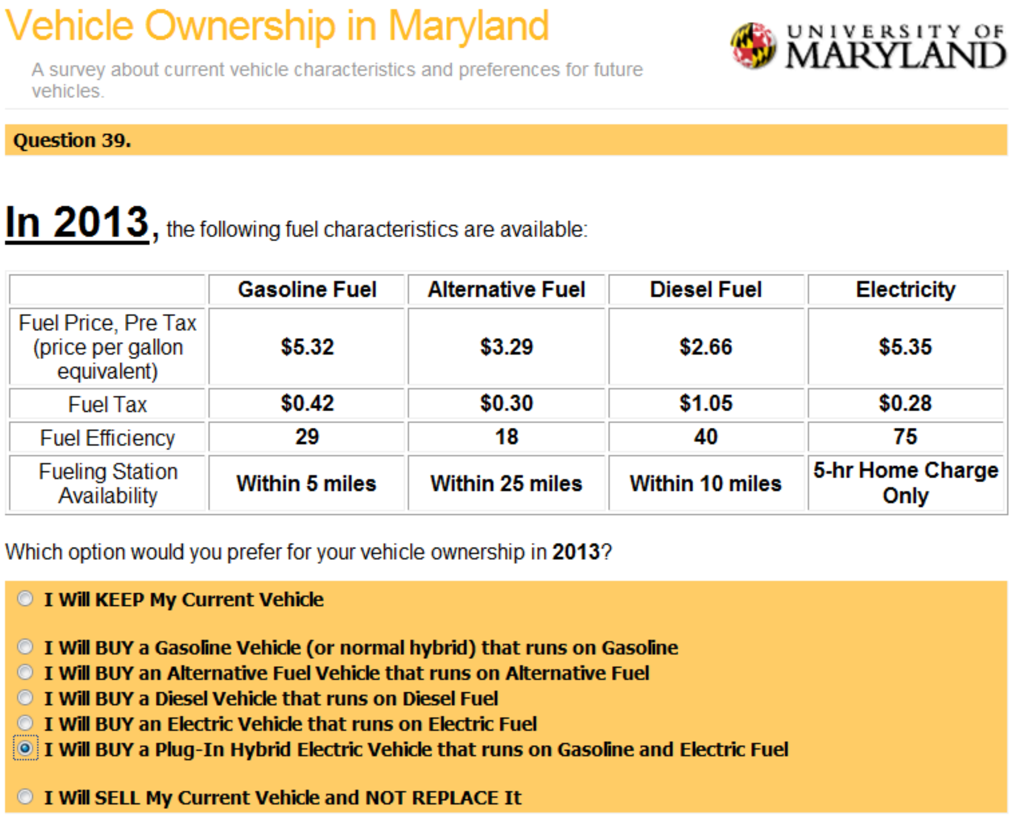
t=3

$$E[D_3] = 0$$

DDCM applied to carownership

- What effect will the following factors have on the vehicle marketplace over the next five years:
  - New vehicle technology
  - Improvements in existing vehicle technology
  - Greater availability of different energy sources
  - Rising fuel prices
  - Transportation and energy policy

Fuel Type Experiment



Vehicle Ownership in Maryland

A survey about current vehicle characteristics and preferences for future vehicles.

Question 39.

In 2013, the following fuel characteristics are available:

	Gasoline Fuel	Alternative Fuel	Diesel Fuel	Electricity
Fuel Price, Pre Tax (price per gallon equivalent)	**$5.32**	**$3.29**	**$2.66**	**$5.35**
Fuel Tax	**$0.42**	**$0.30**	**$1.05**	**$0.28**
Fuel Efficiency	**29**	**18**	**40**	**75**
Fueling Station Availability	**Within 5 miles**	**Within 25 miles**	**Within 10 miles**	**5-hr Home Charge Only**

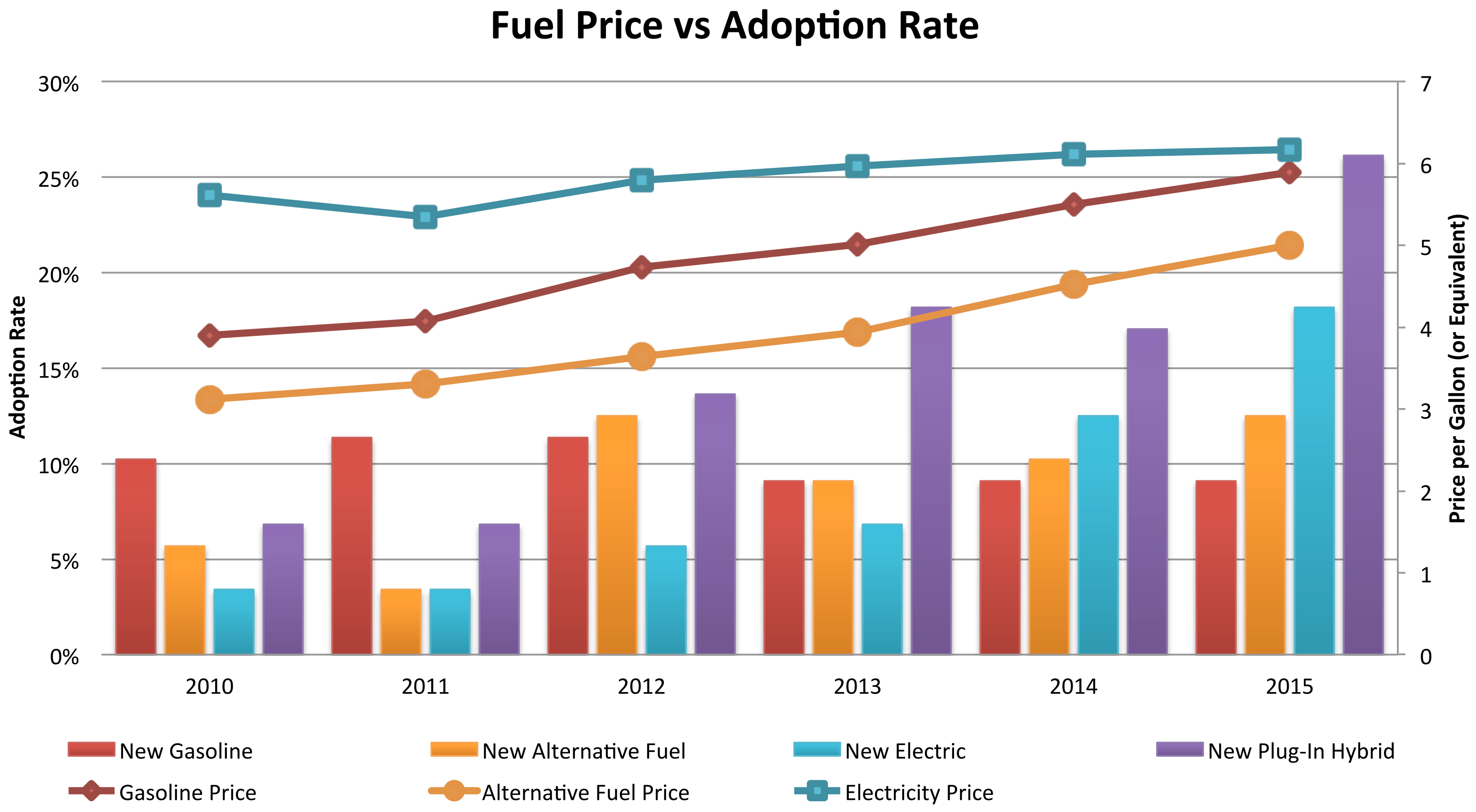
Which option would you prefer for your vehicle ownership in **2013**?

○ **I Will KEEP My Current Vehicle**

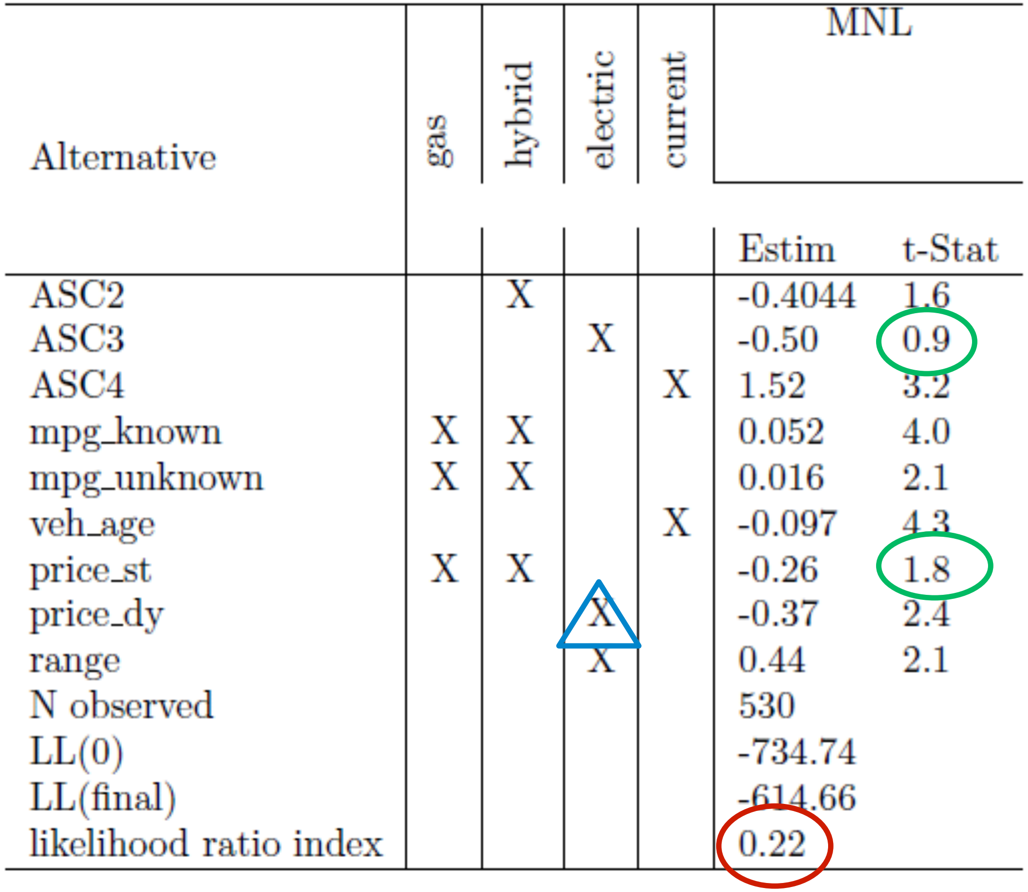
○ **I Will BUY a Gasoline Vehicle (or normal hybrid) that runs on Gasoline**
○ **I Will BUY an Alternative Fuel Vehicle that runs on Alternative Fuel**
○ **I Will BUY a Diesel Vehicle that runs on Diesel Fuel**
○ **I Will BUY an Electric Vehicle that runs on Electric Fuel**
● **I Will BUY a Plug-In Hybrid Electric Vehicle that runs on Gasoline and Electric Fuel**

○ **I Will SELL My Current Vehicle and NOT REPLACE It**

Results – Fuel Technology



Static Model- results

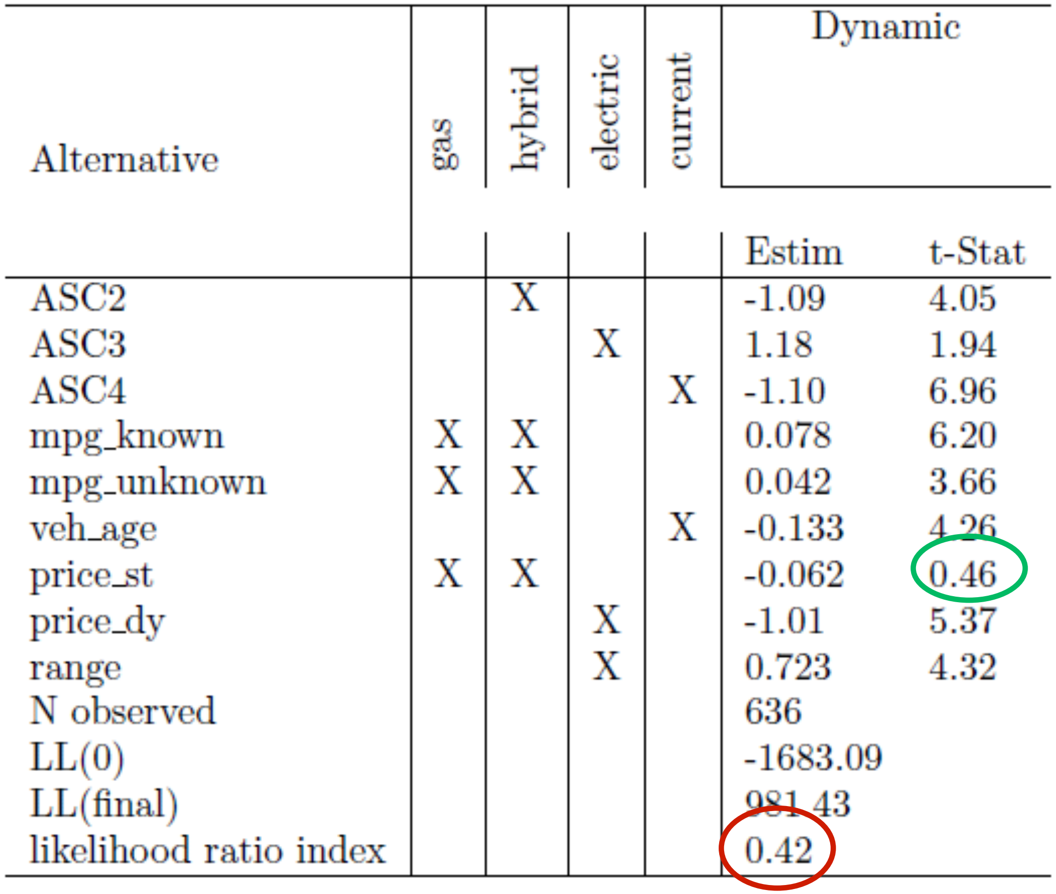


Alternative	gas	hybrid	electric	current	MNL	
					Estim	t-Stat
ASC2		X			-0.4044	1.6
ASC3			X		-0.50	0.9
ASC4				X	1.52	3.2
mpg_known	X	X			0.052	4.0
mpg_unknown	X	X			0.016	2.1
veh_age				X	-0.097	4.3
price_st	X	X			-0.26	1.8
price_dy			X		-0.37	2.4
range			X		0.44	2.1
N observed					530	
LL(0)					-734.74	
LL(final)					-614.66	
likelihood ratio index					0.22	

Dynamic model -results

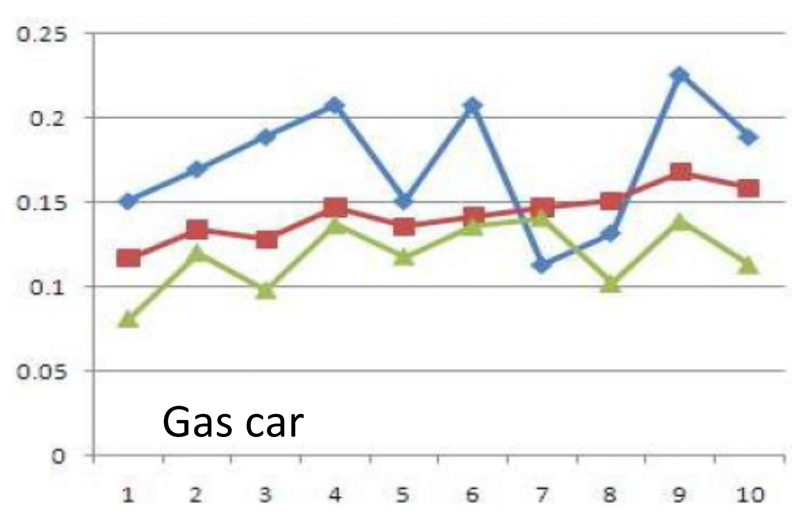
Choose electric car price as the dynamic variable

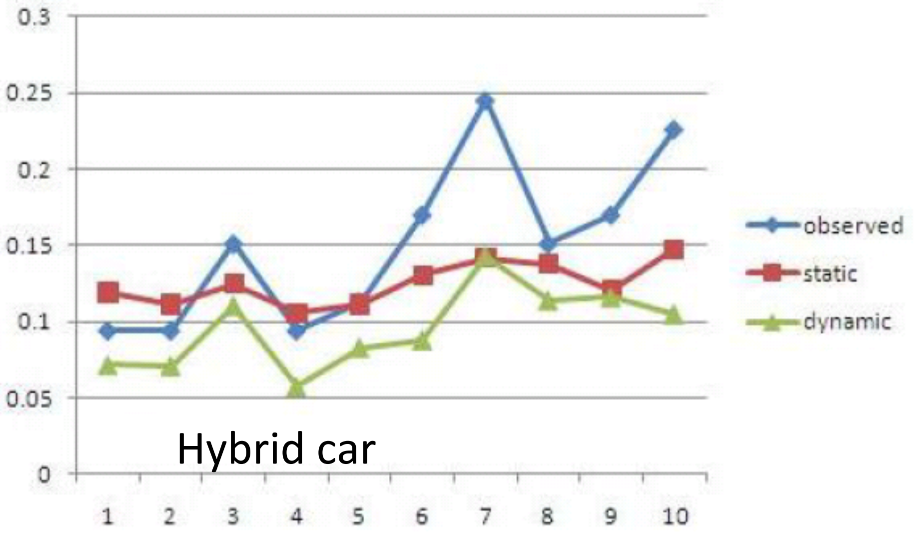
$$y_{j,t+1} = -0.103 \times y_{jt} + 2.617 + N(0,1.78)$$

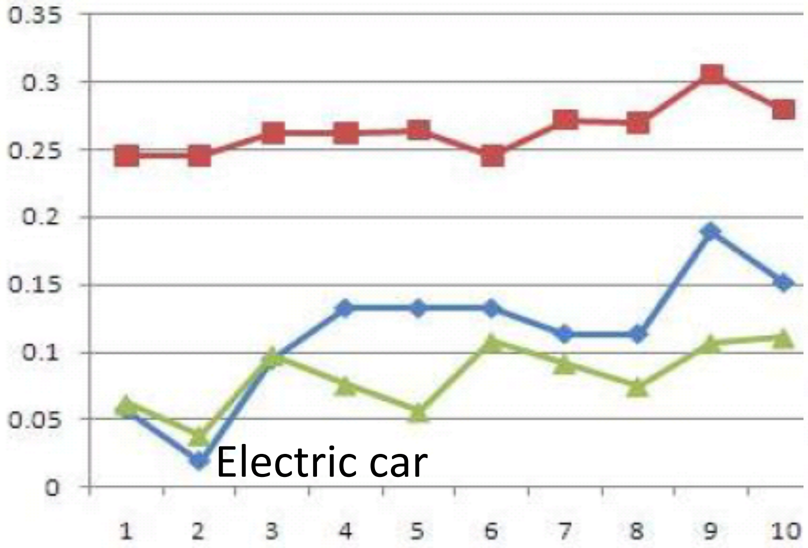


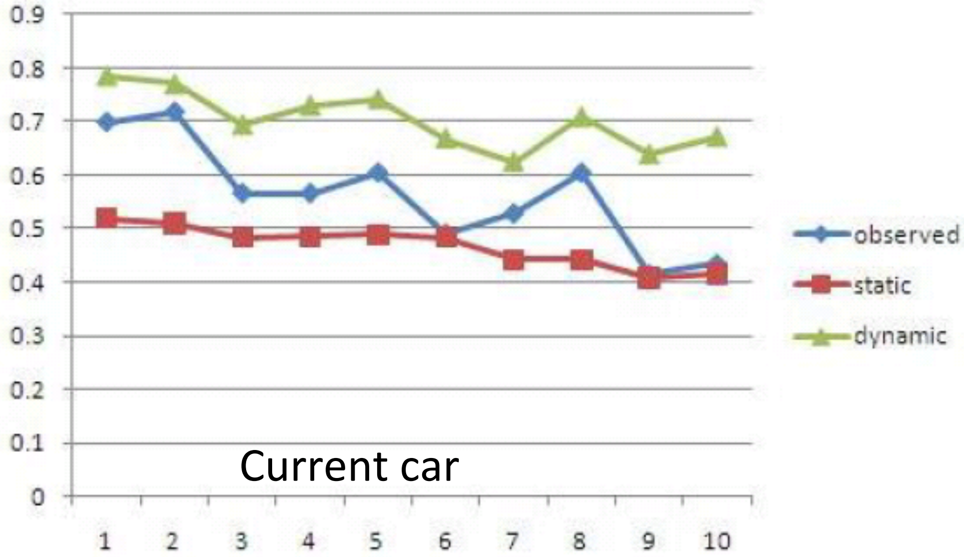
Alternative	gas	hybrid	electric	current	Dynamic	
					Estim	t-Stat
ASC2		X			-1.09	4.05
ASC3			X		1.18	1.94
ASC4				X	-1.10	6.96
mpg_known	X	X			0.078	6.20
mpg_unknown	X	X			0.042	3.66
veh_age				X	-0.133	4.26
price_st	X	X			-0.062	0.46
price_dy			X		-1.01	5.37
range			X		0.723	4.32
N observed					636	
LL(0)					-1683.09	
LL(final)					981.43	
likelihood ratio index					0.42	

Market shares - comparison









Conclusions

- New gasoline vehicles, hybrid and electric vehicles occupy smaller market shares (around 10% each) at the end of the five year period;
- All new typologies become more popular after the fifth time period;
- Static models are incapable of recovering peaks in the demand function;
- MNL model underestimates the market share of the "not buy", and dramatically overestimate the share occupied by electric vehicles in the next five years;
- Dynamic model overestimates the market share of the "not buy", but is capable to reproduce the descending trend for this alternative.

Thanks My students at UMD… (Pratt, Michael, JM, Nayel, Renting, Me, Yangwen) and to my colleague Fabian Bastin (University of Montreal)