ABSTRACT

Title of Dissertation: MARKET PENETRATION OF NEW VEHICLE TECHNOLOGY: A GENERALIZED DYNAMIC APPROACH FOR MODELING DISCRETE-CONTINUOUS DECISIONS

Yan Liu, Doctor of Philosophy, 2017

Dissertation directed by: Associate Professor, Cinzia Cirillo
Department of Civil and Environmental Engineering

Energy consumption and greenhouse gas (GHG) emissions are at their highest levels in history. One of the largest sources of GHG emissions in the United States is from burning fossil fuels for transportation. In developing countries GHG emissions from private vehicles are growing rapidly with their wealth. Government agencies attempt to reduce dependency on fossil fuels by regulating the ownership/usage of private vehicles, promoting vehicles with higher engine efficiency, introducing new fuel types, and defining stricter emission standards. Hybrid and electric vehicles are gaining consumers' interest and trust, and their sale shares are gradually increasing. Meanwhile, environmental awareness, taxes on conventional gasoline cars, and incentives for cars with new technologies, make small and alternative-fuel vehicles more attractive. The future of personal transportation is uncertain; in particular, car ownership, vehicle type preferences and usage behavior are likely to change in surprising ways. In this context, it is important to assess the influence of the vehicle market evolution on consumer's vehicle demands and travel behaviors.

This dissertation proposes a comprehensive modeling framework that is able to analyze different dimensions of the car purchasing and usage problem. A multi-facet approach is taken for the investigation, and different model types are proposed. The investigation starts with a **mixed logit model** that accounts for time-series choices, heterogeneity in preferences and correlation across alternatives. This model is estimated on Stated Preference Survey data collected in Maryland and quantifies market elasticities and willingness-to-pays for improving car characteristics. Afterward, a **dynamic discrete choice model** is developed to predict the diffusion of hybrid and electric cars in Maryland, with consideration of household's forward-looking behavior and stochasticity in vehicle market evolution. This model focuses on vehicle purchase time and vehicle type choice. To further consider vehicle usage decision, an **integrated discrete-continuous choice model** is proposed to jointly estimate household's discrete choices on vehicle type/ownership and continuous choice on vehicle usage. The model is applied to estimate household-level vehicle emissions in Maryland, USA and Beijing, China.

The dissertation concludes with a **sequential discrete-continuous choice model**. The modeling framework is applied to estimate vehicle ownership and usage decisions of forward-looking agents over time in a finite time horizon. In particular, a recursive probit model is formulated to estimate a sequence of vehicle holding decisions, while a regression is used to estimate a sequence of vehicle usage decisions. The proposed model is tested and validated on simulated discrete and continuous choices in a car ownership problem setting.

The dissertation contributes to the theory of dynamic models for discrete-continuous decisions. The sequential discrete-continuous choice model is the first to measure the dynamic interdependency between discrete choice and continuous choice over time. The dissertation also contributes to the understanding of critical transportation issues, including market penetration of new vehicle technology, estimation of household-level vehicle emissions, and policy evaluation for promoting green vehicles and reducing dependency on cars and emissions.

MARKET PENETRATION OF NEW VEHICLE TECHNOLOGY: A GENERALIZED DYNAMIC APPROACH FOR MODELING DISCRETE-CONTINUOUS DECISIONS

by

Yan Liu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Associate Professor Cinzia Cirillo, Chair
Professor Ali Haghani
Associate Professor Ilya Ryzhov
Associate Professor Zhenhong Lin
Dr. Sevgi Erdoğan

Acknowledgements

This work would not have been possible without the financial support from the National Science Foundation and National Transportation Center at University of Maryland. I am especially indebted to my advisor Dr. Cinzia Cirillo, who have been working actively to provide me with protected academic time for my interested research and who have been very supportive of my career goals. No matter my research and life go smoothly or face challenges, she always provides me valuable guidance, believes in me, and encourages me that I can go further. I would not complete this dissertation and seven publications without the help and support from Dr. Cinzia Cirillo, who is also my good friend.

I am grateful to members of my Dissertation Committee, who have provided me extensive personal and professional guidance. As my teacher, mentor, and friend, they have not only provided me advice on scientific research but also shown me what a good scientist should be.

I would also like to thank other contributors whom I have had great pleasure to collaborate and work with during my Ph.D. study. In particular, Dr. Fabian Bastin has given me valuable suggestions on model formulation. Dr. Jean-Michel Tremblay has helped me to solve related R programming issues. Dr. Michael Maness and Dr. Yizhen Gu have provided supportive datasets for the studies in this dissertation.

The understanding and support from my family members are the most important to me in the pursuit of my Ph.D. I would like to thank my parents whose love and encouragement are with me in whatever dream I pursue. They are supportive

and believe that I can do great things. More importantly, I wish to thank my beloved Kelvin, who inspires me to follow my heart and pursue my desirable career.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AR(n)	Auto-Regressive Process of Order n
ASC	Alternative Specific Constant
BEV	Battery Electric Vehicle
BHTS	Beijing Household Travel Survey
BMOPT	Bayesian Multivariate Ordered Probit and Tobit
CES	Constant Elasticity of Substitution
CNY	Chinese Yuan
DC	District of Columbia
DDCCM	Dynamic Discrete-Continuous Choice Model
DDCM	Dynamic Discrete Choice Model
DOE	Department of Energy
DP	Dynamic Programming
EIA	Energy Information Administration
EMBEV	Emission Factor Model for the Beijing Vehicle fleet
EPA	Environmental Protection Agency
EV	Electric Vehicle
GHG	Greenhouse Gas
GEV	Generalized Extreme Value
GIS	Geographic Information System
GV	Green Vehicle
HEV	Hybrid Electric Vehicle
i.i.d.	Independently and Identically Distributed
IIA	Independence of Irrelevant Alternatives
KBB	Kelley Blue Book
MDCEV	Multiple Discrete-Continuous Extreme Value
MMNL	Mixed Multinomial Logit
MNL	Multinomial Logit
MOT	Ministry of Transportation
MOVES	Motor Vehicle Emission Simulator
MPG	Miles per Gallon
MPGE	Miles per Gallon Equivalent
MSI	Metro Service Index
MVN	Multivariant Normal
MVSPS	Maryland Vehicle Stated Preference Survey
NFP	Nested Fixed Point
NHTS	National Household Travel Survey
PEV	Plug-in Electric Vehicle
RMSE	Root-Mean Square Error
RP	Revealed Preference
SC	Stated Choice
SP	Stated Preference
SRS	Stratified Random Sampling

SUR	Seemingly Unrelated Regression
T1EV	Type 1 Extreme Value
TAZ	Traffic Analysis Zones
TCQSM	Transit Capacity and Quality of Service Manual
US	The United States
USD	US Dollar
USDOE	The United States Department of Energy
USEIA	United States Energy Information Administration
VAR(n)	Vector Auto-Regressive Process of Order n
VIN	Vehicle Identification Number
VMT	Vehicle Miles Traveled
VMD	Vehicle Miles Driven

Chapter 1: Introduction

1.1 Background and Motivation

Increasing levels of motorization, congestion, and pollution are inescapable conditions in large and growing metropolitan areas across the world. Modern societies are highly dependent on private vehicles to satisfy demand for activities; while fastest developing societies are moving from industrial-manufacturing economies to more service-oriented economies with greater automobile saturation.

In Europe and the United States (US), transportation accounts for more than a quarter of greenhouse gas (GHG) emissions, and light duty vehicles (LDVs) are the largest contributor (EPA, 2013). China is expected to experience the largest absolute growth in liquid fuels consumption, growing by about 46% in 2020 and doubling in 2040 compared to the 2010 level. India will have the fastest growth rate in liquid fuels consumption from 2010 to 2020 (3.0% per year) and experience the second-largest absolute growth (behind China), primarily driven by diesel fuels used in transportation (US Energy Information Administration, 2014).

The emissions from transportation contribute to global climate change and smog, which are harmful especially to the health of kids and the elderly. To reduce fuel consumption and emissions from vehicles, the development of advanced vehicle technology has become a high priority for many governments and vehicle manufacturers around the world. Nine countries and regions (including the US, Mexico, South Korea, Europe, India, Japan, Brazil, China, Canada), which together account for 75% of global fuel consumption by LDVs, have adopted mandatory or voluntary standards for increasing fuel economy and reducing GHG emissions (US Energy Information Administration, 2015). As a consequence, highly efficient combustion engines, innovative power systems, and greener fuels are gradually available in the marketplace, setting the foundation for clean, efficient, sustainable, and cost-competitive vehicles (Department of Energy, 2015).

In recent years, more fuel-efficient vehicles or alternative energy sources are available in the market and their characteristics are expected to change over time as technology develops. Considering these dynamics and diversity of vehicle types in today and future's market, it is important for governments and auto manufacturers to: (a) understand consumers' vehicle type preference and vehicle demand over time, and (b) optimally manage and regulate vehicle fleet and usage to reduce adverse impacts of transportation. In this context, the role of researchers is to expand the knowledge of the problem and develop better analytical tools for the support of decision making in the dynamic market.

The importance of modeling household vehicle ownership, type, and usage has been recognized for several decades, although new vehicle technologies and dynamic vehicle market have been taken into account more recently (Fontaras et al., 2008; Flamm and Agrawal, 2012; Glerum et al., 2013; Cirillo et al. 2015). These models play a significant role in: (a) determining consumer preference on vehicle types from the perspective of car manufacturers (Bunch et al., 1993; Axsen and Kurani, 2013; Glerum et al., 2014; Cirillo et al., 2017), (b) predicting individuals' activity and travel behavior from the perspective of traffic planners (Ben-Akiva and Bowman, 1998; Bhat and Singh, 2000; Paleti et al., 2013; Oakil et al., 2014), and (c) evaluating policies and regulations to reduce vehicle emissions from the perspective of governments and policy makers (Hayashi et al., 2001; Vyas et al., 2012; Feng et al., 2013; Liu and Cirillo, 2015; Liu and Cirillo, 2016). Intuitively, it is important to accurately predict household vehicle holdings and miles traveled by vehicle type to support critical transportation infrastructure planning and project auto emission levels (Bhat and Sen, 2006).

1.2 Current Research Status

Discrete choice models (i.e., multinomial logit, mixed logit, structural equation models) have been widely used to investigate household vehicle ownership and type choices (Golob et al., 1997; Brownstone et al., 2000, Mabit and Fosgerau, 2011; Jensen et al., 2013; Rasouli and Timmermans, 2013). Unfortunately, most of these models are static and the analysis are usually based on cross-sectional data. Besides, many of these studies ignore vehicle usage behavior which is essential to calculate emissions from private transportation. It should also be noted that given the low market shares of advanced technology vehicles and the rapid changes on the supply side, it is not surprising that many studies on vehicles with new technologies are based on stated preference (SP) data (Hensher, 1994).

To overcome limitations of static models, a number of dynamic models have been developed and applied to the car ownership problem (Ben-Akiva and Abou-Zeid, 2007; Nolan, 2010; Schiraldi, 2011; Cirillo et al., 2015). These studies have addressed a number of interesting modeling issues, such as initial conditions, state dependency, forward-looking behavior, taste heterogeneity, substitution pattern among alternatives, and data collection. For instances, Ben-Akiva and Abou-Zeid

(2007) proposed a dynamic discrete choice model integrated with Hidden Markov Chain to model sequences of decisions; their model accounts for consumer's previous actions (i.e., inertia effect) and for the evolution of latent variables. Dynamic models based on the seminal work of John Rust (1987), use dynamic programming formulation and consider consumer's expectation and market evolution over time. Schiraldi (2011) was the first to introduce a dynamic structural approach to study car replacement decisions for a second-hand vehicle market in Italy. At the same time, Cirillo et al. (2015) proposed a dynamic discrete choice model with regenerative optimal stopping formulation to capture not only the optimal car purchase time but also consumer's vehicle type choice in an evolving market. However, these models only capture discrete choice on car ownership and ignore car usage decision.

In more recent literature, integrated car ownership models have been developed to jointly estimate household car holding, type and usage decisions. Under this family, we classify: the multiple discrete-continuous extreme value (MDCEV) model developed by Bhat and his co-authors (2006), the Bayesian multivariate ordered probit and tobit (BMOPT) model by Fang (2008), and the integrated discrete-continuous choice model by Liu et al. (2014). These models generally have a better performance in estimation and prediction because they consider the correlation between households' discrete choices of vehicle holding and vehicle type and continuous choice of vehicle miles traveled (VMT). However, these models are static; they are not able to capture the changes in households' time-dependent vehicle holding and use decisions.

There are two recent studies that aim at estimating simultaneously household vehicle ownership and usage decisions over time. Gillingham et al. (2015) developed a dynamic structural micro-econometric model to estimate household vehicle ownership, type choice, and usage in Denmark. In particular, a “nested logit” structure is proposed for the discrete choices: the “upper level” models car purchase and type decisions, while the “bottom level” captures trading behaviors of the current car. For the continuous choice, the utility of driving is modeled as a 2nd-order polynomial function of annual kilometers traveled. Their modeling structure allows for forward-looking behaviors, accounts for endogenous scrappage decisions, and captures the evolution of the society and of the market. However, the adopted “two-stage” estimation approach breaks the strict cross-equation restriction that the consumer should care equally about money spent on buying/ selling a car and money spent on driving a car. This estimation approach could lead to insufficient estimated coefficients. Other limitations are that the model only accounts for households with at most one car, and it cannot measure the correlation between car holding and driving decisions.

To overcome the one-car limitation, Glerum et al. (2013) developed a dynamic car ownership model that estimates the joint decision on vehicle transitions, mileage and fuel type in an infinite time horizon. The model is formulated as a discrete-continuous choice model that is embedded in a dynamic programming framework to account for household’s forward-looking behavior in the context of car acquisition. For two-car households, constant elasticity of substitution (CES) utility is adopted to determine the most appropriate allocation of mileage driven by each of the two cars held by households, with a total mileage budget. Despite capturing the dynamic nature of vehicle transaction and use, this model has some limitations: the continuous choice of vehicle mileage is myopic and deterministic; each household can have at most two cars; and the total mileage budget makes the model impossible to evaluate policies related to vehicle use.

1.3 Research Objectives and Contributions

The development and the deployment of advanced vehicle technology has become a high priority for many governments and vehicle manufacturers around the world. These technologies include alternative fuels, plug-in electric vehicles, batteries, electric drive technologies, and efficient combustion engines. They gradually diversify today's vehicle market and influence people's preference on vehicle ownership, type, and usage. Modeling vehicle ownership and usage in the era of advanced vehicle technology becomes important: (a) for transportation analysts aiming at understanding the future of travel behavior and (b) for policy makers called to regulate the energy market and to moderate emissions from the transportation sector. In this dissertation, a multi-facet approach is taken to develop a mature methodology aiming at forecasting the changes in household vehicle ownership, vehicle type choice, and usage behavior over time. The investigation proposes static models for time-series discrete choices and for joint discrete-continuous choice, and generalized dynamic models for discrete choices over time and for joint discrete-continuous choices over time.

The mixed logit formulation proposed in this thesis considers time-series choices from the same individual (panel effect). Meanwhile, random parameters account for taste variation, and flexible correlation patterns among alternatives. The model is estimated on Stated Preference data collected from Maryland residents over a hypothetical nine-year future time period. Car characteristics and fuel prices proposed to respondents in each of the SP scenarios change over time to mimic dynamics in the vehicle market. The analysis attests for the first time that respondents are able to consider trade-offs across vehicles with different technologies and alternative fuels over an extended and future time horizon. The model calculates elasticities with respect to vehicle price, gasoline and electricity vehicle. The results also provide important implications for the understanding of vehicle preferences and for the definition of willingness-to-pay (WTP) for different vehicle characteristics.

The dynamic discrete choice model, based on an optimal stopping formulation derived from dynamic programming, uses a non-linear function that not only captures instantaneous utility in the current market, but also considers expected future utility from future market conditions. In other words, households are forward-looking and the market evolution is modeled through autoregressive vectors of dynamic variables. The entire model framework has been applied to predict the market penetration of "green" vehicles in the State of Maryland from the year 2014 to 2022. The data used for the empirical analysis was again collected from the Maryland Vehicle Stated Preference Survey, with a supplementary historical data of fuel prices from US Energy Information Administration (EIA). The model results have been applied to test different policy scenarios; the variables of interest include fuel price, vehicle purchase price, and characteristics of electric cars. The dynamic discrete model is particularly appropriate to recover peaks/valleys and rapid changes in consumer demand over time, which provide important evidence for vehicle producers. The estimation requires time-series data.

The integrated discrete-continuous choice model jointly estimates household decisions on vehicle holding, type, and usage. The model combines with a motor emission simulator (MOVES) to estimate household-level vehicle emissions. The entire model has been applied to estimate household vehicle ownership, type choice, usage behavior, and emissions in Maryland (US) and then transferred to Beijing (China). The data for the Maryland application are extracted from the 2009 National Household Travel Survey, Maryland Vehicle Stated Preference Survey, and Consumer Reviews. The Beijing application proves the transferability of the integrated model to a large urban area in a developing country. This application uses the 2010 Beijing Household Travel Survey data and GIS shape files of residential location and public transit information. The tools proposed can be used by governments and decision makers to evaluate different policies and regulations for the promotion of "green" vehicles and for reducing vehicle-related emissions.

The sequential discrete-continuous choice model extends the theory of the integrated discrete-continuous choice model on a temporal basis and improves existing dynamic discrete-continuous models based on a pure dynamic programming perspective. The model jointly estimates household vehicle ownership and usage over time. In particular, a recursive probit model is formulated to estimate a sequence of vehicle holding decisions, while a regression is used to estimate a sequence of vehicle usage decisions over time. The inherent Gaussian distributed error component of the recursive probit model enables its integration with regressions. Correlation between the discrete and continuous parts, varying over time, is captured with a full unrestricted variance-covariance matrix of the unobserved error components. The estimation process benefits from a finite-horizon scenario tree technique that is efficient and reduces the dimension of the integrals associated to the probit choice probability calculation. The sequential discrete-continuous choice model has been validated on simulated data sets of car ownership and usage choices and is able to reproduce the evolving trends of households' discrete and continuous demands.

A simulation experiment is performed to check the accuracy of model estimation based on different sample size and forward-looking time periods. Results show that the accuracy of model estimation is mainly based on the number of households and the time difference between total study time periods and forward-looking time periods. The sequential discrete-continuous choice model is appropriate to solve problems with a sequence of discrete and continuous decision variables. Given the dynamic nature, the model requires panel data for estimation. The model can help governments and decision makers to evaluate time-dependent policies and pricing schemes that promote new vehicle technologies and reduce dependency on cars and emissions.

1.4 Dissertation Organization

This dissertation is organized in eight Chapters.

Chapter 2 reviews the literature on vehicle ownership models in an era were major innovations are expected from the automotive industry. The review outlines the progress in the development of vehicle ownership models, from static to dynamic framework, and from single choice models to integrated discrete-continuous choice models.

Chapter 3 describes the datasets that have been collected and used for model calibration and application. They include the Maryland Vehicle Stated Preference Survey data, the 2009 US National Household Travel Survey data, fuel price data from US Energy Information Administration, vehicle characteristics from Consumer Reviews, Beijing Household Travel Survey data, and GIS shapefiles of Beijing residential location and public transit information. The first four data sources provide information on household vehicle ownership and usage for Maryland State and the Washington DC Metropolitan Area in the US, while the last two data sources deliver this information for Beijing in China.

Chapter 4 proposes a modeling framework based on mixed multinomial logit with panel effect for vehicle type choice analysis. This model has been estimated on the Maryland Vehicle Stated Preference Survey data, which was designed to analyze household future preferences for gasoline, hybrid electric, and battery electric vehicles in Maryland. Vehicle market elasticities and willingness-to-pay with respect to a number of vehicle characteristics are calculated and discussed.

Chapter 5 introduces a generalized dynamic discrete choice model to capture the optimal time of vehicle purchase and household's vehicle type choice over a finite horizon. Different model forms are proposed to consider the purchase behavior in different durable good markets: the regenerative optimal stopping formulation allows agents to return to market after a purchase is made, while the regular optimal stopping formulation guarantee agents to be out-of-market after a change in status. Moreover, a vector autoregressive process is built-in to capture market evolution. The proposed model has been applied to forecast “green” vehicle adoption rate for households living in Maryland. Different policy scenarios are evaluated, including changes in fuel price, vehicle purchase price, and improvement of vehicle characteristics.

Chapter 6 introduces an integrated discrete-continuous choice model to jointly estimate households’ decisions on vehicle holding, type, and usage. The proposed model combined with motor emission simulators calculates household-level vehicle emissions. Two applications of this modeling framework have been: the first application aims at exploring the influence of the appearance of “green” vehicles on household car ownership and usage behavior; the second aims at investigating the transferability of this model to developing societies. Household-level vehicle emissions are estimated for both applications and different policy scenarios are evaluated.

Chapter 7 develops a sequential discrete-continuous choice model to jointly estimate household vehicle ownership and usage over time, with the consideration of forward-looking agents in a finite time horizon. Two model forms are proposed based on the number of alternatives in the discrete choice set: binary case and multivariate case. The models have been validated and applied on simulated datasets of car ownership and use choices over time.

Chapter 8 summarizes the main findings, outlines research contributions, and provides future research directions on the theory and application of the car ownership problem.

Chapter 2: Literature Review on Vehicle Ownership Models

The literature on advanced vehicle technology is vast and numerous. References can be found not only in transportation journals but also in applied econometrics, environmental economics, energy and sustainability related journals. This Chapter mainly refers to articles that model people's preferences on vehicle ownership, type, and usage in the era full of GVs with new technologies and alternative fuels.

2.1 Modeling Consumer Preference on New Vehicle Technologies

This section focuses on studies that elicit individual preferences from survey data and estimate market penetration of new vehicles including electric cars and those that run on alternative fuels. Given that their actual market shares are low and that rapid changes are expected on the supply side, it is not surprising that many studies on advanced technology vehicles are based on SP data (Hensher, 1994). In 1991, a three-phase SP survey was implemented in the South Coast Air Basin of California to predict the effect on personal vehicle purchases of attributes that potentially differentiate clean-fuel vehicles from conventional gasoline (or diesel) vehicles. Attributes considered included: limited availability of refueling stations, limited range between refueling or recharging, vehicle prices, fuel operating costs, emissions levels, multiple-fuel capability and performance (Golob et al., 1993). This pioneering data set has been used by several authors to estimate demand for alternative fuel vehicles. These studies often use discrete-choice or structural equations models (Bunch et al., 1993; Golob et al., 1997; Brownstone et al., 2000).

Volatility in gas price, increasing concerns about emissions and global warming, as well as progress in alternative fuel vehicle technology have caused a re-emergence of interest in alternative fuel vehicle data and in behavioral models for demand forecasting and scenario analysis. A SC survey was conducted in Denmark in 2007 by Mabit and Fosgerau (Mabit and Fosgerau, 2011); the sample consisted of new car buyers only. The survey considered five vehicle types: conventional, hydrogen, hybrid non-plugin, bio-diesel and electric vehicles. A mixed logit model was estimated to improve on previous contributions by controlling for reference dependence and allowing for correlation of random effects, which were found to be very important. The monetary attributes considered were purchase price and annual cost, where the annual cost is the sum of maintenance cost, fuel expenses based on intended driving, and annual taxes. The non-monetary attributes were operation range, refueling frequency, acceleration time, and a service dummy. The pollution level of alternative fuel vehicles was specified relative to the reference vehicle.

Jensen et al. (Jensen et al., 2013) collected stated choices and used them to measure the extent to which the experience of using an EV may affect individual preferences and attitudes. The authors set up a "long panel" survey, where data was gathered before and after individuals experienced an EV in real life during a three-month period. They also measured attitudinal effects that might affect the choice of an EV by individuals; their results show that preferences and attitudes are indeed affected by real life experience.

Rasouli and Timmermans (Rasouli and Timmermans, 2013) designed a SC experiment to better understand the decision process of buying an electric car and to derive the relative importance of factors that affects the choice of a special focus on social influence. Attributes considered include vehicle attributes, contextual attributes and social influence attributes. In particular, the social influence attributes describe possible reviews and adoption of this new technology by various elements of social networks (family, friends, colleagues and the larger social network of peers) and the impact of the nature of reviews (positive or negative).

Axsen and Kurani (Axsen and Kurani, 2013) collected data from 508 households representing new vehicle buyers in San Diego County, California in 2011. The mixed-mode survey collected information about access to residential recharge infrastructure, three days of driving patterns, and desired vehicle designs and motivations via design games. The data was used to compare consumers' stated interest in conventional gasoline, hybrid, plug-in hybrid, and pure EV of varying designs and prices.

A survey of consumer attitudes on EV was conducted in Manitoba from late 2011 to early 2012. It utilizes two price assessment methods; it includes direct questions regarding willingness to pay a price premium for an EV, and an indirect question based on the van Westendorp price sensitivity method. The survey derives price ranges for EVs but also supports the hypothesis that EV rollout has focused too much on technology, and not enough on consumers (Larson et al., 2015).

Although SP data are critical for obtaining information about attributes not available in the marketplace, arguments always remain that SP techniques give implausible forecasts. Several studies (Brownstone et al., 2000) highlight the advantages of merging SP and revealed preference (RP) data to predict individual's preferences on alternative-fuel vehicles. Despite that RP data is plagued by multicollinearity and difficulties with measuring vehicle attributes, they appear to be critical for obtaining realistic body-type choice and scaling information.

To summarize, discrete choice models and latent variable models are commonly used to predict the choice of clean-fuel vehicles. Specifically, mixed logit models provide improved fits over logit models by considering for heterogeneity among respondents, controlling for reference dependence, and allowing for correlation of random effects, which are proved to be very important. Latent variable models, such as structure equation models and hybrid choice models, are appropriate to account for attitudinal or social influence attributes of new vehicle technology adoption. Some interesting studies are summarized in Table 2.1.

Based on previous studies on emerging vehicle technologies, seven types of independent variables are important to be considered: (1) the objective characteristics of vehicles (i.e. performance, purchasing and operating costs, and driving range); (2) performance attributes, such as acceleration (Potoglou and Kanaroglou 2007; Mabit and Fosgerau 2011) and top speed (i.e. Dagsvik et al. 2002; Batley et al. 2004); (3) charging speed (Ewing and Sarigollu 1998; Brownstone et al. 2000; Hidrue et al. 2011); (4) fuel availability, mostly as a percentage of conventional fuel stations where it is possible to charge the batteries (Bunch et al. 1993; Batley et al. 2004; Horne et al. 2005; Potoglou and Kanaroglou 2007; Bolduc et al. 2007; Achtnicht 2012; Hackbarth and Madlener 2013); (5) parking availability, mostly like free parking incentive and other political benefits (Adler et al. 2003; Potoglou and Kanaroglou 2007); (6) attitudinal, perceptional and social influence attributes (Mokhtarian et al. 2001;

Vredin Johansson et al. 2006; Bolduc et al. 2007; Atasoy et al. 2010; Valeri and Cherchi, 2016); and (7) sensitivity to emissions (Daziano and Chiew, 2012; Daziano et al., 2017).

Table 2. 1 Summary of Studies on Emerging Vehicle Technologies

Reference	**Data Type**	**Data Source**	**Sample Size**	**Vehicle Type**	**Model Type**	**Independent variables**
Golob et al., 1993	SP	California South Coast Air Basin, 1991	Around 700 respondents	clean-fuel vehicle, conventional gasoline vehicle	Multinomial Logit	availability of refueling stations, refueling range vehicle prices, fuel operating costs, emissions levels, multiple-fuel capability
Bunch et al., 1993	SP	California South Coast Air Basin, 1991	Around 700 respondents	electric vehicle, unspecified liquid and gaseous fuel vehicle (dedicated or multiple-fuel), gasoline vehicle	Nested Multinomial Logit	vehicle purchase price, fuel operating cost, refueling range, availability of fuel, dedicated versus multiple-fuel capability, level of reduction in emissions
Golob et al., 1997	Joint RP & SP	California, 1993	4747 households	electric vehicle, conventional gasoline vehicle	Structural equation	Age of principal driver, gender, employment status, vehicles age, vehicle class, operating cost, fuel type, refueling range, household membership, household income, household head information, number of workers, number of vehicles in household.
Brownstone et al., 2000	Joint RP & SP	California, 1993	4747 households	gasoline, electric, methanol, and compressed natural gas vehicles	Multinomial Logit and Mixed Logit	purchase price, range, acceleration, top speed, pollution, luggage space, operating cost, station availability, vehicle type dummy
Mabit and	SP	Denmark,	2146	conventional,	Mixed Logit	purchase price, annual

Fosgerau, 2011		2007	individuals, 25746 observations	hydrogen, hybrid, bio-diesel, and electric vehicles		cost, operation range, refueling frequency, acceleration time, service dummy, pollution level
Jensen et al., 2013	Joint two-wave SP	Denmark, 2011	369 individuals, 5904 observations	electric vehicle, conventional gasoline vehicle	Hybrid choice model	purchase price, fuel cost, driving range, carbon emissions, top speed, battery stations, battery life, recharging location, vehicle size, respondent's age
Rasouli and Timmermans, 2013	SP	The Netherland, 2012	726 respondents	electric vehicle	Mixed logit models	Vehicle characteristic attributes, contextual and social network attributes
Axsen and Kurani, 2013	SP	San Diego County, California, 2011	508 households	conventional gasoline, hybrid, plug-in hybrid, and electric vehicles	NA	NA
Larson et al., 2015	RP	Manitoba, 2011-2012	240 people	electric vehicle	The van Westendorp price sensitivity method	Battery range, ability to charge at home, total cost, ability to charge at work, ability to charge quickly, government subside and tax exemption
Valeri and Cherchi, 2016	SP	Italy, 2013	121	gasoline, diesel, compressed nature gas, liquefied petroleum gas, hybrid electric, battery electric vehicles	Hybrid choice model (structure equation)	Socio-economics, vehicle characteristic variables, latent habitual behavior

2.2 Dynamic Discrete Choice Models in Economics and Transportation

In previous studies, static models based on disaggregate data are usually used to model and forecast car ownership in transportation planning; they are limited to predicting individual’s behavior and preference in the present regardless of past and future states. However, more efficient and less pollutant vehicles are gradually being available in the marketplace and new opportunities will be created for alternative energy sources over the next five to ten years. Dynamic estimation techniques for analyzing the impact of technological improvements and rapid changes in energy costs are necessary to understand the mobility of tomorrow and future preferences over vehicle characteristics. This section focuses on DDCMs to capture individuals’ time-dependent decisions, accounting for intertemporal substitution effects, look-forward behaviors, and market evolution.

2.2.1 Dynamic Discrete Choice Models in Economics

DDCMs are widely used in economics and related fields. They are useful tools for the evaluation of price elasticity, intertemporal substitution, and new policy in durable goods market. In the structure of DDCMs, agents are forward-looking and maximize expected intertemporal payoffs, with the knowledge of the evolving nature of product attributes such as price and technology. The earliest generation of research on DDCMs includes Wolpin (1984) on fertility and child mortality, Miller (1984) on job matching and occupational choice, Pakes (1986) on patent renewal, and Rust (1987) on machine replacement. Although computational complexity of model estimation becomes a well-known impediment to the development of these dynamic structure, a significant number of interesting applications appears in different areas of economics to solve the empirical issues, e.g., permanent unobserved heterogeneity, initial conditions, censored outcomes and sample selection, measurement error, endogeneity, identification, etc. (Aguirregabiria and Mira, 2010).

With his pioneering work in dynamic modeling, Rust (1987) was the first to formulate the optimal stopping problem and to estimate the optimal time to replace a bus engine. The model was conceived for a single agent, a homogeneous product, and infinite time horizon; random components were assumed to be additively separable, conditionally independent and extreme value distributed. Melnikov (2013) expanded Rust's model to consider a binary decision, whether to buy or to postpone the purchase, based on the expected evolution of printer's quality and price. In his dynamic structure, Melnikov considered heterogeneous products and homogeneous consumers. He assumed that consumers will be out-of-market once they make a purchase, and random components are independently distributed over consumers, products, and time periods. Lorincz (2005) extended the Rust model by proposing the so-called persistent effect, which allows consumers who already had a product to upgrade it instead of replacing it.

Knowing the importance of incorporating consumer heterogeneity, the dynamic structure further improved in a series of later papers (Berry et al., 1995; Shcherbakov, 2008; Carranza, 2010; Gowrisankaran and Rysman, 2012; Dube et al., 2012). Berry et al. (1995) showed that it is necessary to consider consumer heterogeneity to obtain realistic predictions of elasticity and welfare. Their model includes random coefficients, accounts for market-level demand shocks, and endogenous prices, but is static in nature. Dube et al. (2012) recast Berry's estimation as a mathematical program with equilibrium constraints to avoid numerical issues associated with the standard nested fixed point (NFP) algorithm and to make the estimation process more efficient. Gowrisankaran and Rysman (2012) analyzed consumer's preferences over digital camcorder products by combining Berry's modeling techniques of consumer heterogeneity and Rust's optimal stopping technique. Their model explicitly accounted for dynamics in consumer's behavior and allowed for unobserved product characteristics, repeated purchases, endogenous prices, and multiple differentiated products. Another interesting extension of Rust's bus engine replacement model was the integration of an auto-regressive process of order n (AR(n)) type serial correlation of error components into the dynamic structure (Reich, 2013). To make the estimation process more efficient, Reich (2013) decomposed the integral over the unobserved state variables in the likelihood function into a series of lower dimensional integrals, and successively approximated them using Gaussian quadrature rules. More recently, DDCMs have been developed and applied to many other areas such as demand for housing (Bayer et al., 2015), emergency evacuation (Serulle, 2015), and car ownership and purchase (Schiraldi, 2011; Cirillo et al., 2015).

2.2.2 Dynamic Car Ownership Models

Dynamic structures for car ownership include: dynamic transaction and duration models (Gilbert, 1992; de Jong, 1996; Bhat and Pulugurta, 1998; Mohammadian and Miller, 2003; Rashidi et al. 2011), models based on SP panel data (Brownstone et al., 2000; Hensher and Greene, 2001), models that account for past behavior and that use lagged variables (Ben-Akiva and Abou-Zeid, 2007; Nolan, 2010), and approaches based on DP with forward-looking agents (Schiraldi, 2011; Cirillo et al., 2015; Glerum et al. 2013; Gillingham et al., 2015).

Duration models mainly aim to capture dynamics in car ownership, and are used especially to forecast households' vehicle transaction behavior. Gilbert (1992) proposed a hazard model to estimate the distribution of automobile ownership length, and the effects of car characteristics, socioeconomics and market attributes on vehicle holding. de Jong (1996) calibrated a car ownership model system to estimate household's vehicle holding, choice of vehicle type, annual vehicle miles travelled (VMT), and fuel efficiency. He adopted a stochastic duration model based on a hazard function to predict the length of vehicle holding. This model was later combined with the Dutch Dynamic Vehicle Transaction Model to account for car disposal without replacement. Duration models for the time between vehicle transactions have also been used to explain the total number of cars in a household (Bhat and Pulugurta, 1998). Mohammadian and Miller (2003) proposed a market-based transaction approach to solve inconsistency in observed choices. They employed a mixed logit model to investigate the effects of heterogeneity in the dynamic transaction model and to distinguish between heterogeneity-based and state-dependence-based effects for the observed persistence in choice behavior. Rashidi et al. (2011) estimated a system of hazard-based equations in which timing of residential relocation, job relocation, and vehicle transaction were selected as endogenous variables.

The availability of high-quality panel data is always a challenging issue for the calibration and validation of dynamic models. RP panel and pseudo-panel data have been widely used in dynamic models for car ownership. However, both of them have limitations. For panel data, the size and representativeness of the samples decline over time due to attrition, so the data sets are often small (Hanly and Dargay, 2000). An important disadvantage of pseudo-panel data is that averaging over cohorts transforms discrete values of variables into cohort means, therefore individuals' information is lost (Dargay and Vythoulkas, 1999). Due to these limitations, many researchers started to use SP panel data and the combination of SP and RP data for dynamic model estimation. Brownstone et al. (2000) used RP and two waves of SP data to estimate demand for vehicles with alternative fuels. The joint model estimated on both RP and SP data was found to be superior to other specifications. The SP part provides essential information about attributes not available in the marketplace, while the RP part guarantees a plausible model for forecasting. Considering new car types or technologies not commonly used in the marketplace, Hensher and Greene (2001) modeled transactions with new vehicle types which required the collection of SP data. All existing studies based on SP data aim at forecasting market shares for new car types and individual preferences, but are incapable to predict when choices will happen over time (Cirillo et al., 2015).

In transportation, most DDCMs account for consumer's previous actions such as inertia effect. Future plans and random changes in the market conditions are usually not considered. Ben-Akiva and Abou-Zeid (2007) proposed a DDCM with the integration of Hidden Markov Chain to model sequence of choice decisions and the evolution of latent variables. The model, applied to driving behavior analysis, models behavioral dynamics such as individuals' plans, well-being states, and previous actions. Nolan (2010) estimated a dynamic random effects probit model on a micro-level longitudinal data to analyze the determinants of household car ownership in Ireland. This model considers impact from correlated effects, state dependence, unobserved heterogeneity, and initial conditions.

Consumers' expectation and market evolution over time are essential to model purchase decisions in current and future vehicle markets. Although sometimes the future effects are not fully known, or depend on factors that have not yet transpired, it can be assumed that individuals will maximize utility among the available alternatives at that time (Cirillo and Xu, 2011). This knowledge enables consumers to choose the alternative in the current period that maximizes his expected utility over the current and future periods (Train, 2009). Schiraldi (2011) was the first to introduce a dynamic structural approach with optimal stopping problem to study car replacement decisions in a second-hand vehicle market in Italy. His model accounts for consumer's heterogeneity, future expectation, price endogeneity, and infinite time horizon. However, the model is based on aggregate historical data not allowing attributes to change dynamically over time. To overcome the limitation, Cirillo et al. (2015) proposed a DDCM with regenerative optimal stopping formulation to capture not only the optimal car purchase time but also consumer's choices on vehicle types in a dynamically changing vehicle market. Alternatively, Fosgerau et al. (2013) developed the recursive logit model and was the first to apply it to optimal route choice problem by formulating each path as a sequence of link choices. At each node a decision maker chooses the utility-maximizing outgoing link with link utilities given by the instantaneous cost, the expected downstream utility identified by the Bellman equation. The recursive logit model corresponds to a DDCM and can be applied to dynamic car ownership analysis. Table 2.2 summarizes these dynamic structures for car ownership analysis.

Table 2. 2 Summary of Dynamic Discrete Choice Models for Car Ownership Analysis

Reference	Topic	Data	Model characteristics	Model	Optimization Method
Gilbert, 1992	Automobile holding duration	A panel survey that ran 6.5 years, from August 1978 to December 1984 in the US, including about 7500 households	Independency between disposal and replacement, time-varying duration, survivor and hazard function	Duration or hazard model	Maximum likelihood estimation
de Jong, 1996	Vehicle holding duration, type, and usage	First wave of a car panel and vehicle holding duration survey data in Netherland, 1992	Heterogeneity, time-varying covariates	A disaggregate model system: hazard model for vehicle holding duration, logit model for vehicle type, regressions for vehicle use and energy use	Maximum log-likelihood technique and ordinary least squares (OLS) estimation
Bhat and Pulugurta, 1998	Number of cars in a household	Household Activity Survey in Boston Region in 1991, Household Travel Survey in Bay Area in 1990, the Puget Sound Household Travel Panel Survey in 1991, the Dutch Mobility Panel Survey in 1987	ordered- and unordered-response mechanism	Ordered and unordered discrete-choice auto ownership models	Maximum log-likelihood estimation
Dargay and Vythoulkas, 1999	Household car ownership	Pseudo-panel data set from repeated cross-sectional data in UK Family Expenditure Survey since	Lags in adjustment of car ownership, short-run and long-run elasticity of car ownership	A model with random effect specification, and a random effect model with a first order auto-regressive scheme	Generalized least squared method

		1960s			
Hanly and Dargay, 2000	Household car ownership	British Household Panel Survey data in UK, 1993 - 1996,	Dependency of current car ownership on the past state, lagged dummies, uncorrelated error term over time or between households	An ordered probit specification model, and a latent regression	With STATA software
Mohammadian and Miller, 2003	Vehicle acquisition decision	Car Ownership Study in the Greater Toronto Area from 1990 to 1998 with more than 900 households	Dynamic social-demographic variables, dynamic elements with dependency on past behavior	A mixed (random parameters) logit model to estimate vehicle acquisition decision, accounting for heterogeneity and state dependency	Maximum log-likelihood estimation, with LIMDEP software
Ben-Akiva and Abou-Zeid, 2007	To model multiple layers of dynamic decisions	Not mentioned	State dependency, interactions between two-layer decisions	Hybrid choice model (HCM) integrated with a Markovian process to capture individual's plans and actions	Not mentioned
Nolan, 2010	Household car ownership	Longitudinal data from the Living in Ireland Survey (LIIS) from 1995 to 2001	Lagging variables, correlation effects, state dependence and initial conditions	A dynamic random effects probit model	maximum log-likelihood estimator, with STATA software
Rashidi et al., 2011	Vehicle Transaction time decision	The Puget Sound Transportation Panel Survey (PSTPS) dataset of 10 waves from 1989 to 2002, covering Seattle and surrounding areas	A baseline hazard and covariates, use of monotonic and non-monotonic baseline hazard function, time varying covariates, endogenous variables	The Weibull and log-logistic baseline hazard functions	Maximum log-likelihood estimation with non-linear procedure (NLP), and Trust Region Optimization algorithm
Schiraldi, 2011	Vehicle ownership and	Italian Motor Registry data	Heterogeneity across consumers,	A random coefficients discrete	Non-linear search over

	replacement decisions in the second-hand market	from 1994 to 2004	endogeneity of price, forward-looking agents	choice model that incorporates a dynamic optimal stopping problem	the parameters of the model, the contraction mapping
Cirillo et al., 2015	Optimal time of purchase, vehicle type choice	SP survey data of vehicle technology, fuel type and taxation policy experiments in Maryland in 2010	Finite time horizon, forward-looking agents, industry evolution, repeated purchases, differentiated products	Model the time of purchase as a regenerative optimal stopping problem, integrated with a multinomial logit model to estimate vehicle type choice	Maximum log-likelihood estimation
Fosgerau et al., 2013	Transfer path choice to link choice	GIS data of 1832 trips from 24 vehicles on a real network of 466 destinations and 37000 link choices in Borlänge	Infinite alternatives, link additive attributes, dynamic programming	Recursive logit model to determine a sequence of links instead of a traditional path choice problem	Maximum likelihood estimation

2.3 Review of Static Discrete-Continuous Car Ownership Models

Consumer demand choices are sometimes characterized by the choice of multiple alternatives simultaneously. For examples, the choice situation in activity-travel analysis consists of the discretionary type of activity to participate in and the continuous duration of time investment of the participation; the choice situation in car ownership analysis is composed of the discrete choices of vehicle type and quantity as well as the continuous choice of VMT. Discrete-continuous models have been investigated in marketing studies since 1980's. The earliest generation of discrete-continuous models that has investigated vehicle ownership choices was derived from conditional indirect utility function (Mannering and Winston, 1985; Train, 1986;

Hensher et al., 1992; de Jong, 1989b; de Jong, 1989a; de Jong, 1991) which is based on micro-economic theory. The basic concept of this method is to choose the combination of vehicle ownership and usage given the highest utility. Although based on single discreteness, this methodology based on indirect utility function is able to capture the interdependence between the discrete and continuous choices by observed variables. Consistent with economic theory, this elegant formulation is simple to implement. In the following subsections, we concentrate on several empirical studies on (both static and dynamic) integrated discrete-continuous choice models more recently.

2.3.1 Multiple Discrete-Continuous Extreme Value Model

The multiple Discrete-Continuous Extreme Value (MDCEV) Model, developed by Bhat (2005) and further applied by Bhat and Sen (2006) and Bhat et al. (2009), is a utility theory-based econometrics model that jointly estimates a discrete choice of multiple vehicle types and a continuous choice of VMT. The model is formulated based on an assumption that marginal utility diminishes as the level of consumption of any particular alternative increases, yielding multiple discreteness. The dependent variable in this model is the mileage for each vehicle type category. Utility for each household is maximized subject to a total mileage budget. With independently and identically extreme value distributed error terms, the probability function has a simple and elegant closed-form expression. Interestingly, the formulation of the MDCEV model collapses to the familiar multinomial logit (MNL) model in the case of single discreteness.

For each individual, the utility, considering K vehicle types potentially owned, is defined as the sum of the utilities obtained from traveling with each vehicle type:

$$U = \sum_{j=1}^{K} \psi(x_j)(m_j + \gamma_j)^{\alpha_j} \tag{2.1}$$

where m_j is the annual mileage traveled of vehicle type j ($j = 1, 2, \dots, K$); $\psi(x_j)$ is the baseline utility for traveling with vehicle type j; and x_j represents observed characteristics associated with vehicle type j; γ_j and α_j are parameters.

Equation 2.1 is valid if $\psi(x_j) > 0$ and $0 < \alpha_j \leq 1$ for all j. Further, the term γ_j determines if corner solutions are allowed (i.e. an individual does not hold one or more vehicle types) or if only interior solutions are allowed (i.e. an individual is constrained by formulation to hold all vehicle types). The purpose is to find the most appropriate allocation of mileage traveled by each vehicle type by maximizing the utility function with a total mileage budget. The utility form is flexible to accommodate a wide variety of mileage allocation situations based on the values of $\psi(x_j)$ and α_j.

To account for unobserved characteristics that impact the baseline utility, Bhat further introduces a multiplicative random element as follows:

$$\psi(x_j, \varepsilon_j) = \psi(x_j) \cdot e^{\varepsilon_j} \tag{2.2}$$

where ε_j captures the unobserved utility of holding vehicle type j. The exponential form for the introduction of random utility guarantees the positivity of the baseline utility as long as $\psi(x_j) > 0$. To ensure this latter condition, $\psi(x_j)$ is further parameterized to an exponential form as follows:

$$\psi(x_j, \varepsilon_j) = \exp(\beta' x_j + \varepsilon_j) \tag{2.3}$$

Therefore, the overall random utility function takes the following form:

$$\widetilde{U} = \sum_{j} \left[\exp\left(\beta' x_j + \varepsilon_j\right)\right] \cdot \left(m_j + \gamma_j\right)^{\alpha_j}$$

$$Subject\ to\ \ \sum_{j=1}^{K} m_j = M \tag{2.4}$$

M is the total mileage traveled. This type of utility maximization problem is traditionally solved by Lagrangian multiplier method in economics.

As the error term of the utility function is independently and identically Gumbel distributed, the final probability of household holding Q from K vehicle types and traveling certain mileage for each vehicle type is formulated as follows:

$$P\left(m_2^*, m_3^*, \dots, m_Q^*, 0, 0, \dots, 0\right) = \left[\prod_{i=1}^{Q} c_i\right] \left[\sum_{i=1}^{Q} \frac{1}{c_i}\right] \left[\frac{\prod_{i=1}^{Q} e^{V_i}}{\left(\sum_{j=1}^{K} e^{V_j}\right)^{Q}}\right] (Q-1)! \tag{2.5}$$

where $c_i = (\frac{1-\alpha_i}{m_i^* + \gamma_i})$ and $V_i = \beta' x_i + ln\alpha_i + (\alpha_i - 1)\ln(m_i^* + \gamma_i)$. In the case when $Q = 1$ (i.e. only one alternative is chosen), the formulation collapses to the standard MNL model and the continuous component drops out because the mileage traveled will be M.

Generally speaking, the proposed model is an extension of MNL model with the consideration of multiple discrete-continuous choices. It is able to handle a large number of vehicle types and capture interdependence between vehicle type and usage choices. Further, more comprehensive model specifications such as heteroscedasticity and correlation in unobserved characteristics are able to be integrated into the MDCEV model. However, this model has several limitations: (1) households are not allowed to choose multiple vehicles with the same type; (2) mileage of each vehicle is limited by the total mileage budget, making it impossible to evaluate the effect from policies; and (3) there is only one error term to capture the unobserved characteristics of vehicle type and usage choices, without measuring the interdependency between them.

2.3.2 Bayesian Multivariate Ordered Probit and Tobit model

Fang (2008) developed a Bayesian Multivariate Ordered Probit and Tobit (BMOPT) model, which is composed of a multivariate ordered probit for discrete choice and a multivariate Tobit for continuous choice, to jointly estimate vehicle type and usage demand in a reduced form. Specifically, for the discrete part, households' decisions on the number of vehicles in two vehicle type categories (cars and trucks) are jointly estimated by a multivariate ordered probit with an unrestricted correlation. For the continuous part, due to a large percentage of households does not have any truck and their truck usage is zero (truncated data structure), a multivariate Tobit model is adopted to estimate annual miles traveled for each vehicle type simultaneously. The integrated model combines the multivariate ordered equations and Tobit equations by error components with a full unrestricted variance-covariance matrix.

Let two latent continuous variables y_1^* and y_2^* represent the utility levels for holding cars and trucks, let latent variables y_3^* and y_4^* represent uncensored average annual miles driven by cars and trucks. The system for discrete-continuous choice of the vehicles can be written as:

$$y_{1i}^* = w_i'\beta_{11} + \ln(d_i)'\beta_{12} + \varepsilon_{1i} \tag{2.6}$$

$$y_{2i}^* = w_i'\beta_{21} + \ln(d_i)'\beta_{22} + \varepsilon_{2i} \tag{2.7}$$

$$y_{3i}^* = w_i'\beta_{31} + \ln(d_i)'\beta_{32} + \varepsilon_{3i} \tag{2.8}$$

$$y_{4i}^* = w_i'\beta_{41} + \ln(d_i)'\beta_{42} + \varepsilon_{4i} \tag{2.9}$$

where w_i is a vector of characteristics for household i; d_i is an indicator of residential density. The number of cars y_{1i} and trucks y_{2i} held by household i are determined by the value of the corresponding latent utility y_{1i}^* and y_{2i}^*. Specifically, $y_j = 0$ if $y_j^* \leq \alpha_1$; $y_j = 1$ if $\alpha_1 < y_j^* \leq \alpha_2$; $y_j = 2$ or more if $y_j^* > \alpha_2$, where the number of vehicles $j = 1, 2$. Average annual miles driven by cars y_3 is observed only when a household holds at least one car:

$$y_3 = y_3^*,\ if\ y_1 = 1\ or\ 2 \tag{2.10}$$

$$y_3 = 0,\ if\ y_1 = 0 \tag{2.11}$$

The same logic applies to miles driven by trucks y_4:

$$y_4 = y_4^*,\ if\ y_2 = 1\ or\ 2 \tag{2.12}$$

$$y_4 = 0,\ if\ y_2 = 0 \tag{2.13}$$

By fixing the two cut-points α_1 and α_2 in the ordered probit equations, the whole system can be written in to a SUR (seemingly unrelated regression) form. The error component follows a multivariate normal distribution with zero means and full unrestricted covariance matrix:

$$y_i^* = x_i\beta + \varepsilon_i, \quad \varepsilon_i \sim^{i.i.d.} N(\mathbf{0}, \Sigma) \tag{2.14}$$

Indexing households $i = 1, 2, \ldots, N$, the likelihood function is given as follows:

$$L(\beta, \Sigma | y_1, y_2, y_3, y_4) \propto \prod_{\substack{i=1 \\ y_{1i}=0, y_{2i}=0}}^{N} f(y_{1i}^* < \alpha_1, y_{2i}^* < \alpha_1 | \beta, \Sigma) \tag{2.15}$$

$$\times \prod_{\substack{i=1 \\ y_{1i}=0, y_{2i}=1}}^{N} f(y_{1i}^* < \alpha_1, \alpha_1 < y_{2i}^* < \alpha_2, y_{4i} = y_{4i}^* | \beta, \Sigma)$$

$$\times \prod_{\substack{i=1 \\ y_{1i}=0, y_{2i}=2}}^{N} f(y_{1i}^* < \alpha_1, y_{2i}^* > \alpha_2, y_{4i} = y_{4i}^* | \beta, \Sigma)$$

$$\times \prod_{\substack{i=1 \\ y_{1i}=1, y_{2i}=0}}^{N} f(\alpha_1 < y_{1i}^* < \alpha_2, y_{2i}^* < \alpha_1, y_{3i} = y_{3i}^* | \beta, \Sigma)$$

$$\times \prod_{\substack{i=1 \\ y_{1i}=1, y_{2i}=1}}^{N} f(\alpha_1 < y_{1i}^* < \alpha_2, \alpha_1 < y_{2i}^* < \alpha_2, y_{3i} = y_{3i}^*, y_{4i} = y_{4i}^* | \beta, \Sigma)$$

$$\times \prod_{\substack{i=1 \\ y_{1i}=1, y_{2i}=2}}^{N} f(\alpha_1 < y_{1i}^* < \alpha_2, y_{2i}^* > \alpha_2, y_{3i} = y_{3i}^*, y_{4i} = y_{4i}^* | \beta, \Sigma)$$

$$\times \prod_{\substack{i=1 \\ y_{1i}=2, y_{2i}=0}}^{N} f(y_{1i}^* > \alpha_2, y_{2i}^* < \alpha_1, y_{3i} = y_{3i}^* | \beta, \Sigma)$$

$$\times \prod_{\substack{i=1 \\ y_{1i}=2, y_{2i}=1}}^{N} f(y_{1i}^* > \alpha_2, \alpha_1 < y_{2i}^* < \alpha_2, y_{3i} = y_{3i}^*, y_{4i} = y_{4i}^* | \beta, \Sigma)$$

$$\times \prod_{\substack{i=1 \\ y_{1i}=2, y_{2i}=2}}^{N} f(y_{1i}^* > \alpha_2, y_{2i}^* > \alpha_2, y_{3i} = y_{3i}^*, y_{4i} = y_{4i}^* | \beta, \Sigma)$$

Overall, the BMOPT model is easy to implement, convenient to get inferences and hence draw policy implications, able to handle a large total number of vehicles. Within this framework, vehicles are categorized into fuel efficient (cars) and fuel inefficient (trucks) vehicles, which permits implementations of possible environmental and energy saving policies. This model can be extended to incorporate a finer classification of vehicles to suit the needs of particular studies. However, it will become computationally intensive with increasing vehicle categories because the number of equations to be estimated increases proportionally with the number of categories.

2.3.3 Integrated Unordered Discrete-Continuous Choice model

Liu et al. (2014) proposed an integrated model for discrete and continuous choice dimensions with the application to vehicle ownership, type and usage. The model accounts for three different decision variables (first two are discrete choices while the third one is a continuous choice): (1) number of vehicles in a household; (2) vehicle types and vintage for each vehicle in the household with a certain number of vehicles; and (3) total miles traveled by all vehicles in a household. To be consistent, three sub-models are considered in the integrated model: a multinomial probit for vehicle quantity, a MNL for vehicle type and vintage combinations, and a regression for household's total miles traveled. Specifically, the discrete choices of vehicle type and quantity are jointly estimated by treating the logsum of vehicle type logit model as an independent attribute in vehicle quantity probit model; the coefficient of this attribute indicates the impact from the diversity of vehicle types on the number of vehicles holding by households. Because error components of both the discrete and continuous parts follow normal distributions, the model is integrated with error terms following a multivariate normal distribution with full, unrestricted covariance matrix.

A multivariate logit model is adopted to estimate households' vehicle type and vintage combinations. The probability of choosing a certain type/vintage vehicle is:

$$P_{t_k|j} = \frac{exp(V_{t'_k|j})}{\sum_{t_j} exp(V_{t_k|j})} \tag{2.16}$$

where t'_k is the chosen alternative among the full choice set of alternatives t_k. This probability is conditional on the number of vehicles owned by a household (j). Therefore, different models are estimated for households owning 0, 1, 2, and 3 or more vehicles. The expected maximum utility (logsum) that the household would obtain by vehicle type/vintage choices can be written as:

$$J_j = ln \sum_{t_k} \exp(V_{t_k|j}) \tag{2.17}$$

The vehicle ownership model is a multinomial probit and assumes that there are four alternatives in the choice set. The alternatives of owning (0, 1, 2, 3+) cars (or Y_{disc}) have respectively utility (U_0, U_1, U_2, U_{3+}) that consist of one observable part (systematic utility, V) and one non-observable part (error term ε). The observed utility of vehicle quantity is decomposed into two parts $V_j = X_j^T \beta_j$ and $V_{t_k|j} = J_j \lambda$ (j = 0, 1, 2, 3+). V_j is the utility of vehicle holding decision, which depends on factors that vary over j and $V_{t_k|j}$ is the utility of vehicle type choice k conditional on j. The observed utility of zero-car alternative is set to zero for normalization purposes. Therefore, the utility of the discrete choice concerning vehicle holding can be written as:

$$\begin{aligned} U_0 &= \varepsilon_0 \\ U_1 &= X_1^T \beta_1 + J_1 \lambda + \varepsilon_1 \\ U_2 &= X_2^T \beta_2 + J_2 \lambda + \varepsilon_2 \\ U_{3+} &= X_3^T \beta_3 + J_3 \lambda + \varepsilon_3 \end{aligned} \tag{2.18}$$

where, X_1, X_2, ... X_3 are the vector of attributes in the utility functions; β_1, β_2, ... β_3 and λ are the vectors of parameters to be estimated; ε_0, ε_1, ... ε_3 are the error terms. Let $\varepsilon_{disc} = (\widetilde{\varepsilon_1}, \widetilde{\varepsilon_2}, \widetilde{\varepsilon_3})$ and $\widetilde{\varepsilon_j}$ is the difference between ε_j and ε_0, the distribution of ε_{disc} follows a multivariate normal distribution with zero mean and unrestricted covariance matrix.

Regression is adopted to model total miles traveled for each household. In a regression, the dependent variable Y_{reg} is assumed to be a linear combination of a vector of predictors X_{reg} plus some error term (ε_{reg}):

$$Y_{reg} = X_{reg}^T \beta_{reg} + \varepsilon_{reg} \qquad \varepsilon_{reg} \sim N(0, \sigma^2) \tag{2.19}$$

To capture the correlation between the discrete (equation 2.18) and continuous (equation 2.19) parts, the probability of observing Y_{disc} and Y_{reg} can be derived as the product of the probability of observing Y_{reg} and the probability of observing Y_{disc} conditional on observing Y_{reg}.

$$P\left(Y_{disc}, Y_{reg}\right) = P(Y_{reg})P(Y_{disc}|Y_{reg}) \tag{2.20}$$

In the model framework, the error term of the regression (ε_{reg}) is correlated with the differences of error terms from the probit (ε_{disc}), which follows a multivariant normal (MVN) distribution with new mean and variance-covariance matrix.

$$\left(\varepsilon_{disc}, \varepsilon_{reg}\right) \sim MVN(\mathbf{0}, \Sigma) \tag{2.21}$$

Generally, this integrated unordered discrete-continuous choice model accounts for correlations between vehicle quantity, type and vintage, and usage decisions by the error components with a full unrestricted covariance matrix. The model is convenient to implement and transfer to other regions, and is sensitive to taxation policies. The model can be further applied to estimate vehicular emissions and to predict the adoption of clean-fuel vehicles. However, the model system has several limitations: (1) the coefficients of the discrete part is not sufficient because vehicle type logit and vehicle quantity probit models are not estimated simultaneously in one model; (2) Because the probability function in probit model does not have a close form, the number of alternatives in the discrete part should be limited to guarantee the feasibility of model estimation; and (3) the current model framework does not consider the dynamic nature of vehicle ownership and usage behaviors.

2.4 Review of Dynamic Discrete-Continuous Car Ownership Models

2.4.1 Dynamic Discrete-Continuous Choice Model for Car Transition and Use

Glerum et al. (2013) developed a dynamic discrete-continuous choice model (DDCCM) that estimates the joint decision on vehicle transitions, mileage and fuel type in an infinite horizon. The model is formulated as a discrete-continuous choice model that is embedded into a DP framework to account for household's forward looking behaviors in the context of car acquisition. For two-car households, constant elasticity of substitution (CES) utility is adopted to determine the most appropriate allocation of mileage driven by each of the two cars held by households, with a total mileage budget. The integrated model is estimated using the NFP algorithm proposed by Rust (1987) to promise a reasonable computational time.

The DP framework is based on four fundamental elements: the *state space*, the *action space*, the *transition function*, and the *instantaneous utility*. The state space S is constructed based on variables including the age $y_{ctn} \in Y$ and fuel type $f_{ctn} \in F$ of car c of household n in year t. The model assumes each household can have at most two cars. Hence, each state $s_{tn} \in S$ can be represented as:

$$s_{tn} = (y_{1tn}, f_{1tn}, y_{2tn}, f_{2tn}) \tag{2.22}$$

where the car denoted by the index 1 is the oldest car in household n's fleet, and the car denoted by index 2 entered the household in a later stage. The size of the state space depends on the number of ages and fuel types considered. It is important to keep the size as low as possible since the DP problem will be solved repeatedly when estimating the model parameters.

The action space A is constructed based on variables including the transaction $h_{tn} \in H$ in household n's composition of the car fleet in year t, the annual mileage $\widetilde{m}_{ctn} \in R^{+}$ and fuel type $\tilde{f}_{ctn} \in F$ of each car c chosen by household n in year t. Each action $a_{tn} \in A$ can be represented as:

$$a_{tn} = (h_{tn}, \widetilde{m}_{1tn}, \tilde{f}_{1tn}, \widetilde{m}_{2tn}, \tilde{f}_{2tn}) \tag{2.23}$$

Given that a household n is in a state s_{tn} and has chosen an action a_{tn}, the transition function $f(s_{t+1,n}|s_{tn}, a_{tn})$ defines the probability of transferring to the next state $s_{t+1,n}$. The model assumes the transition probability to be degenerate.

Assuming that $a_{tn}^{D} = (h_{tn}, \tilde{f}_{1tn}, \tilde{f}_{2tn})$ gathers the discrete components of an action a_{tn} and $a_{tn}^{D} = (\widetilde{m}_{1tn}, \widetilde{m}_{2tn})$ gathers the continuous components, the instantaneous utility is defined as:

$$u(s_{tn}, a_{tn}^{C}, a_{tn}^{D}, x_{tn}, \theta) = v(s_{tn}, a_{tn}^{C}, a_{tn}^{D}, x_{tn}, \varepsilon_{C}(a_{tn}^{C}), \theta) + \varepsilon_{D}(a_{tn}^{D}) \tag{2.24}$$

where variable x_{tn} contains household socio-economic information, θ is a vector of parameters to be estimated. $v(s_{tn}, a_{tn}^{C}, a_{tn}^{D}, x_{tn}, \varepsilon_{C}(a_{tn}^{C}), \theta)$ is the deterministic part of the utility, $\varepsilon_{D}(a_{tn}^{D})$ and $\varepsilon_{C}(a_{tn}^{C})$ are the unobserved part for discrete and continuous components of actions, respectively.

In previous discrete-continuous choice models (Aguirregabiria and Mira, 2010), the integrated utility function is formulated recursively without a closed-form. In fact, a closed-form formula is possible in the special case where the choice of mileage of each car in the household is assumed myopic. This implied that individuals choose how many mileages they wish to drive with their cars every year, without accounting for the expected discounted utility of this choice for the following years. Under this hypothesis, the integrated utility function is obtained as follows:

$$\bar{V}(s_{tn}, x_{tn}, \theta) = log \sum_{a_{tn}^D} \exp\{\max_{a_{tn}^C}\{ v(s_{tn}, a_{tn}^C, a_{tn}^D, x_{tn}, \varepsilon_C(a_{tn}^C), \theta)\} \quad (2.25)$$

$$+\beta \sum_{s_{t+1,n} \in S} \bar{V}\left(s_{t+1,n}, x_{t+1,n}, \theta\right) f(s_{t+1,n} | s_{tn}, a_{tn}^D)\}$$

Assuming that $v(\cdot)$ is the total utility of vehicle acquisition v_{tn}^D and usage v_{tn}^C, it can be decomposed into two specific parts as follows:

$$v(s_{tn}, a_{tn}^C, a_{tn}^D, x_{tn}, \varepsilon_C(a_{tn}^C), \theta) \quad (2.26)$$

$$= v_{tn}^D(s_{tn}, a_{tn}^D, x_{tn}, \theta) + v_{tn}^C(s_{tn}, a_{tn}^C, a_{tn}^D, x_{tn}, \varepsilon_C(a_{tn}^C), \theta)$$

For the sake of simplicity, the model omits the unobserved random component $\varepsilon_C(a_{tn}^C)$ in the utility function of the continuous choice. As mentioned above, each household is limited to have two cars with a total mileage budget, which motivates the use of a CES utility function to maximize $v_{tn}^C(s_{tn}, a_{tn}^C, a_{tn}^D, x_{tn}, \varepsilon_C(a_{tn}^C), \theta)$ with respect to the annual mileages traveled by each household. Let us denote the mileages of chosen cars with fuel types f_1 and f_2 as $\widetilde{m}_{f_1 tn}$ and $\widetilde{m}_{f_2 tn}$, respectively. The deterministic utility of driving is given by the following CES function, which is valid when $\rho < 1$ and $\rho \neq 0$.

$$v_{tn}^C(s_{tn}, a_{tn}^C, a_{tn}^D, x_{tn}, \theta) = \theta_v(\widetilde{m}_{f1tn}^{\rho} + \widetilde{m}_{f2tn}^{\rho})^{\frac{1}{\rho}} \quad (2.27)$$

The optimal value of mileages for both cars in the household is obtained by solving the following maximization problem, which has two advantages. First, the budget constraint enables us to solve the maximization problem in one dimension. Second, the elasticity of substitution can be directly obtained from the parameter ρ.

$$\max_{\widetilde{m}_{f1tn}, \widetilde{m}_{f2tn}} v_{tn}^C \quad (2.28)$$

$$subject\ to \qquad p_{f1tn}\widetilde{m}_{f1tn} + p_{f2tn}\widetilde{m}_{f2tn} = Inc_{tn}$$

Let v_{tn}^{C*} be the optimal value for the deterministic utility of the continuous actions. It can be inserted back into the integrated utility function, and the Bellman equation becomes:

$$\bar{V}(s_{tn}, x_{tn}, \theta) = log \sum_{a_{tn}^D} \exp\{ v_{tn}^D(s_{tn}, a_{tn}^D, x_{tn}, \theta) + v_{tn}^{C*}(s_{tn}, a_{tn}^D, a_{tn}^{C*}, x_{tn}, \theta) \quad (2.29)$$

$$+\beta \sum_{s_{t+1,n} \in S} \bar{V}\left(s_{t+1,n}, x_{t+1,n}, \theta\right) f(s_{t+1,n} | s_{tn}, a_{tn}^D)\}$$

where $a_{tn}^{C*} = (\widetilde{m}_{1tn}^*, \widetilde{m}_{2tn}^*)$. The integrated utility function can then be computed by value iteration.

The DDCCM is eventually estimated by maximizing the log-likelihood function which is formulated as follows:

$$L(\theta) = \prod_{n=1}^{N} \prod_{t=1}^{T_n} P(a_{tn}^D | s_{tn}, x_{tn}, \theta) \quad (2.30)$$

where N is the total population size, T_n is the number of years household n is observed and $P(a_{tn}^D | s_{tn}, x_{tn}, \theta)$ is the probability that household n chooses a particular discrete action a_{tn}^D at time t.

Although capturing the dynamic nature of vehicle transaction and usage, this model has some limitations: (1) the continuous choice of vehicle mileage is myopic and deterministic, that is, households do not take into account the future expected utility of this choice when they decide how many mileages to drive currently; (2) each household can have at most two cars. Larger household fleets may also be considered but at the cost of increased complexity; and (3) the model has a restrictive assumption that the total mileage budget is exogenously defined.

2.4.2 "Nested Logit" Structural Dynamic Model of Car Choice and Usage

Gillingham et al. (2015) developed a structural microeconomic model to jointly forecast vehicle holding, transaction, and usage in the used-car market, considering the changes of fuel prices and "macro-state" economy. Specifically, a "nested logit" structure is proposed for car choices: the "upper level" models car purchase and type decisions (i.e., not to buy or to buy a car of different types and ages); while the "bottom level" captures trading behaviors of the currently held vehicle (i.e., to sell or to scrap the current vehicle). This structure modeling multiple-layers of discrete choices not only allows for forward-looking behaviors of vehicle type choice, but accounts for endogenous scrappage decisions. For the continuous choice of car usage, the utility of driving is modeled as a 2^{nd}-order polynomial function of annual kilometers traveled. To capture the dynamic nature of car choice and usage decisions, the model specify a stochastic structure of household's income and fuel price by a random walk that follows a log-normal AR(1) process. In addition, the authors outline a "two-stage" strategy to simplify the estimation process: first using Chebychev-polynomials to approximate the expected value function for the continuous state variables; then estimating the remaining parameters by inserting the predicted probability of kilometers driven in the joint likelihood function.

The model is designed for a finite time horizon. Let τ denote the "type" of vehicle and a denote the age, which are used to capture both horizontal and vertical product differentiation. The authors assume a finite number of possible types (gasoline car and diesel car) and ages, denoted as $\tau \in \{1, \dots, \bar{\tau}\}$ and $a \in \{0, 1, \dots, \bar{a}\}$, respectively. Thus, we index the set of cars that consumers can choose from by (τ, a), where τ specifies a particular type of car and a denotes its age. Besides, the model introduces key macro variables that are relevant both for individual choices and for the equilibrium of the market as a whole: defined as (p, m) where p is the current fuel price and m is an indicator of the "macro state" of economy (0 for recession period, 1 for non-recession period). Therefore, consumers' expectations of the price of a typical car (τ, a) when the economy is in state (p, m) are given by the expression $P(\tau, a, p, m)$.

The model focuses on households that own at most one car and households' car choices are updated on an annual basis. At the start of each year, a household makes a decision about whether to buy a new vehicle and/or sell their current vehicle. If a household has an existing vehicle, it cannot purchase another one unless he or she simultaneously sells the existing one. Let $d = (\tau, a)$ denote a household's current car state, where $d = (\emptyset, \emptyset)$ denotes a household does not own any car currently. Let $d' = (\tau', a')$ denote a household purchases a car of type τ' and age a'. If the household chooses not to buy any car, this corresponds to the decision $d' = (\emptyset, \emptyset)$.

Now consider a household that has an existing car $d = (\tau, a) \neq (\emptyset, \emptyset)$. This household actually faces two simultaneous discrete decisions: a sell decision and a buy decision. In order to reflect the sell decision, the authors add a third component d_s to the vector $d' = (\tau', a', d_s)$ where the sell decision d_s takes three possible values: $d_s \in \{-1, 0, 1\}$ where $d_s = -1$ denotes a decision to sell the car for scrap, $d_s = 0$ denotes the decision not to sell the car, and $d_s = 1$ denotes the decision to sell the car in the secondary market. For computational tractability of the model, the authors assume the unobserved factors of these decisions have a multivariate Type 3 generalized extreme value (GEV) distributions that result in a "nested logit" structure for car choice. The upper level of vehicle purchase decisions can be considered as a MNL model with expected future utilities:

$$V_s(d,p,m,x,\varepsilon) = \max_{d' \in D(d)} [v_s(d',d,p,m,x) + \varepsilon(d') + \beta EV_s(d',d,p,m,x,\varepsilon)] \tag{2.31}$$

where $V_s(d,p,m,x,\varepsilon)$ is the value function for a household of age s that owns a car $d = (\tau, a)$ when the "macro state" is m, the fuel price is p, and the household has observed characteristics x and unobserved characteristic factors ε; $v_s(d',d,p,m,x)$ is the indirect utility function; $\varepsilon(d')$ represents the impact of idiosyncratic unobserved factors that affect the consumer's choice; and EV_s is the conditional expectation of $V_{s+1}(\tilde{d},\tilde{p},\widetilde{m},\tilde{x},\tilde{\varepsilon})$ given the current state (d,p,m,x) and decision d'. Since any decision that involves selling the current car d and any unobserved factor that is serially independent do not affect the expected value of future utility conditional on the current choice d', the Bellman formulation of the utility function can be rewritten as:

$$V_s(d,p,m,x,\varepsilon) = \max_{d' \in D(d)} [v_s(d',d,p,m,x) + \varepsilon(d') + \beta EV_s(d',p,m,x)] \tag{2.32}$$

The independently and identically distributed (i.i.d.) extreme value shocks $\varepsilon(d')$ allows for modeling endogenous scrappage decisions in a particularly simple manner. Note that for any alternative d' that involves trading an existing car for another one, the consumer has two possible options: either to scrap the existing car or to sell it in the secondary market. The authors assume a nested logit structure of the unobserved components $\varepsilon(\tau',a',d_s)$ associated with each of the two possible decisions ($d_s = 1$ or $d_s = -1$) for any decision $d' = (\tau',a',d_s)$ involving trading the current vehicle.

Thus, the unobservable components corresponding to the choice of whether to sell or to scrap the currently held vehicle have a bivariate marginal distribution given by

$$F\big(\varepsilon(\tau', a', -1), \varepsilon(\tau', a', 1)\big) = \exp\{-[\exp\left\{-\frac{\varepsilon(\tau', a', -1)}{\lambda}\right\} + \exp\left\{-\frac{\varepsilon(\tau', a', 1)}{\lambda}\right\}]^{\lambda}\} \quad (2.33)$$

where $\lambda \in [0,1]$ is a parameter indexing the degree of correlation in $(\varepsilon(\tau', a', -1), \varepsilon(\tau', a', 1)\ )$. They follow Type 3 extreme value distribution when $\lambda = 1$ and they become highly correlated as $\lambda \to 0$. For each decision d' that involves trading the existing vehicle $d = (\tau, a)$, the consumer will prefer to sell the vehicle in the secondary market if

$$P(\tau, a, p, m)\ +\ \varepsilon(\tau', a', 1) \geq \underline{P}(\tau, p, m)\ +\ \varepsilon(\tau', a', -1) \quad (2.34)$$

where $\underline{P}(\tau, p, m)$ is the scraping value of the car. This implies that conditional on making the upper level choice of trading the current car for another car (τ', a'), the consumer decides to sell his/her current car with the probability of

$$\Pr\{d_s\ =\ 1 | d, d', p, m, x\} = \frac{\exp\{P(\tau,a,p,m)/\lambda\}}{\exp\{P(\tau,a,p,m)/\lambda\} + \exp\{\underline{P}(\tau,a,p,m)/\lambda\}} \quad (2.35)$$

The conditional probability of scrapping the car is just $1 - \Pr\{d_s\ =\ 1 | d, d', p, m, x\}$, and these choice probabilities can be calculated independently of the overall solution of the DP problem given in equation (2.32). Letting $d' = (\tau', a')$, the expected utility of whether to sell or to scrap the current held car can be written as follows:

$$\max\big[v_s\big((d', -1), d, p, m, x\big) +\ \varepsilon(d', -1), v_s((d', 1), d, p, m, x) + \varepsilon(d', 1)\big] =$$

$$\lambda \log\left(\exp\left\{\frac{v_s\big((d', -1), d, p, m, x\big)}{\lambda}\right\} + \exp\left\{\frac{v_s\big((d', 1), d, p, m, x\big)}{\lambda}\right\}\right) + \varepsilon(d') \quad (2.36)$$

where $\varepsilon(d')$ is the Type 3 extreme value random variables with scale parameter $\lambda = 1$ that is distributed independently of $\varepsilon(d)$ for $d' \neq d$. Using this equation, the indirect instantaneous utility $v_s(d', d, p, m, x)$ in equation (2.32) over the two decisions $d_s \in \{1, -1\}$ for any upper level choice $d' = (\tau', a')$ that involves trading the current car for a new one can be redefined as follows:

$$v_s(d', d, p, m, x) = \lambda \log(\exp\left\{\frac{v_s((d',-1),d,p,m,x)}{\lambda}\right\} + \exp\left\{\frac{v_s((d',1),d,p,m,x)}{\lambda}\right\}) + \varepsilon(d') \tag{2.37}$$

To further simplify the Bellman equation by writing it in terms of an "upper level logsum", we can rewrite the expected future utility in equation (2.32) with respect to the independent Type 3 extreme value shocks $\varepsilon(d')$ (simplified as ε') as follows:

$$\begin{aligned} &EV_s(d', p, m, x) \\ &= \int_{\varepsilon'} V_s + 1(f(d'), p', m', x', \varepsilon') q(d\varepsilon') \\ &= \int_{\varepsilon'} \max_{d'' \in D(f(d'))} [V_s + 1(d'', f(d'), p', m', x') + \varepsilon'(d'')] q(d\varepsilon') \\ &= \log(\textstyle\sum_{d'' \in D(f(d'))} \exp\{V_s + 1(d'', f(d'), p', m', x')\} \\ &\equiv \phi(f(d'), p', m', x') \end{aligned} \tag{2.38}$$

where $f(d')$ is given by

$$f(d') = \{\begin{matrix} (\Phi, \Phi) \text{ if } d' = (\Phi, \Phi) \text{ or } d' = (\Phi, \Phi, d_s), d_s \in \{-1,1\} \\ (\tau', \min[\bar{a}, a' + 1]) \text{ if } d' = (\tau', a') \text{ or } d' = (\tau', a', d_s), d_s \in \{-1,0,1\} \end{matrix} \tag{2.39}$$

The model adopts a separate utility function to estimate household's preference on annual kilometers driven. Let k be the total planned kilometers traveled by car over the coming year, and let $p^k(\tau, a, p, c^0)$ be the cost per kilometer traveled. Then, the total costs of driving k kilometers is $p^k(\tau, a, p, c^0)k$. Let $u(k, \tau, a, s, p, m)$ be the conditional direct utility that a household expects from owning a vehicle type τ and driving a planned k kilometers, given by

$$\begin{aligned} u(k, \tau, a, s, p, m) = \theta(y, m)[y - p^k(\tau, a, p, c^0)k - T] + \gamma(y, s, a, m)k + \varphi k^2 \\ - q(a) + \delta_n 1(a = 0) + \delta_\tau \end{aligned} \tag{2.40}$$

where $\theta(y, m)$ is the marginal utility of money; $\gamma(y, s, a, m)$ is the marginal utility of driving; coefficient δ_τ is a car-type fixed effect, δ_n is a coefficient on a new car dummy, and $q(a)$ is a 2^{nd}-order polynomial in car age capturing the rising maintenance costs with car age.

The first-order condition for the optimal kilometers driven implies that

$$k^* = \frac{\theta(y,m)p^{km}(a,\tau) - \gamma(y,s,a,m)}{2\varphi} \tag{2.41}$$

The authors outline a "two-stage" strategy to simplify the estimation of the entire model. In the first stage, Chebychev-polynomial method is used to approximate the expected value function and to estimate the corresponding parameters (i.e., h parameters) of the continuous part. Let $k_{i,t}^*(h)$ denote the predicted kilometers driven for household i at time t. In the second stage, the authors insert this predicted driving from the first stage, and keep the h-parameters fixed while search over the remaining parameters in the utility function of car choice. Formally, the utility of the continuous part is solved as follows:

$$\begin{aligned} u(k^*(h), \tau, a, s, p, m) = \theta(y, m)\big[y - p^k(\tau, a, p, c^0)k^*(h) - T\big] + \gamma(y, s, a, m)k^*(h) \\ + \varphi[k^*(h)]^2 - q(a) + \delta_n 1(a = 0) + \delta_\tau \end{aligned} \tag{2.42}$$

Then, the joint likelihood function of the second stage is given by

$$L^{2step}(\vartheta) = \sum_{i=1}^{N} \sum_{t \in T_i} \log\{\Pr(d_{i,t}|x_{i,t};\ \vartheta, k = k^*(\kappa))[\Pr(d_{i,t,s}|x_{i,t}, k = k^*(\kappa))]^{1\{d_{i,t,s} \neq 0\}}\} \tag{2.43}$$

In a sense, this two-stage estimation approach is similar to thinking of the predicted driving as a characteristic of the chosen car. However, this approach breaks the strict cross-equation restriction that the consumer should care equally much about money spend on buying/ selling a car and money spend on driving a car. Another limitation is that the model only accounts for households with at most one car.

Chapter 3: Data Sources: Survey and Data Description

This Chapter provides a detailed description of data sources used for the dissertation. They include: (a) Maryland Vehicle Stated Preference Survey (MVSPS) data, (b) the 2009 US National Household Travel Survey (NHTS) data, (c) fuel prices from US Energy Information Administration (EIA), (d) vehicle characteristics from Consumer Reviews, (e) Beijing Household Travel Survey (BHTS) data, and (f) GIS shapefiles of Beijing residential location and public transit information. The first four data sources provide information for Maryland State or the Washington DC Metropolitan Area in the US, while the last two data sources deliver information for Beijing in China. In particular, the application in Chapter 4 employs data (a); the application in Chapter 5 employs data (a) and (c); the 1st application in Chapter 6 employ data (a), (b), and (d), while the 2nd application employs data (e) and (f).

3.1 Maryland Vehicle Stated Preference Survey (MVSPS) Data

The main data source used in the dissertation for forecasting vehicle type preferences is the MVSPS data. A stated preference (SP) survey approach was adopted to analyze household vehicle preferences in a dynamic environment. The survey consisted of three parts: (1) current vehicle characteristics, (2) household and respondent characteristics, and (3) a vehicle purchase stated-choice (SC) experiment. The current vehicle characteristics section asked respondents to describe the vehicle that they most often drove. It was assumed that respondents would know the most about this vehicle. The second section of the survey asked respondents to describe their self and their household. Section 3.1.2 provides details on the characteristics that were obtained. The SC experiment section presented respondents with a hypothetical nine-year time frame where they were exposed to various vehicles and asked to keep their current vehicle or acquire a new vehicle. It was assumed that respondents would have the greatest input in decision making for their most driven vehicle.

3.1.1 Survey Design

The survey was conducted under a self-interview, web-based format. Table 3.1 describes the survey methodology employed.

Table 3. 1 Survey Design

Characteristic	Details
Time frame	May-June 2014
Target population	Maryland households
Sampling frame	Households with Internet access in the state of Maryland
Sample design	Recruitment panel
Use of interviewer	Self-administered
Mode of administration	Self-administered via Internet
Computer assistance	Computer-assisted, web-based self-interview
Reporting unit	One person aged 18 or older per household reports for the entire household
Time dimension	Cross-sectional survey with hypothetical longitudinal stated-choice experiment
Frequency	One 2-week phase of collecting responses
Levels of observations	Household, vehicle, person

The SC experiment places respondents in a hypothetical nine-year future time period starting in 2014. Each year includes two scenarios with a total of 18 scenarios possible. In each scenario, respondents are shown a table with the current fuel prices of gasoline and electricity as well as four vehicles – their current vehicle and three new vehicles: a gasoline vehicle, a hybrid electric vehicle (HEV), and a battery electric vehicle (BEV). Respondents then choose whether to keep their vehicle or purchase another vehicle. If the respondents keep their current vehicle, they then go to the next scenario with a new set of vehicles – either the second scenario for the current year or the first scenario for the next year. Otherwise, their chosen vehicle becomes their current primary vehicle and the respondents are accelerated three years into the future. After this acceleration, the respondents are shown the fuel prices in three years and then asked about their satisfaction with their purchase. Then the respondents are returned to the scenario progression with the first scenario for this year (i.e. third year after purchase).

An example sequence is shown in Figures 3.1 through 3.3. The respondent enters in Year 2014. An example scenario is shown for 2015 (Figure 3.1) where the respondent's current vehicle is shown in the left column and new options are shown in the right three columns. The respondent did not purchase a vehicle in 2015, but in 2016, the individual purchased an electric vehicle. The respondent is then accelerated to 2019 where the respondent is shown the new fuel prices in 2019 and asked if he or she is satisfied with his or her purchase (Figure 3.2). The respondent did not purchase a vehicle in 2019 or 2020. The respondent is then shown a scenario for 2021 as shown in Figure 3.3.

In 2015, the following fuel prices are present:

	Gasoline Price	Electricity Price
Fuel Price per Gallon (or equivalent)	**$ 3.83**	**$ 4.28**

And the following vehicles are available:

	Current Vehicle (Runs on Gasoline)	Gasoline Vehicle (Runs on Gasoline)	Hybrid Vehicle (Runs on Gasoline)	Electric Vehicle (Runs on Electricity)
Vehicle Price	**Already Owned**	**$30000**	**$43000**	**$20000**
Fuel Economy	**34 mpg**	**22 mpg**	**28 mpg**	**113 mpge**
Refueling Range	**350 to 450 miles**	**350 to 450 miles**	**350 to 450 miles**	**95 miles**
Vehicle Size	**Small/Compact**	**Mid-Size**	**Large**	**Mid-Size**

Which option would you prefer for your vehicle ownership in 2015?

- I choose to **not buy** a new vehicle.
- I choose to buy the new **Gasoline Vehicle**.
- I choose to buy the new **Hybrid Vehicle**.
- I choose to buy the new **Electric Vehicle**.

>>

Figure 3. 1 Example scenario for year 2015

You have owned your new vehicle for three years and it is now 2019. Fuel prices are now as follows:

	Gasoline Price	Electricity Price
Fuel Price per Gallon (or equivalent)	**$ 4.49**	**$ 4.75**

As a reminder, you purchased the following vehicle:

	Current Vehicle (Runs on Electricity)
Vehicle Price	**$32000**
Fuel Economy	**86 mpge**
Refueling Range	**70 miles**
Vehicle Size	**Mid-Size**

How satisfied are you with your purchase?

- Very Unsatisfied
- Unsatisfied
- Satisfied
- Very Satisfied

>>

Figure 3. 2 Example of time acceleration after a 2016 purchase with satisfaction question

In 2021, the following fuel prices are present:

	Gasoline Price	Electricity Price
Fuel Price per Gallon (or equivalent)	**$ 4.96**	**$ 4.83**

And the following vehicles are available:

	Current Vehicle (Runs on Electricity)	Gasoline Vehicle (Runs on Gasoline)	Hybrid Vehicle (Runs on Gasoline)	Electric Vehicle (Runs on Electricity)
Vehicle Price	**Already Owned**	**$26000**	**$21000**	**$21000**
Fuel Economy	**86 mpg**	**34 mpg**	**54 mpg**	**151 mpge**
Refueling Range	**70 miles**	**350 to 450 miles**	**350 to 450 miles**	**105 miles**
Vehicle Size	**Mid-Size**	**Mid-Size**	**Small/Compact**	**Small/Compact**

Which option would you prefer for your vehicle ownership in 2021?

- ◯ I choose to **not buy** a new vehicle.
- ◯ I choose to buy the new **Gasoline Vehicle**.
- ◯ I choose to buy the new **Hybrid Vehicle**.
- ◯ I choose to buy the new **Electric Vehicle**.

>>

Figure 3. 3 Example of scenario in year 2021 after BEV purchase in 2016

Respondents were given the following instructions for the SC experiment:

- **Make realistic decisions.** Act as if you were actually buying a vehicle in a real life purchasing situation. Take into account the situations presented during the scenarios. If you wouldn't normally consider buying a vehicle, then do not. But if the situation presented would make you reconsider in real life, then take them into account.
- **Assume that you maintain your current living situation** with moderate increases in income from year to year.

- **All prices are adjusted for inflation.** All dollar amounts are in terms of the value of money in 2014. For example, a vehicle priced at $20,000 in 2014 should be of similar quality as a vehicle priced at $20,000 in 2022.
- **Each scenario is dependent on your previous choices.** If you purchase a vehicle, it will replace your current vehicle in future scenarios.

The attributes of an alternative depended on the alternative's vehicle size: small, medium, or large. Vehicles sizes were determined randomly on each draw and the probability of choosing a size depended on the respondent's primary vehicle size. When a respondent's primary vehicle was small:

- A gasoline and hybrid vehicle alternative had a 3/7 chance of being small, 3/7 chance of being medium, and a 1/7 chance of being large
- An EV alternative had a 1/2 chance of being small and a 1/2 chance of being medium

When a respondent's primary vehicle was medium:

- A gasoline and hybrid vehicle alternative had a 1/4 chance of being small, 1/2 chance of being medium, and a 1/4 chance of being large
- An EV alternative had a 1/2 chance of being small and a 1/2 chance of being medium

When a respondent's primary vehicle was large:

- A gasoline and hybrid vehicle alternative had a 1/7 chance of being small, 3/7 chance of being medium, and a 3/7 chance of being large

- An EV alternative had a 1/3 chance of being small and a 2/3 chance of being medium

Attribute levels were set for the base year as shown in Table 3.2. The refueling range and the fuel economy of new vehicles change over time to mimic a dynamic marketplace. Each attribute level changes depending on the vehicle type and size. The fuel economy of gasoline and hybrid vehicles increases annually by 2 miles per gallon (MPG) for small and medium size vehicles and by 1 MPG for large vehicles. For the BEV, the fuel economy increases annually by 3 miles per gallon equivalent (MPGE) and refueling range increases annually by 5 miles.

Table 3. 2 Vehicle Attribute Levels by Type and Size for Base Year 2014

Attribute	Small Size Vehicle			Medium Size Vehicle			Large Size Vehicle		
	GasV	HEV	BEV	GasV	HEV	BEV	GasV	HEV	BEV
Price ($)	13000	18000	18000	18000	20000	20000	28000	28000	–
	16000	21000	21000	22000	24000	24000	35000	35000	–
	20000	24000	24000	26000	28000	28000	43000	43000	–
	25000	28000	28000	30000	32000	32000	50000	50000	–
Fuel Economy (mpg or	24	40	100	20	35	80	14	20	–
	28	44	110	24	40	90	18	23	–
	33	48	120	28	45	100	22	27	–
	38	52	130	32	50	110	26	30	–
Range (mi)	–	–	60	–	–	60	–	–	–
	–	–	70	–	–	75	–	–	–
	–	–	85	–	–	90	–	–	–
	–	–	100	–	–	110	–	–	–

Note: GasV = gasoline vehicle, HEV = hybrid electric vehicle, BEV = battery electric vehicle, mpg = miles per gallon, mpge = miles per gallon equivalent
– denotes that this attribute was not used for this vehicle type and size

The experimental design was generated using the support.CE package in R (Aizaki, 2012). The design was for a three-alternative, three-attribute experiment with four levels per attribute. In the design setup, the attribute levels were used as placeholders that would be filled in by the corresponding year and vehicle size attribute levels chosen for a particular scenario. For example, a label for the third MPG attribute level would result in a shown attribute level corresponding to the third attribute for an alternative of a given size (e.g. using Table 3.2, for year 2014 and size medium, the hybrid MPG shown in a scenario would be 45). This resulted in 32 scenarios being generated which were broken up into two blocks of size 16. Each respondent was given one of these scenario blocks (with equal probability) and then this list was increased to 18 scenarios but randomly selecting two scenarios to repeat. This list was then permutated to obtain an experimental design for the respondent.

The price sequence for gasoline and electricity was created using a random walk dependent on the previous year price and a random draw. The random draws were generated using a modified Irwin-Hall distribution centered at zero and ranging between -1 and +1. The draws were generated by:

1. Summing eight independent uniform draws between 0 and 1
2. Dividing this sum by 4
3. Subtracting this quantity by 1

This was done to ensure that the random shock was bounded and to provide draws that were approximately normally distributed. For the initial year of 2014, a price was randomly chosen (in increments of $0.25) between $3 and $4.25 for a gallon of gasoline and between $4.00 and $5.25 for an energy-equivalent amount of electricity. For each year that followed, the new gasoline and electricity prices were generated as follows:

$$p_t^{gas} = p_{t-1}^{gas} + 0.3 + 1.1 * \eta_t^{gas} \tag{3.1}$$

$$p_t^{elec} = p_{t-1}^{elec} + 0.2 + 0.5 * \eta_t^{elec} \tag{3.2}$$

$p_t^{gas}, p_t^{elec} =$ the fuel price in year t for gasoline and electricity respectively

$\eta_t^{gas}, \eta_t^{elec} =$ the fuel price shock in year t for gasoline and electricity respectively as generated through the modified Irwin-Hall distribution described above

Gasoline prices were assumed to be less stable than electricity prices due to Maryland state regulations in the electricity market. Although these random walks induce both gasoline price and electricity price to have an expected increase annually, gasoline prices are more likely to experience a possible annual decline as compared to electricity prices. This is also intended to mimic the instability of the gasoline market.

3.1.2 Sample Characteristics and Time-Dependent Vehicle Preferences

The dataset, collected from the web-based survey described in Section 3.1.1, contains 3,598 observations of vehicle type choices from 456 households resident in Maryland, US. Table 3.3 and Table 3.4 show descriptive statistics of the collected data, and the comparison between the sample and the population in Maryland. Table 3.3 presents household socio-demographic variables such as gender, age, education level, income, work status, drive to work or not, commuting distance, and home type. The sampling method enforces a nearly even split between male and female respondents. The average education level of households in the sample is higher than that in the population. In addition, the average number of adults and workers within households are slightly higher compared with the population. The percentage of respondents who drive to work is slightly lower than in the population. The percentage of households with extremely high income ($250,000 or more) is lower than in the population. Although there are distinctions, the pattern of household information between the sample and the population is quite similar, which suggests that the data collected from the survey can be representative of the population in Maryland.

Table 3. 3 Household Information

Attributes	Category	Sample Share	Population Share	Attributes	Statistics	Sample Value	Population Value
*Gender	Male	49.9%	48.4%	Commute distance, daily miles traveled	Min.	0	0
	Female	50.1%	51.6%		Max.	120	120
*Age	0-17 years old	0.2%	0.0%		Mean	23.96	32.40
	18-24 years old	10.0%	12.7%		Median	20	26
	25-34 years old	22.2%	17.2%		S.D.	19.39	25.86
	35-44 years old	14.5%	18.0%	Num. of adults	Min.	1	1
	45-54 years old	20.3%	20.2%		Max.	5	5
	>=55 years old	32.8%	31.9%		Mean	2.08	1.87
*Education Level	Less than High School	1.3%	11.78%		Median	2	2
	High School Diploma or GED	17.7%	26.46%		S.D.	0.88	0.68
	Some College	23.5%	21.07%	Num. of workers	Min.	0	0
	Associate Degree	11.7%	7.26%		Max.	5	4
	Bachelor Degree	25.5%	19.10%		Mean	1.54	1.21
	Graduate or Professional	20.3%	14.33%		Median	1	1
Work status	Working Full Time	50.5%	51.6%		S.D.	0.93	0.86
	Working Part Time	13.2%	4.2%	Home parking	Personal Garage	23%	-
	Looking for Work	7.3%	2.5%		Personal Driveway	40.8%	-
	Homemaker	8.9%	9.5%		On-Street	16.5%	-
	Going to School	2.6%	3.4%		Outdoor Parking Lot	15.0%	-
	Retired	13.8%	24.6%		Parking Garage	3.7%	-
	Other	3.7%	4.2%		Other	1.1%	-
*Household income	$0 to $24,999	11.3%	15.2%	Household head	Yes	78.1%	-
	$25,000 to $49,999	24.0%	18.5%		No	21.9%	-
	$50,000 to $74,999	24.6%	17.5%	Driver's license	Yes	95.2%	-
	$75,000 to $99,999	15.5%	13.7%		No	4.8%	-
	$100,000 to $149,000	16.3%	18.2%	*Drive to	Drive to work	84.7%	83.6%

	$150,000 to $249,999	6.8%	8.6%	commute	Drive to transit	5.5%	8.8%
	$250,000 or more	1.5%	8.3%		Not Drive	9.8%	7.6%
*Home type	College Dorm	0.6%	24.2%				
	Apartment	14.0%					
	Condominium	4.3%					
	Townhouse	15.8%	20.8%				
	Rowhouse	3.9%					
	Single-family Home / Detached /Separated House	59.8%	53.5%				
	Other	1.5%	1.5%				
Note: Attributes that start with "*" are compared with American Fact Finder, and other attributes are compared with 2009 NHTS data.							

Considering that households' current vehicle characteristics will affect their vehicle purchasing decisions, Table 3.4 summarizes their vehicle characteristic statistics. Compared with the population, the share of hybrid vehicle in the sample is slightly higher. Thus, the average fuel economy (i.e., MPG) is higher as expected. Half of the vehicles are of medium size and the average number of vehicles within households is fewer in the sample. The sample underestimates the shares of households in two extreme status – without vehicle and with three or more vehicles. Additionally, the sample has more vehicles that are less than three years old.

Table 3. 4 Household Current Vehicle Characteristics

Attributes	Category	Sample Share	Population Share	Attributes	Statistics	Sample Value	Population Value
Hybrid	Yes	6.3%	3.4%	*Num. of Vehicles	No vehicle	0%	4.5%
	No	93.7%	96.6%		1 vehicle	42.5%	21.3%
Vehicle size	Small/Compact	25.5%	-		2 vehicles	41.7%	41.0%
	Mid-Size	51.7%	-		3 vehicles	12.3%	33.3%
	Large	22.7%	-		4 or more	3.5%	
Vehicle MPG	Min.	12	6.4	Vehicle price dollar	Min.	500	-
	Max.	57	57.2		Max.	140000	-
	Mean	26.63	21.79		Mean	20034	-
	Median	26	20.8		Median	20000	-

	S.D.	7.96	7.82		S.D.	12266	-
Model year	Min.	1984	1974	Model year	2011-2014	27.4%	18.4%
	Max.	2014	2009		2008-2010	22.9%	26.7%
	Mean	2007	2001		2004-2007	25.8%	29.3%
	Median	2008	2002		Before 2004	23.8%	25.6%
	S.D.	5.16	5.88				
Note: Attributes that start with "*" are compared with American Fact Finder, and other attributes are compared with 2009 U.S. NHTS data.							

Besides the information on household socio-demographics and vehicle characteristics, decisions to purchase a new vehicle over time provide essential evidence to capture households' preference on switching to new vehicles. According to the survey design, if respondents make a purchase, they will directly jump to the first scenario three years later. Thus, the maximum number of purchases can be made by respondents over a nine-year period is three. From Figure 3.4, we can observe that over the nine-year period with a total of 18 scenarios, around 11% (50) of households always retain their current vehicle, while the shares (number) of households making one, two, and three purchases are around 14% (64), 24% (108), and 51% (234), respectively.

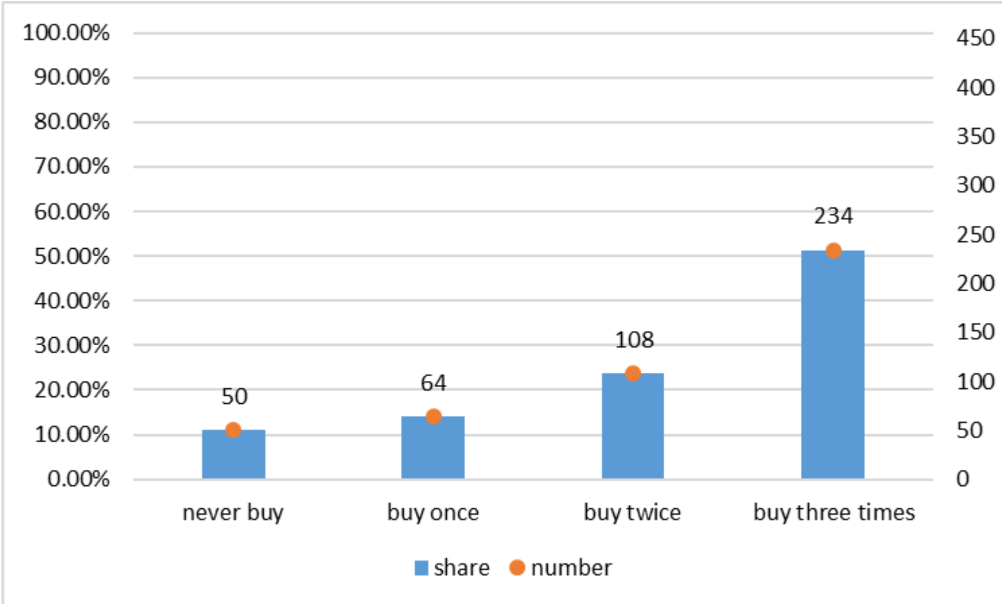


Figure 3. 4 The number (share) of households who never buy a new car, buy once, buy two times, and buy three times

More specifically, at each scenario, households who make a purchase within the previous three years will be out-of-market. Only households in the market have an opportunity to decide either to retain their current car or to buy a new car. To better understand the dynamic nature of households' vehicle type choices, Figure 3.5 reports the shares of households who are in the market, retain their current car, buy a gasoline car, buy a hybrid car, and buy an electric car over the 18 scenarios. In Figure 3.5, the share of households in the market shows a periodic pattern every six scenarios, starting from the first scenario. This can be explained by the assumption of the survey design that once households make a purchase, they will be out-of-market for three years (six scenarios). For example, a large percentage of households makes a purchase at scenario 1, forced to be out-of-market from scenario 2 to 6, then comes back at scenario 7.

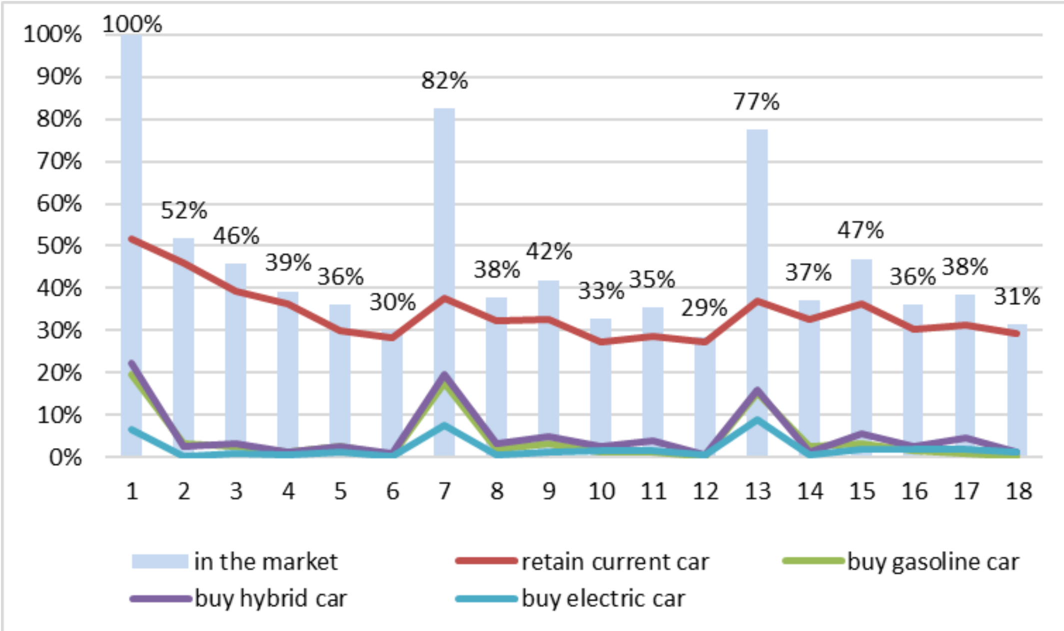
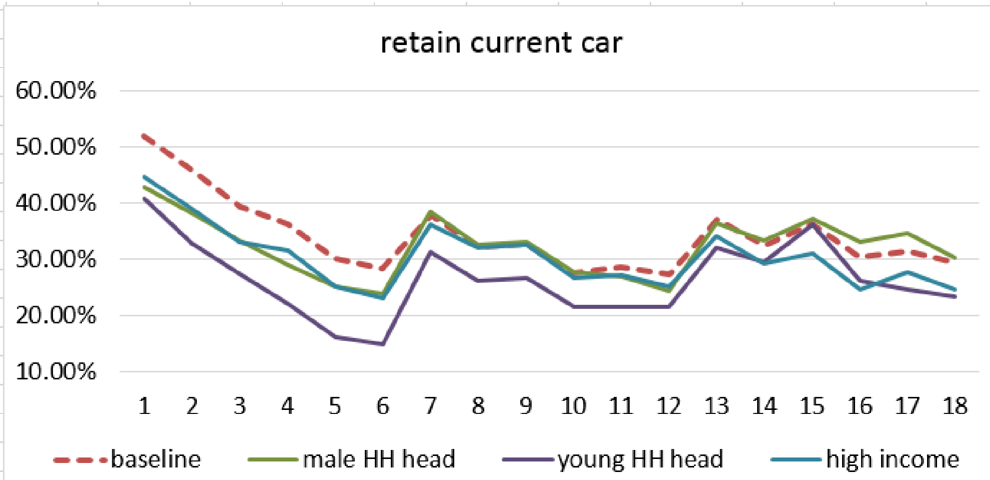
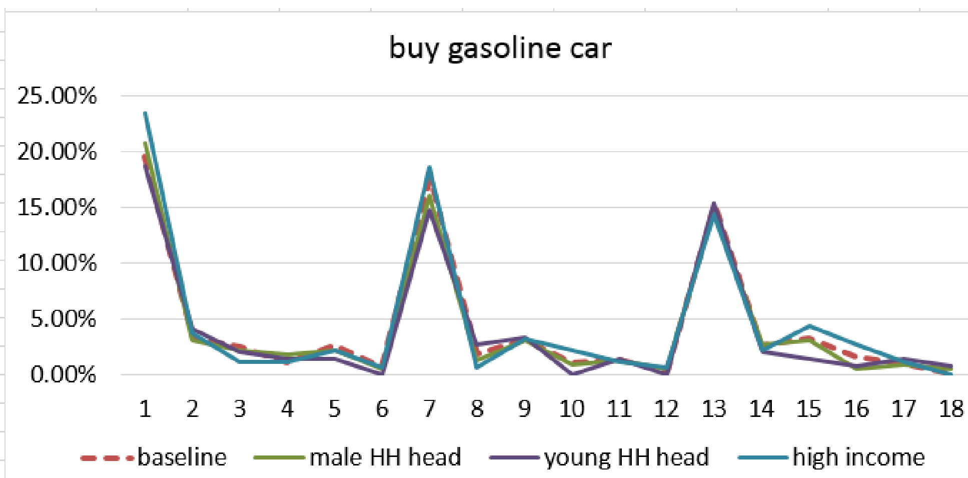


Figure 3. 5 Shares of households in the market and choose different vehicle types over 18 scenarios

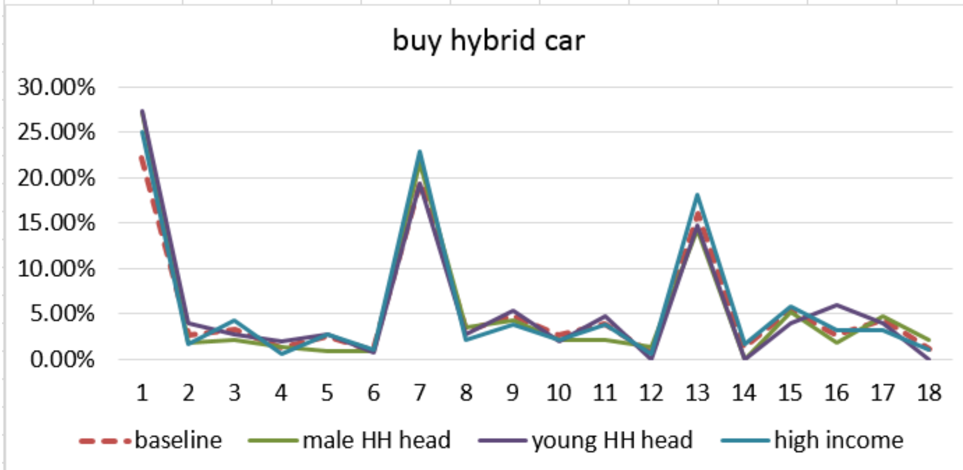
In practice, it's also important for policy makers and automobile producers to identify the potential purchasing group of different vehicle types, especially greener vehicles. Figure 3.6 compares households' time-dependent vehicle type choices between different groups of households. These groups include households: (1) with male household head (male HH head), (2) with household head younger than 35 years old (young HH head), and (3) with annual income more than $75,000 (high income). They are compared with the vehicle type choices of the entire sample (baseline). Observing Figure 3.6 (A) – (D), it's obvious that the group with young household head tends to purchase new cars and has a much high preference on electric ones. The group with male household head tends to purchase new cars only during the first six scenarios and has a moderately high preference on electric cars. For these two groups of households, the choice patterns on gasoline and hybrid cars are quite similar to those of the entire sample. In terms of the high-income group, households would like to purchase new cars as well, but they have a higher preference on gasoline and hybrid cars and are less likely to switch to electric cars. Besides, other socio-demographic variables are also investigated including education level. Compared with the baseline, fewer differences in vehicle type choices are observed for these variables.



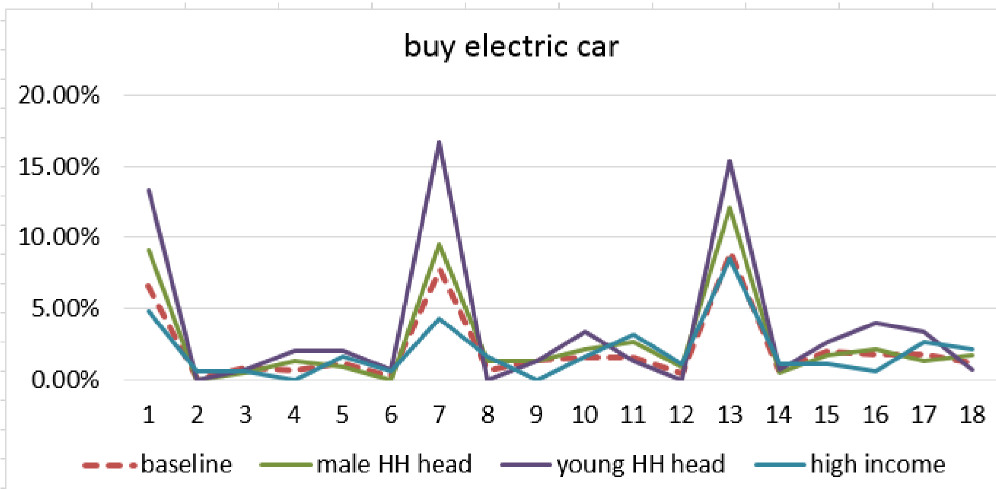
(A)



(B)



(C)



(D)

Figure 3. 6 Households' time-dependent vehicle type choices varying by socio-demographic indicators (gender, age, income)

3.2 US National Household Travel Survey (NHTS) Data

The above MVSPS data does not contain vehicle use information (i.e., annual VMT) which is necessary for the 1st application in Chapter 6. Therefore, the annual vehicle use data is estimated based on the 2009 NHTS data, assuming that households in these two data sources share the same vehicle use pattern.

The NHTS is conducted as a telephone survey, using Computer-Assisted Telephone Interviewing technology. Collected in 2009, the revealed preference (RP) dataset includes all interviews from the national sample and the Add-on partners. The weighting factors have been adjusted to account for the oversampling in the Add-on areas. The 2009 NHTS data is organized into four different data files, including household record, vehicle record, person record, and travel day trip record.

For this study, I mainly interested in the household travel information in the Washington DC Metropolitan area. After data processing and cleaning, 1289 household records are available for the study area. The data file mainly contains information on households' characteristics (i.e. income level, number of adults, number of workers, number of drivers, age, gender, and education level), car ownership (i.e. number of household cars, vehicle make, model, and model year), land use (i.e. housing units per square mile, and population per square mile), and car use (i.e. annual VMT and fuel cost per mile). As shown in Figure 3.7, the average annual VMT for households with one, two, and three or more vehicles are 10168, 25321, and 36855 miles, respectively. Additionally, the number of workers, drivers, and annual VMT increase with household size.

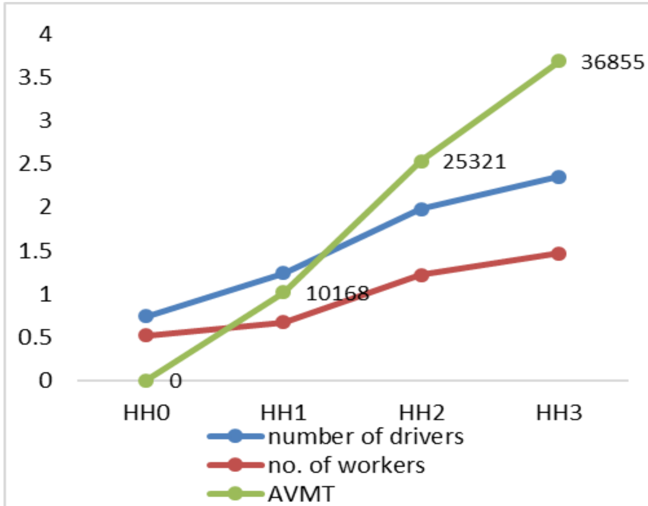


Figure 3. 7 Relationship between household size and annual VMT

3.3 US Energy Information Administration (EIA): Fuel Prices

To capture the evolving nature of fuel prices in the real market, we employ a historical dataset from US EIA, including weekly and monthly gasoline prices from April 1993 to September 2015, and monthly electricity and gasoline prices from January 2003 to September 2015 in Maryland. The unit used for gasoline price is “dollars per gallon”, while the unit for electricity price is transferred to “dollars per one-gallon-equivalent electricity”.

Figure 3. 8 Changes of gasoline and electricity prices from 2003 to 2015

Figure 3.8 plot the changes of gasoline and electricity prices between the year of 2003 and 2015. In terms of gasoline price, there is a climbing trend from 2003 to 2008, and two valleys in the year of 2009 and 2015 due to economic recessions in the US and European markets. More interesting, the fluctuations of electricity price suggest that the relationship between gasoline and electricity is more like substitutes.

3.4 Consumer Reviews: Vehicle Characteristics

Vehicle characteristics, which are important for vehicle type choice estimation in the 1st application of Chapter 6, are collected from two sources; characteristics of gasoline vehicle are from the Consumer Reports, and characteristics of hybrid or electric vehicle are from KBB Consumer Reviews. The collected data includes vehicle specification attributes such as vehicle price, seating space, engine size, transmission, acceleration, shoulder room, etc., which associate with vehicle type decisions.

3.5 Beijing Household Travel Survey (BHTS) Data

To study the behavior of car ownership and use in large metropolitan areas of developing countries such as China, we employ the BHTS data for the 2nd application in Chapter 6.

3.5.1 BHTS Data Description

The Beijing municipal government has organized four large-scale household travel surveys respectively in 1986, 2000, 2005 and 2010. The recent 2010 Survey adopts a multistage sampling strategy with a target of 1% sampling rate (Gu et al., 2015). 1,085 out of 1,911 Traffic Analysis Zones (TAZs) in the whole Beijing City are selected. A face-to-face interview was given to 46,900 households living in this area with a total of 116,142 individuals.

To analyze the relationship between private vehicle ownership/use and public transit use, this study focuses on households who live in the eight districts (Dongcheng, Xicheng, Xuanwu, Chongwen, Chaoyang, Haidian, Fengtai and Shijingshan) within the 5th Ring Road, with a total of 18,492 households being considered. The share of male in the sample is 55.53%, slightly higher than that of the 2010 Census data in Beijing. The average number of household members and workers are 2.46 and 1.2 respectively in the sample, consistent with the Census. Around 80% of the households live in apartments or houses bought by themselves or supported by their companies instead of renting. The average family housing size is about 69.4 square meters. 61.35% of households have an annual income less than 50,000 CNY and 8.45% of households have an annual income over 100,000 CNY. In the sample, the majority of household heads have high school degree or higher education levels. Compared to Census, more elderly people are observed in the sample. It is observed that 26.34% of the respondents have a driver license, and 93.26% of them have a discount public transit pass (IC card). Detailed descriptive statistics are shown in Table 3.6.

In addition, Table 3.6 provides detailed information on household vehicles. In the sample, 24.46% of the households have one car and 3.71% have at least two cars; the percentages of car owners in the sample are much smaller than those reported in the Census data (55%). The percentages of households owning bicycle, electric bicycle and motorcycle are 61.05%, 12.30% and 2.21%, respectively. Besides, the average gasoline cost per month is 700.77 CNY and the average vehicle engine displacement is 1.69 liter. The average annual kilometers traveled per household is 13,776 km.

3.5.2 Sample Selection: Stratified Random Sampling (SRS) Method

From Table 3.6, we can observe some differences in the structure of household socioeconomics and car ownership rates across the sample and the Census, which may result from selection bias when the sample was collected. Therefore, a stratified random sampling (SRS) strategy is employed here to derive a more representative sample of households in Beijing. SRS prevents any serious selection bias by matching the share of households falling into each stratum with those reported in the Census data (Lohr, 2009).

We divide the sample of $\mathrm{N}\ (= 18{,}492)$ households into H strata, with $\mathrm{N_h}$ households in stratum h. For stratified sampling to work, the strata must constitute the entire sample of N households so that each household belongs to exactly one stratum, defined as $\mathrm{N_1 + N_2 + \cdots + N_H = N}$. The share of households in stratum h is $\mathrm{S_h = N_h/N}$.

Using SRS method, we independently take a random sample from each stratum, so that $\mathrm{n_h}$ observations are randomly selected from $\mathrm{N_h}$ households in stratum h. The target sample size is $\mathrm{n = n_1 + n_2 + \cdots + n_H}$ (Lohr, 2009), and the target share of observations in stratum h is $\mathrm{s_h = n_h/n}$, same as the share in Census.

The number of strata H is determined by the attributes selected by the analyst. For example, if Gender and Age are considered, six strata will be produced which are combinations of two genders (male and female) and three age ranges (younger than

30, 30 to 64, and older than 64 years old), as described in Table 3.5. With reference to a final sample size of 5000, the number and share of observations in six strata are reported as follows:

Table 3. 5 A Simple Example of SRS

Stratum	Gender	Age	N_h	S_h	n_h	s_h
1	Male	< 30	2722	8.31%	1026	20.53%
2	Male	30 - 64	12111	36.99%	1313	26.26%
3	Male	>= 65	3314	10.12%	217	4.35%
4	Female	< 30	2222	6.79%	961	19.22%
5	Female	30 - 64	9447	28.85%	1231	24.63%
6	Female	>= 65	2927	8.94%	251	5.02%

For this study, a two-step SRS strategy is employed to determine the sample for estimation. The attributes considered are households' car ownership distribution, gender, and residential location. In the first step, we define three strata categorized by the number of cars in each household – zero car, one car, and two or more cars. In the second step, there are sixteen strata which are combinations of households' gender (male and female) and residential location (eight districts). The obtained stratified random sample consists of 8,540 observations. Table 3.6 compares household social-demographics and vehicle-related variables between the two samples before and after applying SRS strategy, as well as Census data. The distributions of household's residential location, gender, age, and number of cars in the sample after applying SRS strategy are shown to be more appropriate compared to the Census.

Table 3. 6 Comparison of Descriptive Statistics between the Two Samples and Census

Attributes	Category	Before	After	Census	Attributes	Category	Before	After
Households' Social-demographics and Land Use Information								
Gender	Male	55.53%	52.21%	51.13%	Household	One member	14.41%	14.52%

	Female	44.47%	47.79%	48.87%	size	Two members	41.60%	35.44%
Age	<30 years old	15.10%	25.20%	39.75%		Three members	31.60%	35.32%
	30-64 years old	65.84%	65.67%	50.88%		Four members	8.29%	9.94%
	>= 65 years old	19.06%	9.13%	9.37%		Five or more	4.10%	4.78%
Education Level	Not educated	1.47%	0.09%		Number of workers	No workers	28.96%	11.31%
	Elementary school	5.81%	0.97%			One worker	31.45%	33.98%
	Middle school	19.41%	10.49%			Dual workers	34.32%	45.50%
	High school	27.22%	27.13%			Three workers	4.62%	7.99%
	Junior college	17.73%	28.63%			Four workers	0.65%	1.22%
	Undergraduate or higher	28.37%	32.68%		Work status	Working Full Time	52.95%	74.41%
Household income	0-50,000yuan	61.35%	46.93%			Working Part Time	1.65%	3.29%
	50,000-100,000	30.20%	38.43%			Looking for Work	6.74%	0.21%
	100,000-150,000	5.68%	9.38%			Homemaker	1.40%	0.05%
	150,000-200,000	1.64%	3.20%			Going to School	0.59%	1.01%
	200,000-250,000	0.53%	0.94%			Retired	36.23%	20.91%
	250,000-300,000	0.25%	0.44%			Other	0.45%	0.12%
	300,000 or more	0.36%	0.68%		Home Ownership	Private owned	64.25%	59.77%
Residential location	Dongcheng	7.44%	4.72%	4.68%		Company owned	15.54%	12.00%
	Xicheng	9.54%	5.75%	5.62%		Lodge housing	1.82%	2.25%
	Chongwen	6.35%	3.10%	3.06%		Rented housing	17.88%	25.56%
	Xuanwu	7.10%	5.30%	5.23%		Low-rent housing	0.18%	0.18%
	Chaoyang	21.23%	31.79%	32.42%		Other	0.33%	0.25%
	Haidian	22.40%	24.05%	24.02%	Home type	Formal building	85.57%	87.48%
	Fengtai	17.30%	19.70%	19.56%		Moderate building	0.22%	0.16%
	Shijingshan	8.64%	5.59%	5.41%		Simplified building	2.80%	2.12%
Live > half year	No	1.76%	2.03%			House	0.02%	0.02%
	Yes	98.24%	97.97%			Apartment	0.16%	0.21%
Driver's license	Yes	26.34%	40.98%			One-floor house	11.15%	9.93%
	No	72.10%	57.13%			Other	0.07%	0.07%
	Missing	1.56%	1.89%		IC Card	Yes	93.26%	91.72%
						No	6.74%	8.28%
Households' Vehicle-related Information								
Number of Cars	No cars	71.83%	55.74%	45%	Vehicle Engine Displacement	Min.	0.80	0.80
	One car	24.46%	37.90%	55%		Max.	9.90	4.4
	Two or more	3.71%	6.36%			Mean	1.69	1.72

Bicycle	Yes	61.05%	63.47%		(liter)	S.D.	0.40	0.42
	No	38.95%	36.53%		Annual Kilometers Traveled	Min.	0	200
Fuel Cost per month (￥)	Min.	30.00	30.00			Max.	140000	98000
	Max.	5000.00	5000.00			Mean	13776	13830
	Mean	700.77	730.77			S.D.	9278	8819
	S.D.	430.86	449.36					

Note: "Before" represents the entire sample before applying SRS strategy, the sample size is 18,492; "After" represents the sample after applying SRS strategy, the sample size is 8,540.

3.6 Proxy of Beijing Public Transit Services

Four data sources are used to calibrate indicators of public transit services in Beijing, China: the BHTS data, GIS shapefile of 1911 traffic analysis zones (TAZs) in Beijing in 2010, GIS shapefile of public bus stops in Beijing in 2010, and GIS shapefile of Beijing metro network in 2015 (adjusted to December 2010). The BHTS dataset and GIS shapefiles are linked by a key ID – TAZ reference number. Specifically, the GIS shapefile of public bus stops in 2010 contains more than 800 bus routes and 57,250 bus stops in Beijing. The adjusted shapefile of Beijing metro network includes 12 metro lines. The integrated GIS shapefiles include spatial information on the length of metro lines and the number of metro stations/bus stops in each TAZ. The temporal information such as service duration and headway of each metro line are obtained from Wikipedia. To measure public bus and metro services, four spatial indicators and one temporal indicator are derived from the data, including (1) density of bus stops, (2) density of metro stations, (3) percentage of bus stop coverage, (4) percentage of metro station coverage; and (5) metro service index (MSI) for each TAZ.

The density of bus stops and metro stations are defined as follows: a scale factor is applied to guarantee these attributes have similar order of magnitude.

$$Density\ of\ bus\ stops\ in\ TAZ\ i = \frac{number\ of\ bus\ stops\ in\ TAZ\ i}{total\ area\ of\ TAZ\ i} \times scale\ factor \quad (3.3)$$

$$Density\ of\ metro\ stations\ in\ TAZ\ i = \frac{number\ of\ metro\ stations\ in\ TAZ\ i}{total\ area\ of\ TAZ\ i} \times scale\ factor \quad (3.4)$$

The percentage of bus stop coverage or metro station coverage follows the Transit Capacity and Quality of Service Manual (TCQSM) recommendations with adjustments for developing countries such as China. In particular, a service buffer is created for each bus stop or metro station to identify the area where potential public transit users are located. Centered at a bus stop or a metro station, the service buffer is defined as the circular area with a radius of 500 meters or 800 meters respectively. The radius is determined to include the majority of walking trips to a bus stop or metro station based on willingness to travel studies. The specific formula is defined as follows:

$$Coverage\ of\ bus\ stop\ in\ TAZ\ i = \frac{area\ of\ the\ union\ of\ bus\ stop\ service\ buffer\ in\ TAZ\ i}{total\ area\ of\ TAZ\ i} \quad (3.5)$$

$$Coverage\ of\ metro\ station\ in\ TAZ\ i = \frac{area\ of\ the\ union\ of\ metro\ station\ service\ buffer\ in\ TAZ\ i}{total\ area\ of\ TAZ\ i} \quad (3.6)$$

MSI accounts for both spatial and temporal measurements of metro service (Liu, 2013). With the knowledge of daily service duration and headway of each metro line, the MSI is formulated as follows:

$$MSI\ in\ TAZ\ i = \frac{Coverage\ of\ metro\ station\ in\ TAZ\ i}{average\ metro\ headway} \times metro\ service\ duration \tag{3.7}$$

In this analysis, we do not calculate bus service index because the time schedule of buses in Beijing is very unstable and highly depends on road congestion. Besides, at the current stage, we are missing the data of daily service duration and headway for many bus routes in 2010.

Chapter 4: Methodology Part 1: Mixed Multinomial Logit Model

4.1 Introduction

Technological innovation is a major driving force in the automotive industry and related sectors. In recent years, automakers have introduced several innovations for their products and adopted innovative technologies to achieve CO2 emissions, fuel economy, and performance goals (EPA, 2016). More improvements are expected in the short and medium-long run. Meanwhile, customers will be confronted to the choice of buying more efficient and less polluting vehicles.

This Chapter introduces the framework of a mixed multinomial logit (MMNL) model with panel effect to analyze consumers' preference on gasoline, hybrid, and electric vehicles. Given that the actual market shares of advanced technology vehicles are low and that rapid changes are only expected on the supply side, it is not surprising that many studies on vehicles with new technologies are based on stated preference (SP) data (Hensher, 1994) especially for the US market. The study in this Chapter bases on the MVSPS data that places respondents in a nine-year hypothetical time period, with car characteristics changing over time to mimic the dynamic vehicle market. The MMNL model is estimated on the SP dataset; the estimation coefficients have been applied to calculate vehicle market elasticities with respect to price and consumers' willingness-to-pays for improving car characteristics. The study shows that respondents are able to consider trade-offs between gasoline, hybrid, and electric cars over an extended time horizon.

4.2 Taste Heterogeneity and Panel Effect

Discrete choice models are used to estimate households' vehicle preferences over a nine-year time horizon and across the four alternatives presented to each respondent: keep the current vehicle, buy a new gasoline vehicle, buy a new HEV, and buy a new BEV. Besides MNL models, we adopted MMNL models with panel effect in order to account for random taste heterogeneity in model coefficients, correlation across alternatives, and state dependency due to repeated measurements over time. The general utility function of the MMNL model is formulated as follows:

$$U_{njt} = \beta X_{njt} + \mu Z_{nj} + \varepsilon_{njt} \tag{4.1}$$

where:

U_{njt} represents the utility of individual n choosing alternative j at time t;

β is a vector of either fixed (β_0) or random coefficients (β_n) corresponding to a sequence of attributes X_{njt};

μ is a vector of normally distributed random terms with zero mean;

Z_{nj} are error components that specify the correlation pattern;

and ε_{njt} is a vector of error terms that are i.i.d. type 1 extreme value (T1EV) over different households, alternatives, and time periods.

The random coefficients vary over the population with density $f(\beta)$ (Train, 2009), where the parameters of the density depend on the specification of the random distributions adopted by the analyst.

In order to account for the panel nature of the dataset the likelihood that household n makes a sequence of choices over time is the product of logit formulas:

$$L_{nj}(\beta) = \prod_{t=1}^{T_n} \frac{\exp(\beta X_{njt} + \mu Z_{nj})}{\sum_{i \in C} \exp(\beta X_{nit} + \mu Z_{nj})} \tag{4.2}$$

where C is the choice set including all alternatives; T_n is the number of time periods for household n. The unconditional probability that household n choosing alternative j (P_{nj}) is the integral of the product of logit probabilities over all values of β:

$$P_{nj} = \int \left[\prod_{t=1}^{T_n} \frac{\exp(\beta X_{njt} + \mu Z_{nj})}{\sum_{i \in C} \exp(\beta X_{nit} + \mu Z_{nj})}\right] f(\beta) d\beta \tag{4.3}$$

The dimension of the integrals equals the number of random coefficients and error components in the model specification.

4.2 Application: Measuring Vehicle Type Preference in Maryland

4.2.1 Model Estimation Results

Four specifications are proposed to model vehicle type preferences; results are presented in Table 4.1. The attributes that are considered to explain household decisions are essentially socio-demographic and vehicle characteristics. Model 1 is a MNL model with vehicle characteristics only, while in Model 2 we add to the previous specification socio-demographic characteristics. Model 3 is a MMNL model that includes an error component common to the new vehicle alternatives and that accounts for panel effect. Besides panel effect, Model 4 considers random coefficients for fuel economy; these coefficients are specific to the gasoline and electric vehicle alternatives and are assumed to be log-normally distributed. A log-normal distribution is applied because these response coefficients are expected to be positive. All MMNL models account for panel effect, the random coefficients being constant over choices made by the same respondent.

Table 4. 1 Vehicle Type Preference Estimation Results from Four Different Models

Variables [units]	Utility				Multinomial Logit		Mixed Logit (error component, panel effect)	Mixed Logit (random coeffs, error component, panel effect)
	Current	Gasoline	Hybrid	Electric	Model 1 Coefficient (t-stat)	Model 2 Coefficient (t-stat)	Model 3 Coefficient (t-stat)	Model 4 Coefficient (t-stat)
ASC_current	X				0.909 (1.9)	1.180 (2.2)	0.648 (0.9)*	0.785 (1.0)*
ASC_hybrid			X		1.170 (3.8)	0.976 (3.1)	0.950 (2.7)	0.613 (1.7)
ASC_electric				X	-0.983 (-0.8)*	-1.230 (-1.0)*	-3.350 (-2.6)	-3.290 (-2.2)
Young_ele				X	-	1.240 (5.5)	1.200 (4.9)	1.330 (4.7)
Young_hev			X		-	0.353 (2.2)	0.302 (1.6)*	0.397(2.0)
Educ_female_hev			X		-	0.215 (1.3)*	0.328 (1.8)	0.366 (1.8)
Educ_male_ele				X	-	0.476 (2.0)	0.471 (1.9)	0.686 (2.1)
Num_workers	X				-	-0.354 (-4.3)	-0.619 (-4.0)	-0.677 (-4.1)
Num_vehicles	X				-	0.180 (2.1)	0.346 (2.2)	0.380 (2.2)
Electricity_price[$]				X	-0.291 (-1.3)*	-0.424 (-1.8)	-0.263 (-1.1)*	-0.316 (-1.1)*
Gas_price[$]		X	X		-0.124 (-1.3)*	-0.134 (-1.4)*	-0.216 (-1.9)	-0.187 (-1.5)*
Fuel_economy_know (mean) [100MPG]	X	X	X		0.346 (0.9)*	0.522 (1.4)*	2.120 (3.9)	0.151 (0.3)*^
Fuel_economy_know (s.d.) [100MPG]	X	X	X		-	-	-	1.820 (4.7)^
Fuel_economy_unknown[100MPG]	X	X	X		-0.690 (5.5)	-0.729 (-5.9)	0.092 (0.5)*	0.223 (1.2)*
Ele_economy_know (mean) [100MPGE]				X	1.110 (1.7)	1.540 (2.3)	2.50 (3.5)	0.642 (1.4)*^
Ele_economy_know (s.d.) [100MPGE]				X	-	-	-	0.521 (2.1)^
Ele_economy_unknow[100MPGE]				X	1.260 (2.0)	1.37 (2.1)	1.980 (2.8)	2.140 (2.8)
Vehicle_size		X	X		0.068 (0.8)*	0.079 (1.0)*	0.167 (1.9)	0.227 (2.4)
Electric_vehicle_size				X	0.541 (2.4)	0.634 (2.8)	0.659 (2.9)	0.663 (2.6)
Recharging_Range[100 miles]				X	0.181 (0.5)*	0.206 (0.5)*	0.682 (1.6)*	0.838 (1.7)
Current_vehicle_price[$10,000]	X				-0.213 (-2.1)	-0.200 (-2.1)	-0.209 (-2.0)	-0.215 (-2.1)
Gasoline_vehicle_price[$10,000]		X			-0.388 (-4.7)	-0.391 (-4.7)	-0.345 (-3.7)	-0.386 (-4.0)
Hybrid_vehicle_price[$10,000]			X		-0.700 (-6.3)	-0.692 (-6.3)	-0.701 (-5.6)	-0.675 (-5.2)
Electric_vehicle_price[$10,000]				X	-0.628 (-3.1)	-0.685 (-3.4)	-0.666 (-3.0)	-0.750 (-3.1)
Dummy_short_run_gasol_veh_price		X			0.138 (2.7)	0.138 (2.7)	-0.025 (-0.4)*	-0.057 (-0.8)*
Dummy_short_run_hybrid_veh_price			X		0.069 (1.4)*	0.070 (1.4)*	-0.080 (-1.3)*	-0.109 (-1.6)*
Nest_effect_for_buying_group		X	X	X	-	-	2.440 (15.6)	2.570 (15.1)
Number of estimated parameters					18	24	25	27
Number of observations / individuals					3598 / 456	3598 / 456	3598 / 456	3598 / 456
Null log-likelihhod					-4987.887	-4987.887	-4987.887	-4987.887
Initial log-likelihood					-3003.521	-3052.559	-2713.730	-2592.452
Final log-likelihood					-2968.202	-2899.109	-2562.654	-2523.349
Rho-square					0.405	0.419	0.486	0.494
Adjusted Rho-square					0.401	0.414	0.481	0.489

Note: "" means the coefficient is not significant at significant level of 0.1;*

"^" means the coefficient is log-normally distributed by assumption.

For consistency, all four models are estimated on the same sample and assume "buying a gasoline vehicle" as the base alternative.

The coefficients of vehicle purchase price are negative and significant as expected. The absolute value of the price coefficient for the current vehicle is the lowest, followed by that of the new gasoline vehicle, and by those for new HEV and BEV, that are the highest. This pattern indicates that households are more sensitive to the purchase price of HEV and BEV, possibly because these vehicles are more expensive and their technology not fully known to the respondents. Besides, the dummy variables for the short-run purchase prices are only significant for the gasoline vehicle alternative in the MNL models, suggesting that: (1) no significant difference is observed between short-run and long-run purchase prices for HEV and BEV and (2) the random disturbances in the more flexible MMNL models may wipe out the distinction between how households value their vehicles in a short and long run.

The coefficients for vehicle size of both gasoline-powered vehicle and electricity-powered vehicle have positive sign as households prefer larger vehicles. In addition, households care more about the size of BEV than gasoline vehicle or HEV. The recharging range of BEV is positive as expected since a greater range allows for longer trips. Lower estimated values of recharging range are obtained from the MNL models, suggesting that the MNL models more conservatively predict how households value the recharging range. The change in the value of recharging range coefficients between MNL and MMNL models is consistent with Maness and Cirillo (Maness and Cirillo, 2012) and Bhat (Bhat, 2000).

For fuel economy measured by MPG (or MPGE for BEV), households are split into two groups based on their knowledge of current vehicle fuel economy. For households who know their vehicle MPG (or MPGE), the coefficients are positive as expected. For households choosing BEV, the coefficients of fuel economy are similar between households who know and do not know the fuel economy of their current vehicles. However, for households choosing gasoline-powered vehicles, fuel economy has little influence for households without knowledge of their current vehicle fuel economy. Additionally, lower estimation values are observed from the MNL model, suggesting that the MNL models are unable to capture household preferences for fuel economy. When fuel economy is treated as a log-normally distributed random variable in Model 4, the estimated mean is insignificant while the estimated standard deviation is significant, which may be attributed to a wide variation in preferences for fuel economy.

The coefficients of fuel price, including electricity price and gasoline price, are negative; increases in electricity and gasoline prices lead to decreases of households' preferences for electricity-powered and gasoline-powered vehicles respectively. The absolute values of the parameters for electricity price are higher than those of gasoline price, thus indicating that fuel price has a greater impact on the purchase of electricity-powered vehicles.

In Models 2, 3 and 4, coefficients of all household socio-demographic variables have reasonable sign, and the estimation results between the MNL and the MMNL models are consistent. Results show that young people prefer green vehicles with new technology, especially BEV. Women with a high education level (bachelor degree or higher) have a greater preference for HEV, while men with a high education level are more likely to choose BEV. Additionally, households with more workers or with fewer vehicles prefer to choose a new vehicle rather than to keep their current one.

4.2.2 Price Elasticity and Willingness to Pay (WTP)

Understanding how prices and other factors affect travel behavior is critical for transportation planning and for transportation demand management, including pricing reforms and pollution reduction (Litman, 2013). Transportation analysts measure responses to policies by elasticities, which is the percentage change in choice probabilities associated with one-percent change in the variable of interest. Direct elasticity is the change in the choice probability of choosing a particular alternative with respect to an observed variable of the utility of the same alternative; indirect or cross elasticity refers to the change in the choice probability when an observed variable relating to another alternative changes (Train, 2009).

In this Section, I use model estimations to calculate direct and indirect elasticities to vehicle price and to fuel price. In other terms, I want to understand how vehicle purchasing behavior changes when vehicle price or energy price increases. I present short-term and long-term elasticities; the first are calculated over a period of five years and the latter over the nine year time horizon for which observations were collected in our sample.

In MNL models, direct and cross elasticities are given in equations (4.4) and (4.5) respectively.

$$E_i x_{ni} = \frac{\partial P_{ni}}{\partial x_{ni}} \frac{x_{ni}}{P_{ni}} = \frac{\partial V_{ni}}{\partial x_{ni}} x_{ni}(1 - P_{ni}) \tag{4.4}$$

$$E_i x_{nj} = \frac{\partial P_{ni}}{\partial x_{nj}} \frac{x_{nj}}{P_{ni}} = \frac{\partial V_{ni}}{\partial x_{nj}} x_{nj} P_{nj} \tag{4.5}$$

Elasticities for the mixed logit model are calculated following Train (2009).

$$E_{ni} x_{nj}^{m} = -\frac{x_{nj}^{m}}{P_{ni}} \int \beta^{m} L_{ni}(\beta) L_{nj}(\beta) f(\beta) d\beta \tag{4.6}$$

where:

β^m is the element of the vector of coefficients for which we calculate the elasticity (i.e. vehicle price and fuel price);

L are the logit probabilities;

and P are the mixed logit probabilities.

Simulations are necessary to approximate the integral in equation (4.6) associated to the elasticity calculations; in this study we use 1000 draws from the random distributed coefficients β^m for which we calculate the elasticity.

Table 4.2 presents market elasticity with respect to vehicle price, which measures a change in the market share of gasoline, hybrid, or electric vehicles in response to a one-percent change in the purchase price of the corresponding vehicle. For example, the values for "Model 1" in Table 4.2 suggest that (1) over the long-run and short-run one-percent increase in the purchase price of a new gasoline vehicle will decrease the market share of gasoline vehicles by 0.94% and 0.60%, respectively; (2) in the long run one-percent increase in the purchase price of a new HEV or BEV is expected to reduce the market share of the corresponding vehicle by 1.79% or 1.46%, respectively; (3) for HEV and BEV, results show little difference of vehicle own price elasticities between the long-run and the short-run; and (4) Results from the MNL models suggest that the long-run market elasticity for gasoline vehicle is greater than the short-run elasticity by a factor of 1.5, as households are reluctant to switch their preferred vehicle in shorter time periods.

We can observe that elasticities to the market price are -0.60 in the short run and in the range of -0.70 to -0.95 in the long run for gasoline vehicle; -1.45 to -1.80 for HEV; and -1.30 to -1.60 for BEV. The values calculated indicate that gasoline vehicles are price inelastic while HEV and BEV are price elastic. The results obtained for gasoline vehicle are consistent with results estimated by Lave and Train (Lave and Train, 1979), Levinsohn (Levinsohn, 1988), and McCarthy (McCarthy, 1996). However, these studies are quite dated and do not provide elasticities for advanced technology vehicles. By comparing Models 1-2 and Models 3-4, we can also observe that MMNL models estimate more moderate market elasticities with respect to vehicle purchasing price.

Table 4. 2 Market Elasticity with respect to Vehicle Price

Market Elasticity		Model 1	Model 2	Model 3	Model 4
Vehicle Price	GasV (long-run)	-0.94	-0.95	-0.69	-0.73
	GasV (short-run)	-0.60	-0.60	-	-
	HEV	-1.79	-1.77	-1.44	-1.46
	BEV	-1.46	-1.60	-1.43	-1.32

Similarly, Table 4.3 reports market cross-elasticity with respect to vehicle price. This type of elasticity measures how one-percent increase in the purchase price of one type of vehicle affects the market share of other types of vehicles. The market cross-elasticities for different vehicle types are identical in Models 1-3 because of the independence of irrelevant alternative (IIA) property holds for the “buying group”. The MNL models (Model 1-2) generally underestimate market cross-elasticities, while the MMNL models (Model 3-4) provide cross-elasticities that are in the range from 0.20 to 0.60. Market cross-elasticities estimated with Model 4 indicate that increasing the purchase price of gasoline vehicle will induce more households to turn to HEV rather than BEV. Similarly, increasing the purchase price of HEV will induce more households to choose gasoline vehicles rather than BEV.

Table 4. 3 Market Cross-Elasticity with respect to Vehicle Price

Market Cross Elasticity			Model 1	Model 2	Model 3	Model 4
Vehicle Price	GasV Price	HEV	0.09	0.09	0.23	0.27
		BEV	0.09	0.09	0.23	0.20
	HEV Price	GasV	0.22	0.22	0.57	0.47
		BEV	0.22	0.22	0.57	0.38
	BEV Price	GasV	0.07	0.08	0.20	0.25
		HEV	0.07	0.08	0.20	0.25

Vehicle market elasticities with respect to fuel price are reported in Table 4.4. This type of elasticity measures how one-percent increase in gasoline price or electricity price affects the market share of either gasoline-powered or electricity-powered vehicles. When comparing results from the four models, the estimated elasticities from the MMNL models are similar or higher than the values obtained with MNL. In Model 3, one-percent increase in gasoline price produces respectively 0.58% and 0.54% decrease in the market shares of gasoline vehicle and HEV, while the market share of BEV will increase by 0.21%. On the other hand, increasing electricity price by one-percent will decrease by 1.14% the market share of BEV, while the market shares of gasoline-powered vehicles will increase by 0.17% only. Results show that the market elasticities with respect to the electricity price are much greater than those calculated with respect to gasoline price, which indicates that households are more sensitive to electricity price when buying BEV. Overall, compared with the literature (Johansson and Schipper 1997, p. 209; Goodwin 1992;

Goodwin et al., 2011), our results provide higher estimations of the market elasticity with respect to fuel price. These results may be due to different travel patterns, car ownership, vehicle fees, and fuel prices in our geographical area (Maryland) (Giuliano and Dargay, 2006; Litman, 2013).

Table 4. 4 Market Elasticity with respect to Fuel Price

Market Elasticity		Model 1	Model 2	Model 3	Model 4
Gasoline Price	GasV	-0.49	-0.53	-0.58	-0.58
	HEV	-0.48	-0.51	-0.54	-0.60
	BEV	0.05	0.06	0.21	0.19
Electricity Price	GasV	0.07	0.11	0.17	0.22
	HEV	0.07	0.11	0.17	0.22
	BEV	-1.38	-2.01	-1.14	-1.12

Further analysis presents households' valuation of vehicle attributes such as fuel economy, range, and size. **Error! Reference source not found.**4.5 summarizes t he willingness to pay (WTP) values estimated; in general, it can be observed that MNL models underestimate the WTP for vehicle fuel economy and size. Model 3 shows that WTP to increase one MPG (or MPGE) for gasoline, hybrid electric, and battery electric vehicles are $614, $302, and $375, respectively. These results indicate that households are willing to pay more to increase the fuel economy of gasoline vehicles as they are in general less fuel-efficient. Additionally, results show that the WTP to increase vehicle size by one level for gasoline, hybrid electric, and battery electric vehicles are $4841, $2382, and $9895, respectively. The WTP to increase the size of electric vehicle is the highest, as probably the main concern of potential buyers of electric vehicles is about the small size of this type of vehicle in the current market. The WTP to increase one mile in the recharging range of battery electric vehicles is between $102 and $112 as predicted by the MMNL model; it should be noted that this variable is not significant in the MNL models.

Table 4. 5 Willingness to Pay for Vehicle Fuel Economy, Range, and Size

Attributes	Vehicle Type	Model 1	Model 2	Model 3	Model 4
Fuel Economy (dollars to increase 1 mpg)	GasV	89	134	614	-
	HEV	49	75	302	-
	BEV	177	225	375	253
Range (dollars)	BEV	-	-	102	112
Size (dollars to increase one size level)	GasV	1753	2020	4841	5881
	HEV	971	1142	2382	3363
	BEV	8615	9255	9895	8840

The results provide important implications for the understanding of vehicle preferences and for the definition of WTP for different vehicle characteristics. These can be summarized as follows: (1) the market share of advanced vehicle technology (hybrid and electric vehicle) is affected by their market price; (2) the propensity to buy a new vehicle depends on fuel price; (3) the WPT to increase the fuel economy of gasoline vehicles is double with respect to the WPT to increase the efficiency of hybrid and electric vehicles; (4) perspective buyers are concerned about the size of electric vehicle and the WTP to increase their size is relatively high.

From a policy perspective, moderate prices for new technology vehicles or economic incentives will accelerate their diffusion in the market place. Low prices for fuel and electricity will tend to decrease the interest in new and more efficient vehicles. Finally, vehicle size is a very important factor for the US market; potential buyers are willing to pay a high price in order to own larger electric vehicles.

4.3 Chapter Conclusions

The proposed MMNL model framework has a flexible structure that can approximate any random utility model (McFadden and Train, 2000). It obviates the limitations of standard MNL by allowing for random taste variation, unrestricted substitution pattern, correlation between unobserved factors over time, and panel effect on observations of the same individual.

Specifically, under some derivations of the MMNL model, the values of random coefficients represent different tastes of decision makers. The most popular distributions of random coefficients used in applications of MMNL models are normal, log-normal, uniform, triangular and gamma distributions.

The unobserved random portion of mixed logit utility can be correlated over alternatives depending on the specification of observed variables associated with random coefficients. In the standard logit model, all coefficients are fixed and the covariance matrix of error components is assumed to be an identity matrix, so that there is no correlation in utility between alternatives. The lack of correlation gives rise to the IIA (independence of irrelevant alternatives) property and the restrictive substitution pattern among alternatives. However, the MMNL structure overcomes the IIA property and provides sufficiently realistic substitution patterns by accounting for correlations between alternatives. It is important to note that the mixing distribution, whether motivated by random parameters or by error components, captures variance and correlations in unobserved factors (Train, 2009).

Besides, the specification of MMNL model is easily generalized to allow for repeated choices by each sampled decision maker. The simplest specification treats the coefficients that enter utility as varying over people but being constant over choice situations for each person (Train, 2009).

The model framework in this study is appropriate to evaluate impact of vehicle-related policies on household future vehicle preferences, and to calculate price elasticities and WTPs for improving car characteristics. Additionally, the model can be further extended into a dynamic choice model which forecasts vehicle market share over a short, medium, and long term. It provides policy makers a valuable reference for medium to long term urban planning.

However, the proposed MMNL model framework has some drawbacks. First, the estimation process needs simulation because the log-likelihood function of mixed logit does not have a closed form. The estimation time will exponentially increase as the number of random coefficients increases, exhibiting computational complexity of the model. Second, although past and future exogenous variables can be added to the utility in a given time period to represent lagged response and anticipatory behavior, the model formulation is static and without a notice about market evolution.

Chapter 5: Methodology Part 2: Generalized Dynamic Discrete Choice Model

5.1 Introduction

To overcome the static nature of mixed logit models, this Chapter formalizes a general dynamic discrete choice framework to capture the optimal time of vehicle purchase and household's vehicle type choice in a dynamic market. In the framework, forward-looking agents optimize their utility over time; two options are available at each time: keeping the current vehicle or buying a new vehicle among the options available in the market. Different model forms are proposed to consider the purchase pattern of different durable goods in the market: the regenerative optimal stopping formulation allows agents to return to the market after a purchase is made, while the regular optimal stopping formulation guarantee agents to be out-of-market after a change in status. Moreover, the model accounts for dynamically evolving market conditions by a stochastic diffusion process that captures time-series correlations between market indicators.

The proposed modeling framework has been applied to estimate green vehicle adoption rate for households living in Maryland. The estimation results have been applied to test different policy scenarios, including changes in fuel price, vehicle purchase price, and improvement of vehicle characteristics. These policies have a high impact on the adoption of electric cars and on their diffusion in the marketplace.

The following sections present the formulation of the dynamic modeling framework, the specification of the dynamic attributes, the estimation strategies for solving the underlying Maximum Likelihood problem, and the application to forecast time-dependent green vehicle adoption in Maryland.

<u>5.2 Generalized Consumer Stopping Problem</u>

Consumers are indexed by $i = 1, 2, \dots, M$. Time is assumed to be discrete and indexed by $t = 0, 1, \dots, T$. In each time period t, consumer i faces two options if he or she is in the market: (a) to buy one of the products $j \in \mathcal{J}_t = \{1, 2, \dots, J_t\}$ available in the market at time t and obtain a terminal payoff u_{ijt}; or (b) to postpone the purchase and obtain a one-period utility payoff $c_{it}(x_{it}, q_{it};\ \theta_i, \alpha_i)$, where x_{it} is a vector of social demographic attributes for consumer i at time t, q_{it} is a vector of characteristics of consumer's owned products, θ_i and α_i are vectors of parameters corresponding to x_{it} and q_{it}.

If consumer i buys product $j \in \mathcal{J}_t$, he or she obtains a terminal payoff formulated as follows:

$$u_{ijt} = f\left(x_{it}, z_{jt}, y_{jt};\ \theta_i, \gamma_i, \beta_i\right) + \varepsilon_{ijt} \tag{5.1}$$

where:

x_{it}, θ_i are $(1 \times Q)$ vectors defined as above;

z_{jt} is a $(1 \times K)$ vector of static or time-dependent characteristics for product j in the market in time period t;

γ_i represents a vector of parameters related to z_{jt};

y_{jt} is a $(1 \times H)$ random vector of dynamic attributes for product j in the market in time period t, such as energy price, vehicle price, and environmental incentives which describe industry/market evolution;

β_i represents a vector of parameters related to y_{jt};

ε_{ijt} is an individual-specific random utility component, which follows a GEV distribution. The random utility components are assumed to be i.i.d. over consumers, products, and time periods.

We assume consumer preferences on characteristics of products are homogenous, then parameters $\theta_i, \gamma_i, \beta_i$ reduce to θ, γ, β respectively. Specifically, if consumer i decides to buy a vehicle at time t instead of postponing, vehicle type choice is estimated by a MNL model with an error component following T1EV distribution. Correspondingly, for consumer i, $v_{it} = max_{j\in\mathcal{J}_t} u_{ijt}$ follows T1EV distribution with cumulative distribution (F_v) and probability density functions (f_v) as follows:

$$F_v(u;\, r_{it}) = exp\left(-e^{-(u-r_{it})}\right) \tag{5.2}$$

$$f_v(u;\, r_{it}) = e^{r_{it}} exp\left(-e^{-(u-r_{it})} - u\right) \tag{5.3}$$

where r_{it} is the mode of this distribution, formulated as:

$$r_{it} = lnG\left(exp\left(f\left(x_{it}, z_{jt}, y_{jt};\, \theta, \gamma, \beta\right)\right)\right) \tag{5.4}$$

where $G\left(f\left(x_{it}, z_{jt}, y_{jt};\, \theta, \gamma, \beta\right)\right) = \sum_{j\in\mathcal{J}_t} f\left(x_{it}, z_{jt}, y_{jt};\, \theta, \gamma, \beta\right)$ for MNL model with a Gumbel-distributed error component. Alternatively, r_j can be represented as follows:

$$r_{it} = ln\sum_{j\in\mathcal{J}_t} exp(f\left(x_{it}, z_{jt}, y_{jt};\, \theta, \gamma, \beta\right)) = E_t\left[max_{j\in\mathcal{J}_t}(u_{ijt})\right] = E_t[v_{it}] \tag{5.5}$$

where $E_t(*)$ is the expectation given vehicle set $\mathcal{J}_t$ in the market. We consider r_{it} because it is a scalar-valued sufficient statistic for the distribution of future payoffs

(Melnikov, 2013), and it contains the information available to the consumer i at time t.

In each time period, and based on the available information, the consumer is called to decide when to buy a vehicle and which type of vehicle to buy. The frameworks models jointly the decisions of whether to postpone the purchase until the next period or to buy a new vehicle; in the latter case the consumers also chooses product j_t^* from $\mathcal{J}_t$ that maximizes his or her utility of purchase (u_{jt}). We assume consumers are able to look forward and maximize their expected inter-temporal payoffs. Denoting the time period the consumer decides to buy a product by τ, the consumer's optimization problem can be formulated as:

$$D_{it}(v_{it}, r_{it}, c_{it}) = max_{\tau \geq t}\{\sum_{k=t}^{\tau-1} \beta^{k-t} c_{ik} + \beta^{\tau-t} E_{\tau}[v_{i\tau}|r_{it}]\} \tag{5.5}$$

where D_{it} represents the decision process of consumer i at time t; $\beta \in [0,1]$ is a common discount factor; and $E_{\tau}[* |r_{it}]$ denotes a conditional expectation given the information set available for consumer i at time t.

5.3 Recursive Decision Process

Given the definition of v_{it}, r_{it}, c_{it}, an alternative way to formulate the consumer's decision process recursively is as follows:

$$D_{it}(v_{it}, r_{it}, c_{it}) = \max\{v_{it}, c_{it} + \beta E_{t+1}[D_{it+1}(v_{it+1}, r_{it+1}, c_{it+1})|\, r_{it}]\} \tag{5.7}$$

If consumer i postpones his or her purchase at time t, the reservation utility can be written as:

$$W_{it}(r_{it}) = c_{it} + \beta E_{t+1}[D_{it+1}(v_{it+1}, r_{it+1}, c_{it+1})|\, r_{it}] \tag{5.8}$$

Therefore, the recursive formulation of the consumer decision process can be simplified as:

$$D_{it}(v_{it}, r_{it}, c_{it}) = \max\{v_{it}, W_{it}(r_{it})\} \tag{5.9}$$

5.4 Formulation of Choice Probability

The consumer decision D_{it} remains random because the random component ε_{ijt} exists in the utility function. We assume ε_{ijt} randomly take a specific realization for each consumer i, which indicates ε_{ijt} is simply the unobserved part of the utility function and is independent of dynamic attributes. Based on utility maximization, consumer i will make a purchase at time t when $v_{it} > W_{it}(r_{it})$. Otherwise, he or she will postpone the purchase until the next period. For a randomly choosing consumer i, the probability of postponing the purchase at time t can be written as:

$$\pi_{i0t}(r_{it}) = P\big(v_{it} \leq W_{it}(r_{it})\big) = F_v(W_{it}(r_{it});\, r_{it}) = exp\big(-e^{-(W_{it}(r_{it})-r_{it})}\big) \tag{5.10}$$

Consequently, the probability that consumer i buys a product at time t is $h_{it}(r_{it}) = 1 - \pi_{i0t}(r_{it})$. And the probability of the consumer purchasing product j at time t is the product of $h_{it}(r_{it})$ and the conditional probability of choosing $j \in \mathcal{J}_t$ given consumer i makes a purchase.

$$\begin{aligned} \pi_{ijt}(r_{it}) &= P\big([v_{it} > W_{it}(r_{it})] \cap [v_{it} = u_{ijt}]\big) \quad (5.11) \\ &= P\big(v_{it} > W_{it}(r_{it})\big) \cdot P\big(v_{it} = u_{ijt} \geq u_{ikt}, \forall k \in \mathcal{J}_t \; and \; k \neq j\big) \\ &= h_{it}(r_{it}) \cdot P\big(u_{ijt} \geq u_{ikt}, \forall k \in \mathcal{J}_t \; and \; k \neq j\big) \\ &= h_{it}(r_{it}) \cdot p_{ijt} \end{aligned}$$

where p_{ijt} represents the conditional probability of buying product j given that consumer i makes a purchase at time t. Obviously, if the consumer makes a purchase, $\sum_{j\in\mathcal{J}_t} p_{ijt} = 1$; otherwise, $\sum_{j\in\mathcal{J}_t} p_{ijt} = 0$.

It should be noted that the calculation of the expected utility in the future is based on a finite horizon scenario tree. At each time period, it assumes that a respondent can anticipate possible alternative characteristics over a limited number of future time periods. For example, if three future time periods are considered, the respondent is assumed to have no knowledge of the 4th time period starting from time 0, and the expected utility from the 4th time period is assumed to be zero. For more details, we refer to Cirillo et al. (2015).

In the following sections, the framework described above will be generalized to relax some of the assumptions and to accommodate different behavioral processes. In particular, different model specifications are formulated for one-time purchases, repeated purchases, and industry evolution based on one dynamic attribute and multiple correlated dynamic attributes.

<u>*5.5 Transition Probability Matrix*</u>

- *Scenario 1: One-Time Purchase*

In this case, let's assume that consumers can only make one purchase in the considered time horizon and will leave the market immediately after their first purchase. We will use the probability transition matrix of vehicle ownership as an example to describe this scenario in detail. Denoting vehicle ownership status of consumer i at time t by $S_{it} \in \{0, 1, 2, 3, 4\}$, where $S_{it} = 0$ if the consumer does not purchase any vehicle, $S_{it} = 1$ if buys a gasoline vehicle, $S_{it} = 2$ if buys a hybrid vehicle, $S_{it} = 3$ if buys an electric vehicle, and $S_{it} = 4$ if the consumer is out-of-market. The transition between the states is governed by a Markov probability matrix $H_1: \{0, 1, 2, 3, 4\} \rightarrow \{0, 1, 2, 3, 4\}$ specified as:

$$H_1(r_{it}) = \begin{bmatrix} \pi_{i0t}(r_{it}) & \pi_{i1t}(r_{it}) & \pi_{i2t}(r_{it}) & \pi_{i3t}(r_{it}) & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.12}$$

If the consumer does not purchase any vehicle, $S_{it} = 0$, he or she has a probability of $\pi_{i1t}(r_{it})$, $\pi_{i2t}(r_{it})$, or $\pi_{i3t}(r_{it})$ to purchase a gasoline, hybrid, or electric vehicle, and a probability of $\pi_{i0t}(r_{it})$ to postpone the purchase to the next time period. If the consumer makes a purchase, $S_{it} = 1, 2, or\ 3$, he or she will be out-of-market where $S_{it} = 4$. Intuitively, state 4 is an absorbing state, which indicates that once the consumer is out-of-market, he or she will never return.

- *Case 2: Multiple Purchases*

Notice that the model in scenario 1 can be extended to incorporate repeated purchases, that is, the consumer will stay in market or return to market after buying a product. More specifically, repeated purchases can be modeled by solving a regenerative optimal stopping problem. When the consumer reaches a terminal state, the decision process is restarted and attributes describing characteristics of the consumer's owned product are reinitialized. Note that "regenerative" takes its statistical meaning (Ross, 1997), so it is sufficient to discuss the sequence of choices from one regeneration time to the next. Taking vehicle ownership problem as an example, if consumer i always stays in market, the transition of consumer states can be represented by a Markov probability matrix $H_2: \{0, 1, 2, 3\} \to \{0, 1, 2, 3\}$:

$$H_2(r_{it}) = \begin{bmatrix} \pi_{i0t}(r_{it}) & \pi_{i1t}(r_{it}) & \pi_{i2t}(r_{it}) & \pi_{i3t}(r_{it}) \\ q_{10} & (1 - q_{10})p_{i1t} & (1 - q_{10})p_{i2t} & (1 - q_{10})p_{i3t} \\ q_{20} & (1 - q_{20})p_{i1t} & (1 - q_{20})p_{i2t} & (1 - q_{20})p_{i3t} \\ q_{30} & (1 - q_{30})p_{i1t} & (1 - q_{30})p_{i2t} & (1 - q_{30})p_{i3t} \end{bmatrix} \tag{5.13}$$

where q_{j0} represents the transition probability from state j to state 0. In this case, whether a consumer makes a purchase or not, he or she will have a chance to buy or to postpone. In the diverse market of durable products, a consumer usually does not consider repurchase immediately after owning a new product. Therefore, in a more comprehensive framework, when a consumer buys a product, he or she will be out-of-market for a certain time period and then return to market.

5.6 Industry Evolution

- *Case 1: An Autoregressive Process for A Single Dynamic Attribute*

As defined above, y_{jt} represents the evolution of product j's characteristics in the market or environmental incentives offered by producers or policy makers. Given the dimensionality of the product characteristic space and the diversity of products in a typical market, it is computationally infeasible to generate y_{jt} directly (Melnikov, 2013). Therefore, it assumes that a reduced set of state variables can adequately describe the state of market at time t. In this case, a stochastic diffusion process is used to model the change of a single dynamic attribute to mimic the evolving market.

$$y_{j,t+1} = \mu(y_{jt}) + \sigma(y_{jt})\upsilon_{j,t+1} \qquad (5.17)$$

where $\upsilon_{j,t+1}$ are i.i.d. and follow standard normal distributions; $\mu(y_{jt})$ and $\sigma(y_{jt})$ are continuous and almost everywhere differentiable; $0 < \sigma(y_{jt}) < \infty$; $\mu(y_{jt}) > y_{jt}$; and $\lim_{n\to\infty} \beta^n \mu^n(y) < \infty$ where $0 \le \beta < 1, \mu^0(y) = \mu(y), \mu^n(y) = \mu(\mu^{n-1}(y))$.

Notice that the above formulation is quite flexible and encompasses many specifications used to model economic growth and technological change. Considering the dynamic pattern in vehicle ownership problem, I use a stable autoregressive process of order one (AR(1)), a specific type of diffusion process, to generate state variables such as energy price and vehicle price. The AR(1) specifies that the dynamic variable depends linearly on its own previous values and a stochastic term. The formulation can be expressed as follows:

$$y_{j,t+1} = \delta_j + \eta_j y_{jt} + \sigma v_{j,t+1}, \ \left|\eta_j\right| < 1 \tag{5.15}$$

where δ_j and η_j are parameters to be estimated, σ is the standard deviation of the stochastic term.

- *Case 2: Vector Autoregressive Process for Multiple Dynamic Attributes*

The AR(1) process can be extended to a vector autoregressive process of order one (VAR(1)) to model multiple correlated dynamic variables to mimic market evolution. The VAR(1) is a generalized form of AR(1). It captures the linear interdependencies among multiple time-series variables by building the evolution of one variable on its own lags and the lags of the other variables. In the case of two correlated dynamic variables, the process can be specified as follows:

$$y_{1,t+1} = \delta_1 + \eta_{11} y_{1,t} + \eta_{12} y_{2,t} + \sigma_1 v_{1,t+1} \tag{5.16}$$

$$y_{2,t+1} = \delta_2 + \eta_{21} y_{1,t} + \eta_{22} y_{2,t} + \sigma_2 v_{2,t+1} \tag{5.17}$$

where δ_1, δ_2, η_{11}, η_{12}, η_{21}, η_{22} are parameters to be estimated; σ_1 and σ_2 are the standard deviations of the stochastic parts. Alternatively, the process can be written in a matrix form:

$$y_{t+1} = B + Ay_t + \epsilon_{t+1} \tag{5.18}$$

where $B = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$ and $A = \begin{bmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{bmatrix}$ are parameters to be estimated; ϵ_{t+1} is the stochastic term which follows multivariate normal distribution with mean $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and variance $\begin{bmatrix} {\sigma_1}^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & {\sigma_2}^2 \end{bmatrix}$.

Different scenarios of model structures are presented in Section 5.4 and 5.5 to identify diverse purchase behaviors and markets. The combinations of these scenarios can also be used to model more complex market situation.

5.7 Optimization Process

The proposed model is estimated by a maximum likelihood technique. The estimated parameters of $\theta, \alpha, \gamma, \beta$ are obtained by maximizing the likelihood of purchase decisions over all the consumers and time periods; the final likelihood function is defined as follows:

$$L(\theta, \alpha, \gamma, \beta) = \prod_{i=1}^{M} \prod_{t=1}^{H} P_{it}[D_{it}(v_{it}, r_{it}, c_{it}) | \theta, \alpha, \gamma, \beta] \tag{5.19}$$

where H defines the number of time periods; and P_{it} represents the probability that household i makes a decision D_{it} at time t.

To obtain the above likelihood function, we should first calculate the probability of “not to buy” ($\pi_{i0t}(r_{it})$) and the probability of “buy” and choosing product j ($\pi_{ijt}(r_{it})$). The key point for the whole process is to calculate the expected future utility by a finite horizon scenario tree, which is a commonly used technique in DP and stochastic programming (Bertsekas, 2005; Shapiro et al., 2009). At each time period, a respondent is assumed to have a perspective about future scenarios in the short-term horizon, which is characterized by the changing attributes of alternatives and evolving market conditions.

As a simple illustration, let's suppose that, starting from the generic time period t, the respondent faces two possible alternatives – buy a car of certain type and not to buy. At time $t+1$, each of the two scenarios from time t generates another two scenarios – to buy and not to buy, resulting a total of four scenarios. Iteratively, the decision process is formulated by means of a scenario tree in Figure 5.1. In this example, the expected future utility at time t is rewritten as $E[D_t]$ for simplification purpose.

At time 0, a respondent has two alternatives – either not to buy or to buy the car with the highest utility. If not to buy, the respondent will obtain a reservation utility of $W(y_0) = c_0 + E[D_1]$; otherwise, he or she will obtain the highest utility of purchase v_0. The decision of this respondent depends on which of these two utilities has a higher value. In order to calculate the reservation utility $W(y_0)$, the expected utility for the next time period $E[D_1]$ must be calculated. This expected utility for time 1 should be the expected maximum utility from the two alternatives (to buy or not to buy) at time 1; expressed as $E[D_1] = E\{\max\{v_1, c_1 + \beta E[D_2]\}\}$. Think recursively, we can also calculate $E[D_2] = E\{\max\{v_2, c_2 + \beta E[D_3]\}\}$, where $E[D_3]$ is assumed to be zero based on our assumption that the maximum forward-looking periods is 3. Specifically, starting from time 0, the expected utilities from the third or later time periods is assumed to be zero. If we start from time 1, the expected utilities from time 4 or later will be zero. The same is true for all other time periods where we collect the observations.

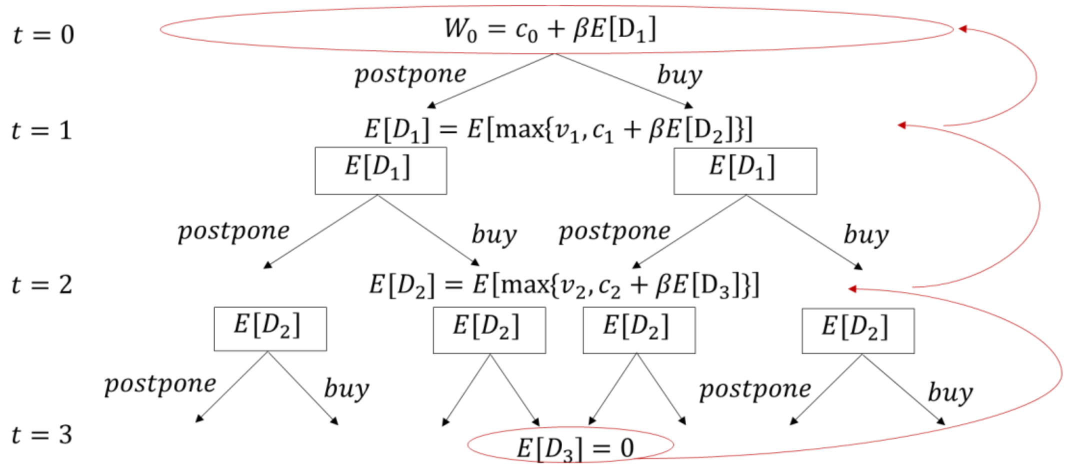


Figure 5. 1 An example of finite-horizon scenario tree

5.8 Application: Green Vehicle Adoption in Maryland

5.8.1 Model Estimation Results

Five scenarios of models have been estimated to analyze households' preferences on new vehicle types and their characteristics in Maryland. The first model is a MNL, estimated for comparison purpose. The second one is the proposed dynamic structure with repeated purchases and no market evolution. The third model is the dynamic structure with repeated purchases and evolving gasoline price generated using an AR(1) process. The development of the forth model is based on the third one. It accounts for market evolution through the generation of gasoline and electricity prices with the VAR(1) process. The market evolution of the last model is the same as the forth model; however, we consider one-time purchases, which means households will be out-of-market immediately after their first purchase. The five estimation results are presented and compared in Table 5.1.

In each time period, respondents either keep their current vehicle or choose from three alternatives: a new gasoline vehicle, a new hybrid vehicle, and a new electric vehicle. Out of 500 respondents participated in the survey, 456 of them provided complete information and were included in the final sample for estimation. More importantly, although respondents are supposed to express their decisions for eighteen time periods over nine years, only the decisions from the first fifteen time periods are effective for the estimation, and decisions of the rest three are sacrificed for calculating the expected utility of the future. It is found that the most appropriate look-forward time period is 3 by comparing the likelihood ratio index, the sign and t-value of estimated coefficients between models with look-forward time period equaling to 1, 2, 3, 4 and 5. In this case, the sample for estimation contains 3598 observations. The variables include vehicle price, size, fuel economy, refueling range, gasoline and electricity prices, number of vehicles held by a household, number of workers, and other social-demographic attributes.

Table 5. 1 Model Estimation Results: Consumer’s Preference on Vehicle Type

Attributes [units]	current	Gasoline	Hybrid	Electric	MNL Estimate (t-stat)	Repeated Purchases (Dyn_R) Estimate (t-stat)	Repeated Purchases (AR_R) Estimate (t-stat)	Repeated Purchases and (VAR_R) Estimate (t-stat)	One-time Purchase (VAR_S) Estimate (t-stat)
Vehicles [number]	X				0.094 (2.0)	0.185 (15.8)	0.159 (15.5)	0.157 (13.8)	0.214 (11.9)
Workers [number]	X				-0.101 (-2.5)	-0.027 (-2.2)	-0.035 (-3.1)	-0.036 (-1.9)	-0.020 (-1.0)*
VehPrice.gas [$10,000]		X			-0.582 (-6.3)	-0.492 (-8.1)	-0.394 (-9.1)	-0.372 (-2.9)	-0.099 (-3.2)
size.gas [small, medium, large]		X			0.194 (1.5)*	0.905 (10.9)	0.344 (5.1)	0.333 (1.5)*	0.987 (6.0)
mpg_known.gas [100mpg]		X			-1.151 (-1.3)*^	10.111 (14.2)	7.682 (7.4)	7.394 (5.8)	14.570 (15.9)
mpg_unknown.gas [100mpg]		X			-1.619 (-1.8)^	8.631 (13.1)	0.485 (1.3)*	0.590 (1.2)*	13.782 (14.2)
GasPrice.gas [$1]		X			-0.270 (-3.4)	0.547 (7.5)^	-0.091 (-4.9)	-0.127 (-2.0)	-0.443 (-24.0)
ASC.hev			X		-2.263 (-4.6)	2.148 (4.5)	-1.053 (-9.3)	-0.927 (-1.0)*	2.071 (4.4)
D_Young.hev [1/0]			X		0.178 (1.6)*	0.489 (4.0)	0.377 (2.4)	0.314 (1.8)	0.524 (9.8)
D_EducFemale.hev [1/0]			X		0.218 (1.9)	0.434 (3.9)	0.158 (1.4)*	0.181 (1.2)*	0.169 (1.0)*
VehPrice.hev [$10,000]			X		-0.464 (-4.6)	-0.592 (-6.9)	-0.500 (-8.5)	-0.536 (-2.5)	-0.535 (-6.9)
size.hev [small, medium, large]			X		0.158 (1.4)*	0.706 (8.3)	0.408 (7.6)	0.382 (2.3)	0.372 (2.3)
mpg_known.hev [100mpg]			X		1.691 (2.5)	8.569 (14.3)	8.297 (6.8)	7.955 (4.9)	5.445 (3.3)
mpg_unknown.hev [100mpg]			X		0.803 (1.2)*	5.630 (10.1)	2.090 (11.6)	1.965 (1.7)	2.523 (1.8)

ASC.bev	X	-5.684 (-5.0)	-3.198 (-5.0)	-3.088 (-3.0)	-2.013 (-1.9)	4.385 (1.7)
D_Young.bev [1/0]	X	1.059 (6.3)	1.651 (9.9)	1.478 (8.5)	1.496 (8.2)	1.615 (5.7)
D_EducMale.bev [1/0]	X	0.396 (2.1)	0.739 (4.6)	0.497 (2.9)	0.436 (2.4)	0.350 (1.6)*
VehPrice.bev [$10,000]	X	-0.726 (-3.6)	-0.794 (-5.1)	-0.573 (-3.1)	-0.637 (-3.3)	-1.181 (-5.8)
size.bev [small, medium, large]	X	0.714 (3.3)	0.752 (4.6)	0.769 (3.8)	0.578 (2.7)	0.205 (0.4)*
range.bev [100miles]	X	0.544 (1.3)*	2.010 (5.6)	0.960 (2.5)	0.880 (2.1)	0.708 (1.1)*
mpg_known.bev [100mpg]	X	2.494 (3.7)	4.838 (9.8)	3.998 (6.5)	3.120 (4.4)	1.593 (0.7)*
mpg_unknown.bev [100mpg]	X	2.516 (3.8)	4.060 (8.4)	2.003 (3.2)	1.293 (1.9)	0.849 (0.4)*
ElePrice.bev [$1]	X	-0.107 (-0.6)*	0.123 (2.1)^	-0.327 (-2.0)	-0.321 (-1.8)	-0.557 (-2.2)
LL(0)		-5621.471	-8201.659	-8201.659	-8201.659	-5621.471
LL($\hat{\beta}$)		-3557.327	-2779.839	-2808.669	-2805.058	-1423.27
Likelihood ratio index		0.367	0.661	0.658	0.658	0.747

Note: "" means the coefficient is not significant at significant level of 0.1;*

"^" means the sign of the coefficient is not as expected.

All models are estimated on the same data set and with the same specification for consistency; the estimation results are shown in Table 5.1.

- *MNL Model Results*

The estimation of MNL model is for comparison purposes: results are reported in the column "MNL". The model is static; the panel data is treated as a cross-sectional data. It can be observed that the estimated coefficients have reasonable signs except for fuel economy of gasoline vehicle. Most coefficients are statistically significant except for the size and fuel economy of gasoline and hybrid vehicles, the range of electric vehicle, price of electricity, and the indicator of young people for the hybrid vehicle alternative. The coefficient related to the number of vehicles held by a household is positive, indicating that households with more cars are more likely to keep their current vehicles and to postpone the purchase of new vehicles. The coefficient associated with the number of workers is negative, which suggests that households with more workers tend to purchase new vehicles. As expected, the purchasing price coefficients are negative for all types of vehicles, and their magnitudes suggest that households are more sensitive to the price of electric vehicles, followed by gasoline vehicles, and least sensitive to the price of hybrid vehicles. Size coefficients are positive for all vehicle types attesting that households prefer large cars. On the other hand, the coefficients of fuel economy for hybrid and electric vehicles are positive, indicating that households prefer higher fuel efficiency. With reference to the operating cost, the magnitude of the estimated coefficients show that households are more sensitive to gasoline price. Besides, we can observe that female with a bachelor or higher degree are more likely to purchase hybrid vehicle, while young people or male with a bachelor or higher degree tend to buy electric vehicle.

- *Dynamic Model Results without Market Evolution*

The dynamic structure captures the sequence of decisions made by a household over time; however, no market evaluation is considered and all attributes are static. The model specification remains the one adopted for the MNL case, and the estimation results are presented in the column "Dyn_R" of Table 1. We can observe that all coefficients are statistically significant. However, the sign of gasoline price and electricity price is incorrect. As already stated, gasoline price and electricity price are static variables generated from the scenarios presented in the SP survey. But the generated values from the SP survey might not necessarily reflect the values of fuel prices anticipated by the respondents. Compared to the MNL model results, the magnitude of coefficients related to the number of vehicles and number of workers indicates that households' purchase decisions are more sensitive to the number of vehicles and less sensitive to the number of workers in this dynamic structure. Different from MNL model results, the magnitudes of vehicle purchasing price coefficients suggest that households are more sensitive to electric vehicle price, then to the price of hybrid vehicles, and least sensitive to gasoline vehicle prices. This pattern seems to be more reasonable because households usually are reluctant to buy vehicles with new technologies, and a lower vehicle price will attract more buyers. The remaining coefficients of the dynamic model suggest that households prefer larger vehicle size, higher fuel economy, and longer refueling range.

- *Dynamic Model Results with Market Evolution*

This sub-section presents three dynamic discrete choice models, in addition to the specification presented as above, with consideration of market evolution. In the first model only one attribute (gasoline price) is dynamic over the considered time horizon and repeated purchases are possible. For each SP scenario, gasoline price follows an AR(1) model, and the residuals are standard normal distributed. The parameters of the AR(1) model are calibrated using historical data; in particular I used gasoline prices from April 1993 to September 2015 (1169 observations). The calibrated AR model presents the following specification:

$$y_{j,t+1} = 0.046458 + 0.98607 * y_{jt} + 0.05318 * \upsilon_{j,t+1} \tag{5.20}$$

where $y_{j,t+1}$ and y_{jt} correspond to gasoline price (unit: \$/gallon) of adjacent time periods; and $\upsilon_{j,t+1}$ follows a standard normal distribution. From this formula, we can observe that the autoregressive factor is very close to one while the drift and standard deviation of the error are close to zero. The pattern indicates that gasoline prices have been relatively stable in the real market from 1993 to 2015. I use this formula to generate households' perspective dynamic gasoline price in each scenario for dynamic model estimation; the corresponding results are shown in the column named "AR_R" of Table 5.1.

All of the estimated coefficients are significant and have a reasonable sign except for fuel economy of gasoline vehicle and the indicator for educated female. Unlike the MNL model, the magnitudes of gasoline price and electricity price indicate that households are more sensitive to electricity price than to gasoline price. Another important observation is that the marginal effects of fuel economy are quite different between the two groups considered, those who know the fuel economy of their current vehicle and those who do not. Compared to the previous dynamic structure without market evolution, households are less sensitive to vehicle size and range. Although the magnitudes of the remaining coefficients slightly change, the signs and effects of these coefficients are consistent with the previous models.

The second and third dynamic models with market evolution are extensions of the first dynamic model presented and assume that gasoline price and electricity price vary simultaneously over time; one allows repeated purchases and the other allows one-time purchase only. By assuming that gasoline price and electricity price for each SP scenario follow a vector auto-regressive model, I used monthly gasoline and electricity prices from January 2003 to September 2015 (153 pairs of observations) to calibrate the factors of the vector auto-regressive model. Drifts, and variance-covariance matrix of errors are determined under the hypothesis that the residuals follow a standard multivariate normal distribution. The final specification for the vector autoregressive model is presented as follows:

$$\begin{bmatrix} y_{1,t+1} \\ y_{2,t+1} \end{bmatrix} = \begin{bmatrix} 0.071 \\ 0.529 \end{bmatrix} + \begin{bmatrix} 0.966 & -0.024 \\ 0.088 & 0.838 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} + \begin{bmatrix} 0.032 & -0.003 \\ -0.003 & 0.131 \end{bmatrix} \begin{bmatrix} v_{1,t+1} \\ v_{2,t+1} \end{bmatrix} \quad (5.21)$$

where $\begin{bmatrix} y_{1,t+1} \\ y_{2,t+1} \end{bmatrix}$ and $\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}$ correspond to gasoline price (unit: \$/gallon) and electricity price (unit: \$/gallon-equivalent electricity) of adjacent time periods; and $\begin{bmatrix} v_{1,t+1} \\ v_{2,t+1} \end{bmatrix}$ follows a standard multivariate normal distribution. From this formula, we can observe that the autoregressive factor of electricity price is 0.838, smaller than that of gasoline price 0.966. The drift of gasoline price is very close to zero while that of electricity price is 0.529. The variance of the errors for gasoline price is close to zero while that of electricity price is 0.131. This pattern indicates that gasoline price is relatively stable in the market, while electricity price fluctuated from 2009 to 2015. The formula is used to generate households' perspective gasoline and electricity prices at each scenario. The dynamic models for repeated purchases or one-time purchase are estimated, and the corresponding results are shown in the column named "VAR_R" or "VAR_S" in Table 5.1.

When repeated purchases are considered, all coefficients have a reasonable sign and most of them are significant except for vehicle size, fuel economy of gasoline vehicle, and the indicator of educated female. Although small changes are observed, the estimation results of the dynamic structure considering two evolving variables are quite consistent with the one obtained considering one dynamic variable. In general, the fit of the model improves when we consider the dynamic nature of this problem; the rho-squared increases from 0.367 in the MNL model to 0.658 in the dynamic model with market evolution and repeated purchases.

When one-time purchase is considered, all coefficients have a reasonable sign. However, most of the households' social-demographic variables and the characteristics of electric vehicles are not significant. Obviously, this model is not appropriate to forecast households' vehicle purchase decisions based on the MVSPS data. This is because that the survey allows respondents to return to market and make another purchase every three years, which cannot be captured by the dynamic model with one-time purchase.

5.8.2 Market Share Forecast

The estimated coefficients are used to predict the market share of different vehicle types, which measures the prediction power of both static and dynamic models allowing repeated purchases. Figure 5.2 presents and compares the observed and predicted trends of market share of keeping the current vehicle, buying a new gasoline vehicle, a new hybrid vehicle, and a new electric vehicle along the 15 scenarios offered to the respondents over the nine-year period. In Figure 5.2, the red line represents the observed market share; the green line is associated with the MNL model; the purple, blue, and orange lines are associated with dynamic model without market evolution, evolving gasoline price, and evolving gasoline and electricity prices, respectively. The probability of keeping the current vehicle is relatively high: it starts at 50% in the first scenario, it increases up to 90% for the following three years, then it returns to 55% in the seventh scenario, jumps to 90% again for the following three years, and goes down to 60% in the thirteenth scenario. New gasoline, hybrid, and electric vehicles occupy smaller market shares: starting at 20%, 23%, and 7% respectively in the first scenario, they decrease to less than 5%, and then go up again in the third year. The big fluctuations in our data are due to the survey design; respondents who purchase a new vehicle are assumed to be out-of-market for the following three years and during this time period they are restricted to keep their current vehicles. By observing the peak values over the 15 scenarios, the market shares of keeping the current vehicle and buying an electric vehicle slightly increase from 50% to 60%, and from 7% to 9% respectively. On the other hand, the market shares of choosing new gasoline and hybrid vehicles decreases from 20% to 15%, and from 23% to 16% respectively during the same period.

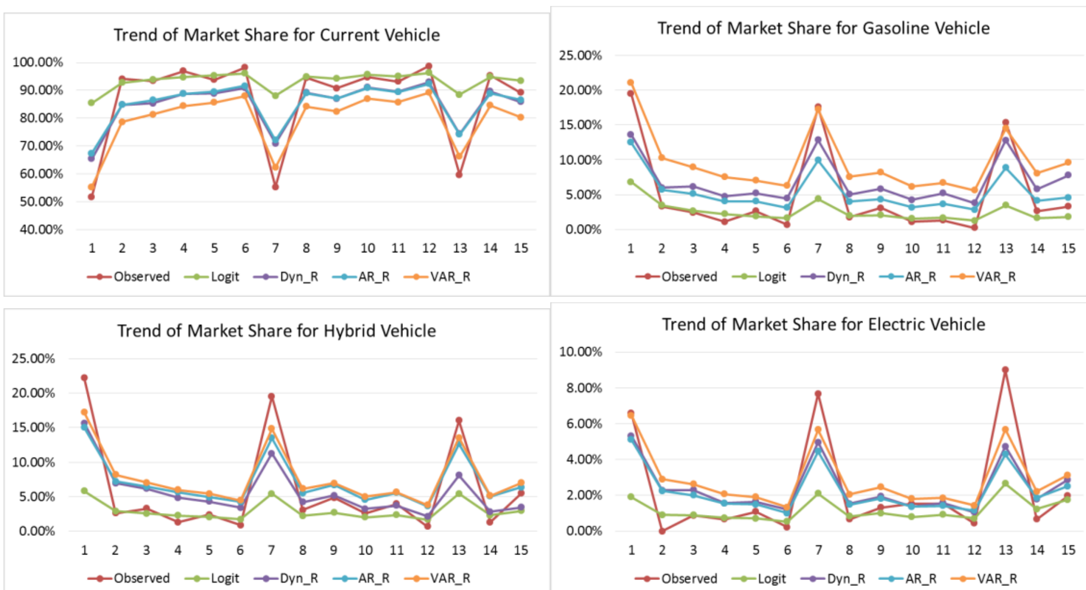


Figure 5. 2 Comparison of market prediction across static and dynamic models with repeated purchases

From Figure 5.2, we can observe that the static MNL model predicts a very stable market share and it is incapable to capture fluctuations and peaks of the market share. More specifically, it predicts well only the upper bounds of market share of keeping the current vehicle and the lower bounds of buying new gasoline, hybrid, and electric vehicles respectively. All three dynamic models are able to recover the fluctuations of the real market share, especially the model with evolving gasoline and electricity prices that approximates all the peaks over the 15 scenarios. However, the dynamic models underestimate the upper bound of market share of keeping the current vehicle, and overestimate the market share lower bound of buying new gasoline, hybrid, and electric vehicles. To summarize, the dynamic models are excellent to predict fluctuations and peaks in market shares while the MNL model averages market shares over time and fails to detect sudden changes in consumer demands.

Figure 5.3 compares the prediction power of two dynamic models with evolving gasoline and electricity prices: the blue line allows repeated purchases and the green line allows one-time purchase. We can observe that the one-time purchase model averages the market shares over time and is incapable of predicting fluctuations, peaks, upper bounds and lower bounds in the real market share. On the other hand, the model allowing repeated purchases does an excellent job in predicting fluctuations and peaks of the actual market share.

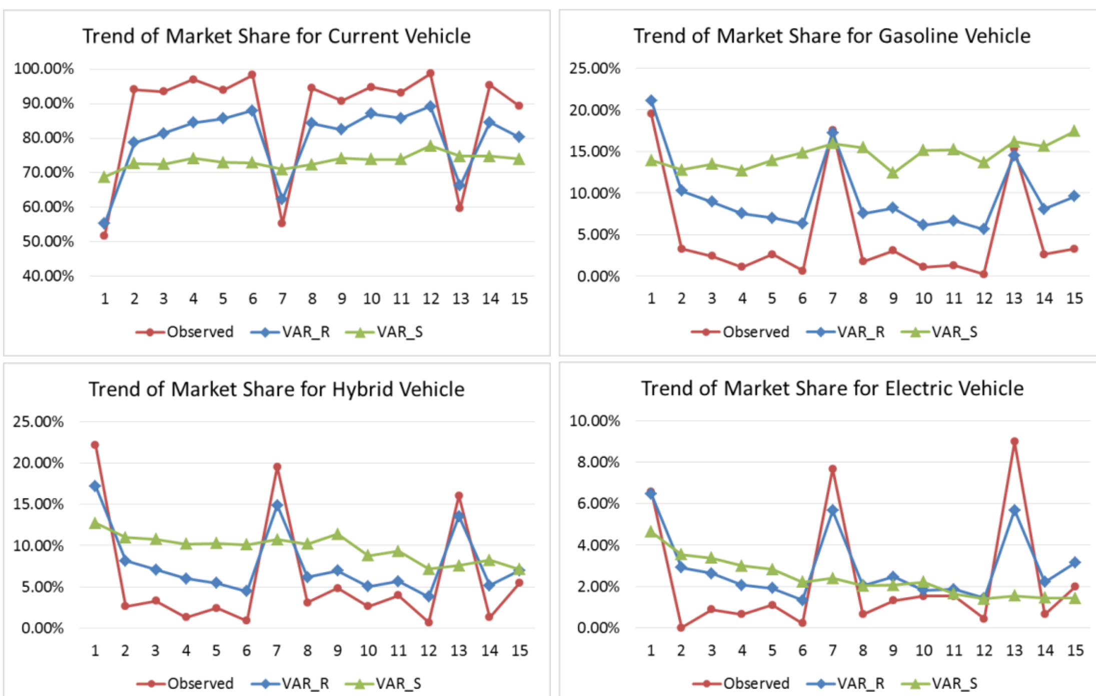


Figure 5. 3 Comparison of market share prediction across dynamic models allowing repeated purchases and one-time purchase

5.8.3 Cross-Sample Validation and Policy Implications

In order to validate the model results, I re-estimated both static and dynamic models on 80% of the sample and applied the model estimates to the remaining 20% of the sample. Figure 5.4 reports the Root Mean Square Error (RMSE) of market shares calculated respectively for the static logit model and the three dynamic models over the fifteen time periods considered. The RMSE values suggest that the logit model has the highest prediction error, especially in reproducing the market share of the current vehicle; while in comparison the dynamic models performs equally well.

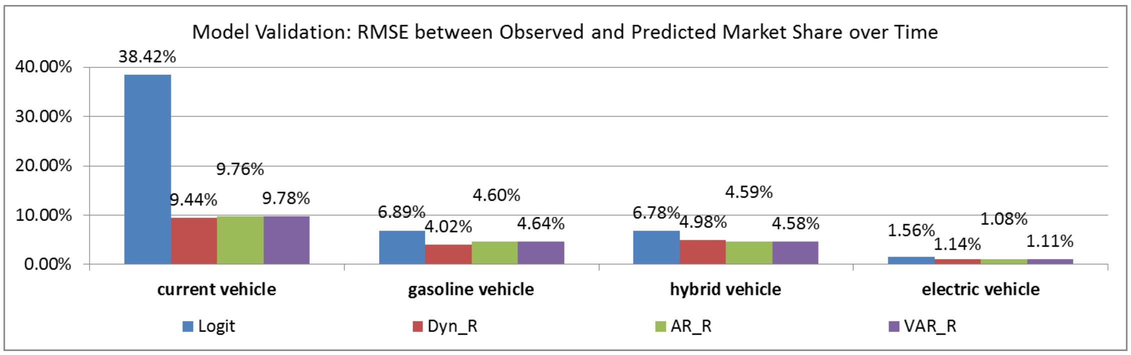


Figure 5. 4 Model validation results: RMSE between observed and predicted vehicle market share over time

The estimation results of the dynamic model with evolving gasoline price and electricity price have been applied to test policy scenarios; the variables of interest are fuel price (i.e., gasoline price and electricity price), vehicle purchase price (i.e., hybrid vehicle price and electric vehicle price), and characteristics of electric cars (i.e., MPG equivalent electricity and recharging range). More specifically, the scenarios investigated are as follows:

- *Fuel price*

  Gasoline price over 15 time periods: 10% decrease, 10% increase

  Electricity price over 15 time periods: 10% decrease, 10% increase

- *Vehicle purchase price*

  Price of hybrid car over 15 time periods: 10% decrease, 10% increase

  Price of electric car over 15 time periods: 10% decrease, 10% increase

- *Technology improvement*

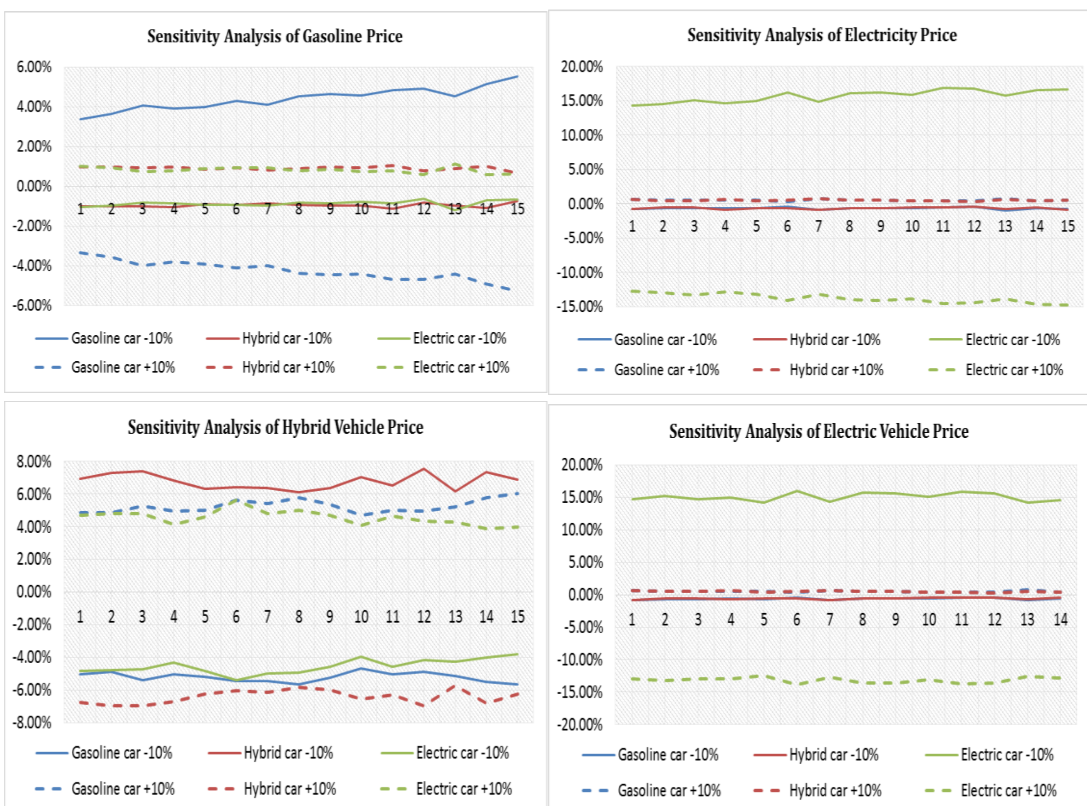
  MPG equivalent electricity over 15 time periods: 10% decrease, 10% increase

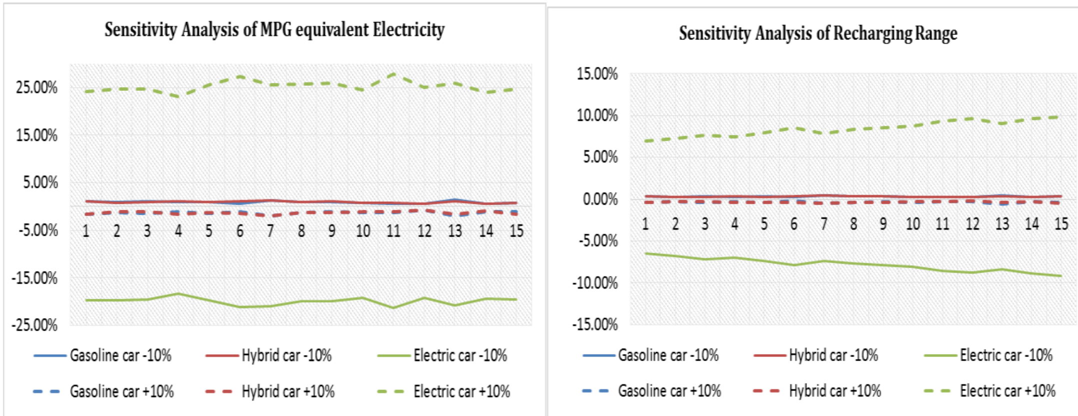
  Recharging range of electric car over 15 time periods: 10% decrease, 10% increase

Results in Figure 5.5 show how the changes of these variables influence households' decisions of purchasing gasoline, hybrid, or electric cars over time at an aggregate level. Overall, the impact of all tested variables on vehicle type decisions is significant. Changes in fuel price have a large effect on the purchase of the corresponding vehicle type, especially the changes of electricity price on the purchase of electric cars. We observe that the effect of gasoline-price changes on gasoline-vehicle purchase gradually increases over the 15 time periods, while it is not obvious to identify a trend for the change of electricity price.

Changes in vehicle price also have a large effect on vehicle type choices. The effects have different patterns under the changes of hybrid and electric vehicle prices. The decrease/increase of hybrid vehicle price encourages/discourages households to buy hybrid cars, while discourages/encourages them to buy gasoline and electric cars. For example, at time period 1, a 10% decrease in the price of hybrid car leads to a 7% increase in the purchase of hybrid car and a 5% decrease in the purchases of gasoline and electric cars. We observe that the impact of hybrid vehicle price on vehicle type choices fluctuates over time. The changes in the price of electric car only influence the choice of electric car, and the effects on gasoline and hybrid cars are negligible. For example, a 10% increase in electric vehicle price leads to a 13-14% decrease in the purchase of electric car, and less than 1% increase in the purchase of other vehicle types. We can observe little variation in the effect of electric vehicle price on vehicle type choices over time.

Additionally, I test some variables related to the technology improvement of electric car, such as MPG equivalent electricity and recharging range. We observe that the purchase of electric car is very sensitive to the change of car characteristic variables, especially MPG equivalent electricity. To summarize, compared to the purchase of gasoline and hybrid cars, the purchase of electric car is more sensitive to the change of fuel price, vehicle price, and car characteristics.





Note: The above six pictures describe the changes in the market share of gasoline, hybrid and electric cars when the target variables increase or decrease by 10%. For example, in the first picture, "Gasoline car -10%" or "Gasoline car +10%" means the change in the market share of gasoline car when gasoline price decreases or increases by 10%.

Figure 5. 5 Application results of dynamic models with two evolving attributes: sensitivity analysis of fuel price, vehicle price, and electric car characteristics

5.9 Chapter Conclusions

This Chapter formalizes a general dynamic discrete choice framework in which forward-looking agents optimize their utility over time in a finite time horizon. The main strengths of the proposed model can be summarized as follows:

- In the dynamic model framework, the utility function is non-linear, which accounts for information both on current alternatives and on individual expectations about future alternatives.
- The model framework allows decision makers to have more than one starting conditions, and it considers heterogeneous population and products.
- The model is generalized to consider purchase behaviors in different markets; agents can either return to market or leave the market after a purchase is made.

- The number of agent’s forward-looking time periods, considered for the calculation of the expected future utility, is flexible.
- The dynamic discrete choice model is integrated with a stochastic diffusion process to jointly capture market evolution.

The proposed model framework has been successfully applied to predict the adoption rates of different vehicle types including gasoline, hybrid, and electric vehicles. Model estimations are coherent with general expectations. Model validation shows that dynamic models are particularly appropriate to recover peaks/valleys and rapid changes in consumer demand over time. On average, the dynamic models have a better performance in predicting vehicle market shares.

Although the proposed model has a dynamic nature, it does not consider state dependency or panel effect over choices made by the same individual. More specifically, the model restrictively assumes that the error components are i.i.d. over household, product, and time period. Besides, as the model structure is developed based on the logit model, it fails to capture the correlations between different alternatives. Moreover, the model only considers household’s discrete choices of purchase time and vehicle type decisions. These decisions are in fact highly influenced by households’ vehicle usage behavior, which is not captured here. In the future, the model can be further improved by considering a joint decision of car ownership and usage.

Chapter 6: Methodology Part 3: Integrated Discrete-Continuous Choice Model

6.1 Introduction

There are wide applications on joint discrete and continuous choices in different areas such as car ownership and use (Liu et al., 2014), activity type and duration (Cirillo et al., 2015-2), energy appliance type and demand (Vaage. 2000). Recently many researchers develop models that can simultaneously capture discrete and continuous decisions, among which Liu and Cirillo (2016) proposed an integrated discrete-continuous car ownership model combined with MOVES for vehicle emission estimation and green policy evaluation. This Chapter generalizes their model framework: (a) to forecast the penetration of "green" vehicle (i.e., hybrid and electric cars) in the market, and (a) to predict vehicle ownership, use, and emission patterns both in developed countries (i.e., the US) and in developing countries (i.e., China).

The generalized model framework integrates four sub-models: (a) vehicle type and vintage choice (discrete); (b) vehicle quantity choice (discrete); (c) vehicle usage choice (continuous); and (d) vehicle GHG emission rates estimator. Vehicle quantity sub-model accounts for vehicle type/vintage preferences by incorporating the mode of vehicle type sub-model. Vehicle type sub-model is flexible to account for information of conventional vehicles and "green" vehicles. Regressions are used to estimate the annual VMT of each household car. The vehicle quantity choice and vehicle usage choice are integrated by an unrestricted full variance-covariance matrix, which considers the interdependence between discrete and continuous choices. The model framework combines with MOVES, which calculates emission rates for different vehicle types, to estimate household-level vehicle emissions.

The following sections include a generalized formulation of the integrated discrete-continuous choice model, an application exploring the impact of “green” vehicle adoption on vehicle ownership, use, and emission patterns in Maryland, and another application predicting residents’ behavior on vehicle ownership, use, and emissions in Beijing, China. The first application is based on the MVSPS data in Section 3.1, while the second application is based on the BHTS data in Section 3.5.

6.2 Discrete Choice Sub-Model

MNL models are employed to capture household decisions on vehicle type. For example, in the first application, we consider five different vehicle types categorized by their fuel type (gasoline, electricity, hybrid) and model year (less or equal to 3 years old, greater than 3 years old). The alternatives include new gasoline vehicle, new HEV, new BEV, old gasoline vehicle, and old HEV, among which old gasoline vehicle is the most popular alternative. The utility function of choosing any vehicle type can be formulated as follows:

$$U_{t_j} = V_{t_j} + \varepsilon_{t_j} \quad \text{and} \quad V_{t_j} = X_{t_j}^T \beta_{t_j} \tag{6.1}$$

where t_j is the full choice set for households held j vehicles; U_{t_j} and V_{t_j} are the indirect and direct utilities for households choosing any vehicle type among the full choice set t_j, respectively; ε_{t_j} is the unobserved error term of the utility function, which follows T1EV distribution with scale parameter normalized to 1; X_{t_j}

represents a list of independent variables of car characteristics which are important indicators for households' decisions; and β_{t_j} is a list of marginal utilities, associated with independent variables, to be estimated.

A multinomial probit model is employed to forecast the number of vehicles held by households. For example, in the first application, there are three alternatives including one, two, and three or more vehicles; while in the second application, we have three alternatives including zero, one, and two or more vehicles. Households are assumed to be rational and choose the alternative with the maximum utility. The utility function of vehicle quantity choice is formulated as follows:

$$U_j = V_j + \alpha L_j + \varepsilon_j \quad and \quad V_j = X_j^T \beta_j \tag{6.2}$$

where U_j and V_j are the indirect and direct utilities for households held j vehicles respectively; ε_j is the unobserved error term which follows normal distribution with mean zero; L_j presents the expected maximum utility, the so-called logsum, calculated from the vehicle type sub-model. It serves as an indicator describing whether and how the diversity of vehicle type will influence households' decisions on vehicle quantity; X_j is a list of independent variables including household social-demographics and land use information; and α, β_j are the corresponding coefficients to be estimated.

6.3 Continuous Choice Sub-Model

Linear regression models are adopted to estimate the annual VMT of each household vehicle; the formulation is as follows:

$$Y_{VMT,s} = X_s^T \beta_s + \varepsilon_s \tag{6.3}$$

where s represents primary, secondary, or tertiary vehicle; $Y_{VMT,s}$ is the dependent variable describing annual VMT of households' primary, secondary, or tertiary vehicle; X_s is a list of explanatory variables including household social-demographics, residential density, and driving cost; β_s is a list of coefficients to be estimated; and ε_s is an unobserved error term which follows normal distribution with zero mean.

6.4 Integration of Discrete and Continuous Choices

To estimate vehicle quantity and usage decisions simultaneously, the joint probability is expressed as the product of the marginal probability of driving certain miles and the conditional probability of choosing a certain number of vehicles based on the miles driven (Liu, 2013).

$$P(Y_{disc}, Y_{VMT}) = P(Y_{VMT})P(Y_{disc}|Y_{VMT}) \tag{6.4}$$

where Y_{disc} is households' discrete choice of choosing certain number of vehicles; Y_{VMT} is households' continuous choice of driving certain miles; $Y_{disc}|Y_{VMT}$ represents choosing certain number of vehicles conditional on the miles driven; and P represents the probability function.

Taking the advantage that both error terms of the regressions (ε_s) and the probit model (ε_j) follow a multivariate normal distribution, the conditional term ($\varepsilon_j|\varepsilon_s$) also follows a multivariate normal distribution with new mean and variance-covariance matrix (Liu et al., 2014).

6.5 Calculation of Vehicle Greenhouse Gas Emissions

To calculate vehicle GHG emissions, we should first estimate households' vehicle type, quantity, annual VMT, and GHG emission rates for different vehicle types from the joint modeling framework. The GHG emission rates can be obtained from vehicle emission simulators such as MOBILE, MOVES, EMBEV, and CORPERT IV. In particular, the first application assumes that the direct emission from battery electric cars is zero. The annual GHG emissions of gasoline and hybrid vehicles can be calculated according to Equation 6.5 and 6.6, respectively.

$$AGHGEs\left(\frac{grams}{vehicle}\right) = RERs\left(\frac{grams}{vehicle-mile}\right) \times AVMT\left(\frac{miles}{year}\right) + SERs\left(\frac{grams}{vehicle-day}\right) \times 365\left(\frac{days}{year}\right) \quad (6.5)$$

$$AGHGEs\left(\frac{grams}{vehicle}\right) = RERs\left(\frac{grams}{vehicle-mile}\right) \times AVMT\left(\frac{miles}{year}\right) \quad (6.6)$$

where $AGHGEs$ is annual GHG emissions; $RERs$ is running emission rates; $SERs$ is start/extended idle emission rates; and $AVMT$ is annual VMT. The start/extended idle emission rates are zero for HEVs because they use electricity to start. For simplification purpose, let's assume the driving days is 365 per year and no difference between weekday and weekend. This assumption can be relaxed in future works.

6.6 Application 1: Green Vehicle Ownership, Use, and Emission in Maryland

This application adopts the integrated discrete-continuous approach to model households' future preferences on vehicle type, quantity, use, and the relevant GHG emissions under the consideration of GVs in a dynamic market. The integrated model is estimated on a dynamic panel data derived from the MVSPS in Maryland that collected households' inter-temporal preferences over a nine-year time period. As a supplementary data support, I derive vehicle use information from the 2009 NHTS data.

6.6.1 Model Estimation Results

- *Estimation Results of Vehicle Type Sub-Models*

A multinomial logit model has been employed to investigate households' time-dependent preferences on GV adoption and trade-offs between different vehicle characteristics such as purchase price, fuel economy, refueling range, cargo space, and fuel capacity. Specifically, I estimate short-run and medium-long-run vehicle purchasing patterns based on the MVSPS data relative to the first four years (2014-2017) and the entire nine years (2014-2022). The expected maximum utility (logsum) from the vehicle type model serves as an important indicator for the diversity of vehicle types in the market.

Based on the model setting, there are 5^j different vehicle type choices (alternatives) for households with j vehicles. However, it is considered to be infeasible to estimate the model on a full set of alternatives especially for households with three or more vehicles. By taking the advantage of the independence of irrelevant alternatives (IIA) property of logit model, we estimate trade-offs between characteristics of different vehicle types on a randomly selected subset of alternatives. Train (1986) stated that beyond a minimal number of alternatives, the estimated parameters are not sensitive to the number of alternatives in the estimation. Table 6.1 reports and compares the estimation results of vehicle type sub-models between the short run and medium-long run.

Table 6. 1 Vehicle Type Sub-Model Estimation Results for the Short and Medium-Long Runs

Variables [unit]	Utilities					1-car household		2-car household		3-car household	
	New gasoline	New hybrid	New electric	Old gasoline	Old hybrid	Medium Run Est.	Short Run Est.	Medium Run Est.	Short Run Est.	Medium Run Est.	Short Run Est.
Purchase price (inc.<75k) [$10,000]	X	X	X	X	X	-0.612	-0.656	-0.222	-0.247	-0.150	-0.161
Purchase price (inc.>=75k) [$10,000]	X	X	X	X	X	-0.292	-0.340	-0.135	-0.169	-0.070	-0.075
Fuel economy [100 MPG/MPGE]	X	X	X	X	X	0.828	0.887	0.437	0.448	0.350	0.134*
Recharging range [100 miles]			X			-	-	0.575	0.605	0.830	0.852
Cargo space [cu.ft.]	X	X	X	X	X	0.040	0.058	0.057	0.065	0.067	0.065
Number of seats	X	X	X	X	X	0.099	-	0.048	-	0.052	0.040*
Engine size [liter]	X	X		X	X	0.045*	-	0.092	0.059*	0.103	0.062
Fuel capacity [gallon]	X	X		X	X	0.083	0.095	0.078	0.084	0.075	0.074
Log(num. of models in the class)	X	X	X	X	X	0.230	0.251	0.208	0.230	0.179	0.184
Shoulder room [inch]	X	X	X	X	X	-0.022	-	-0.024	-0.012	-0.024	-0.018
Head room [inch]	X	X	X	X	X	0.016	0.013	0.020	0.019	0.020	0.017
Leg room [inch]	X	X	X	X	X	-0.071	-0.058	-0.090	-0.082	-0.107	-0.100
Length [inch]	X	X	X	X	X	0.004*	0.003*	0.006	0.005	0.014	0.016
Width [inch]	X	X	X	X	X	-	-	-0.019	-0.017	-0.046	-0.054
Height [inch]	X	X	X	X	X	-0.072	-0.065	-0.074	-0.072	-0.080	-0.073
Number of observations						1107	615	1539	942	1031	621
Initial likelihood						-1781.6	-989.8	-3543.7	-2169.0	-3088.6	-1860.3
Final likelihood						-1360.4	-749.5	-2599.7	-1595.5	-2045.2	-1233.8
Rho-Squared						0.236	0.243	0.266	0.264	0.338	0.337

Note: "" means the coefficient is not significant at the significance level of 0.05; "X" means the variable in this row is considered in the utility function of the vehicle type in this column.*

We can observe that households generally prefer vehicles with larger space, higher fuel economy and larger engine size, longer refueling range, and lower sale price, which is consistent with previous studies (Bhat, 2000; Maness and Cirillo, 2012; Cirillo et al., 2017):

- ✓ Households tend to choose vehicles with larger space including larger cargo space, more seats, and higher fuel capacity as the coefficients of these variables are positive and significant. Most households choose to hold conventional gasoline vehicles because they have larger size compared with hybrid and electric cars. It is important to improve the size of GVs to attract more potential buyers.

- ✓ Households with fewer vehicles care more about fuel economy and the number of makes and models in certain vehicle class. Households with more vehicles care more about engine size, indicated by the magnitude of coefficients for engine size, fuel economy, and logarithm of vehicle makes and models.
- ✓ The coefficients of refueling range are significant and positive for households with more than one vehicle. And the increasing magnitude indicates that households with more vehicles prefer higher refueling range. This may be due to the fact that households with more vehicle have higher probability to hold a BEV.
- ✓ The coefficients related to vehicle sale price are negative and significant, and the magnitude is larger for households with lower income or with fewer vehicles. This pattern indicates that households with lower income or fewer vehicles are more sensitive to vehicle purchase price because they may have less money to support other living expenses. Additionally, the coefficient of vehicle purchase price is more negative in the short run, which indicates households are more sensitive to the purchase price in a shorter time period. This explains the lack of GV adoption in a shorter run.

- *Estimation Results of the Integrated Discrete-Continuous Choice Model*

The proposed model framework jointly predicts households' future preferences on vehicle type, quantity, and annual VMT for their primary, secondary, and tertiary vehicles. For each household, primary vehicle is defined as the one used the most, followed by second and tertiary vehicles if any. The integrated discrete-continuous choice model is estimated on the MVSPS data for a short run and a medium-long run. The sample for the short run contains 1844 observations between the year of 2014 and 2017, while the sample for the medium-long run contains 3677 observations over a nine-year period from 2014 to 2022. Table 6.2 reports the estimation results for the two scenarios.

Table 6. 2 Joint Choice Model Estimation Results: Short Run V.S. Medium-Long Run

Variable	Alternative	Medium-long run			Short run		
		Coefficient	S. D.	P-value	Coefficient	S. D.	P-value
Logsum of vehicle type	All	0.523	0.001	<0.001	0.185	0.084	0.027
Constant	2 cars	-6.709	0.053	<0.001	1.788*	1.002	0.074
	3 cars	-17.254	0.034	<0.001	3.755*	2.608	0.150
HH head Gender	2 cars	-0.355	0.045	<0.001	-0.206	0.103	0.045
	3 cars	-0.276	0.031	<0.001	0.176*	0.245	0.473
Education	2 cars	-0.087			0.026*	0.030	0.384
	3 cars	-0.114			0.016*	0.039	0.691
Income	2 cars	0.275	0.013	<0.001	0.178	0.041	<0.001
	3 cars	0.476	0.003	<0.001	0.338	0.097	0.001
Num. of kids	2 cars	0.202	0.019	<0.001	-0.118	0.057	0.039
	3 cars	0.145	0.019	<0.001	-0.319	0.063	<0.001
Res. density	2 cars	-0.054	0.009	<0.001	-0.044	0.015	0.004
	3 cars	-0.049	0.004	<0.001	-0.174	0.064	0.006
Constant	Regression for primary vehicle	3.218	0.002	<0.001	2.999	0.049	<0.001
HH head gender		-0.298	0.002	<0.001	-0.279	0.013	<0.001
HH head age		-0.231	0.002	<0.001	-0.211	0.005	<0.001
Income		0.042	0.001	<0.001	0.086	0.002	<0.001
Res. density		-0.047	0.001	<0.001	-0.041	0.001	<0.001
Driving cost		-6.359	0.002	<0.001	-6.170	0.250	<0.001

Constant	Regression for secondary vehicle	2.498	0.003	<0.001	2.260	0.048	<0.001
HH head gender		-0.370	0.001	<0.001	-0.281	0.016	<0.001
HH head age		-0.264	0.003	0.001	-0.216	0.006	<0.001
Income		0.052	0.001	<0.001	0.094	0.002	<0.001
Res. density		-0.048	0.002	<0.001	-0.047	0.001	<0.001
Driving cost		-7.026	0.002	<0.001	-6.339	0.290	<0.001
Constant	Regression for tertiary vehicle	2.191	0.003	<0.001	1.266	0.196	<0.001
HH head gender		-0.136	0.004	<0.001	0.005*	0.043	0.909
HH head age		-0.135	0.001	<0.001	0.014	0.024	0.564
Income		0.018	0.001	<0.001	0.079	0.027	0.003
Res. density		-0.029	0.001	<0.001	-0.071	0.012	<0.001
Driving cost		-4.711	0.012	<0.001	-1.218	0.455	0.007
Log-likelihood at zero		-5830.7			-2856.5		
Log-likelihood at convergence		-789.8			-514.1		
Number of observations		3677			1844		
R square		0.865			0.820		

Note: "" means the coefficient is not significant at the significance level of 0.05. HH head gender: 1 for female and 0 for male.*

The estimation results of the integrated discrete-continuous car ownership model can be interpreted as follows:

The variable named "logsum of vehicle type" represents the expected maximum utility of the vehicle type sub-model. It is an important indicator illustrating how the introduction and diversity of GVs influence households' vehicle quantity and use decisions. We can observe that the corresponding coefficients are significant, positive, and between zero and one, which is consistent with previous study (Liu and Cirillo, 2016). Besides, the value of this coefficient in the medium-long run is 0.523, almost two times larger than the value estimated for the short run 0.185. This pattern indicates that the diversity of vehicle types considering both gasoline vehicles and GVs has a higher positive impact on car ownership and use for a longer time period.

The coefficients of households' income are positive and significant both in vehicle quantity and usage parts, which indicates that households with higher income tend to hold more vehicles and drive more.

The negative coefficients of household head gender indicate that male household heads are more likely to hold more vehicles and to drive the primary, secondary, and tertiary vehicles more frequently. However, the coefficients are not significant for the vehicle quantity part in the short run.

The coefficients of education level are significant and negative only for the medium-long run, which indicates that household head with higher education level prefer fewer vehicles in the medium-long run. However, the parameters are not significant for the short run.

The coefficients of household head age are negative for the vehicle usage part, which indicates households with younger head drive more.

The negative coefficients of residential density indicate that households living in areas with higher population density prefer to have fewer vehicles and to drive less. In other words, households living in suburban or rural areas are more likely to have more vehicles.

The coefficients of driving cost are negative and significant, which indicates that households tend to drive less when fuel cost increases. From the magnitude of the coefficients, we observe the use of households' primary and secondary vehicles is more sensitive to the fuel cost than the use of tertiary vehicles.

- *Estimation of Vehicle GHG Emissions*

To calculate vehicle GHG emissions, I first derive GHG emission rates both for conventional gasoline and hybrid vehicles from the reported values by Liu and Cirillo (2016) and US DOE. In particular, the average running and start/extended idle GHG emission rates of gasoline vehicles are 401 grams/mile and 678 grams/day, respectively (Liu and Cirillo, 2016). The average running emission rate of hybrid vehicles is 0.51 pound/mile, which is equivalent to 231 grams/mile, from US DOE. Then, we calculated annual GHG emissions for different vehicle types according to equations 6.5 and 6.6. In Figure 6.1, we can observe that the average annual GHG emissions in the short run are 5.17 tons, 3.71 tons, and 3.62 tons for households' primary, secondary, and tertiary vehicles respectively. The average annual GHG emissions in the medium-long run are slightly lower, which might indicate that more households are willing to consider GVs or drive less in a longer time period. Besides, taking GVs into account, the average annual GHG emission is lower than the value (5.2 tons) reported by the Environmental Protection Agency (EPA, 2013).

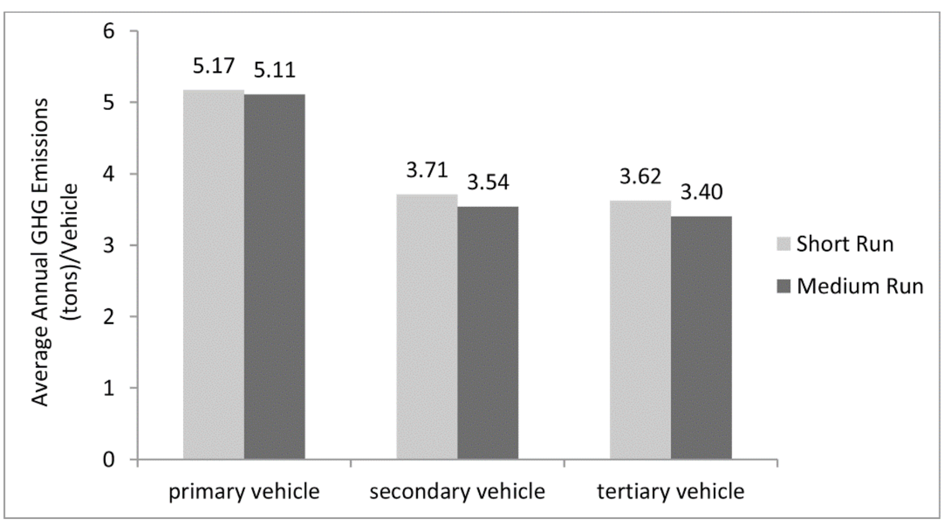


Figure 6. 1 Average annual GHG emissions per vehicle: short run V.S. medium run

The average annual GHG emission of households' primary vehicle is much larger than that of the secondary and tertiary vehicles. This is because the predicted average annual VMT for the primary vehicle is 16123 miles, much larger than that of the secondary vehicle (10450 miles) and tertiary vehicle (7229 miles). Compare with the average emissions between the secondary and tertiary vehicles, households' secondary vehicles probably have a higher percentage of GVs.

6.6.2 Sensitivity Analysis and Policy Implications

This section evaluates the impact of two "green" taxes including gasoline tax and vehicle ownership tax, and quantifies their influences on households' car ownership, use, and GHG emissions. For each type of tax, we proposed three plans with increasing taxation rates, named "policy 1", "policy 2", and "policy 3". The following figures present and compare the average change rates of annual VMT, vehicle quantity, and annual GHG emissions under the implementations of different taxation plans in the short and medium-long runs.

- *Sensitivity Analysis for Gasoline Tax*

Gasoline tax is a tax on gas price applied to conventional gasoline vehicles and HEVs. The proposed three plans increase gas price by 5%, 10%, and 20% of the original price. This type of tax is designed to reduce vehicle GHG emissions mainly by decreasing vehicle usage. Figure 6.2 shows the annual VMT reduction rates under the three gasoline taxation plans in the short and medium-long runs.

We can observe gasoline taxes effectively reduce the average annual VMT especially in the medium-long run: the average reduction rates under the three taxation plans are 4.34%, 8.59%, and 15.41%. Gasoline taxes have the greatest impact on the secondary vehicles followed by the primary vehicles, which may be because households' secondary vehicles contain more conventional gasoline vehicles that are more sensitive to gasoline taxes.

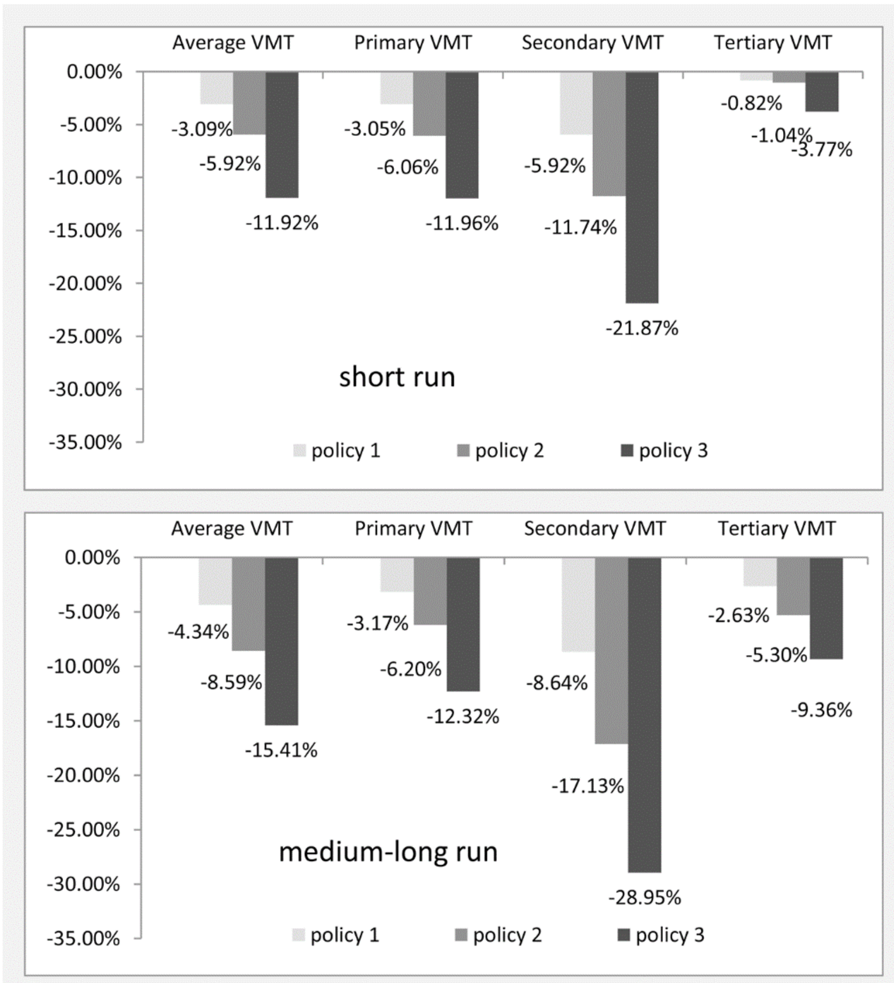


Figure 6. 2 Annual VMT reduction rates under gasoline taxes in the short and medium-long runs

Figure 6.3 shows the change rates of households' vehicle quantity under the three gasoline taxation plans in the short and medium-long runs. We can observe that these taxes have little influence on the reduction of vehicle quantity in both scenarios. Although proposing a 20% increase to gas price, the reduction rates of vehicle quantity are lower than 1%.

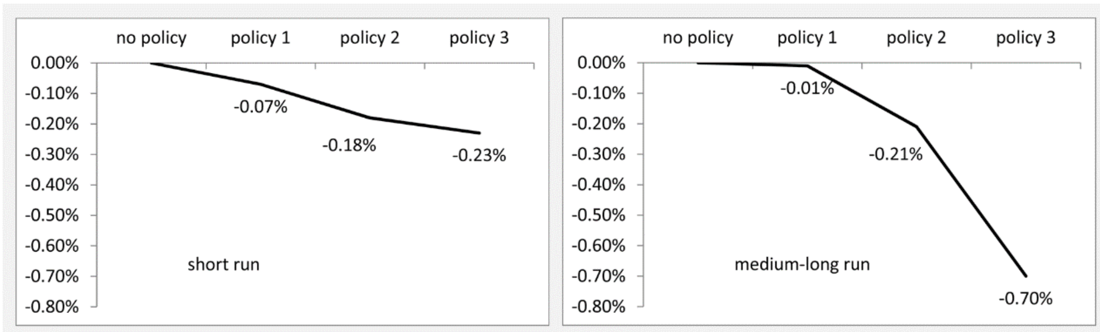


Figure 6. 3 Car quantity change rates under gasoline taxes in the short and medium-long runs

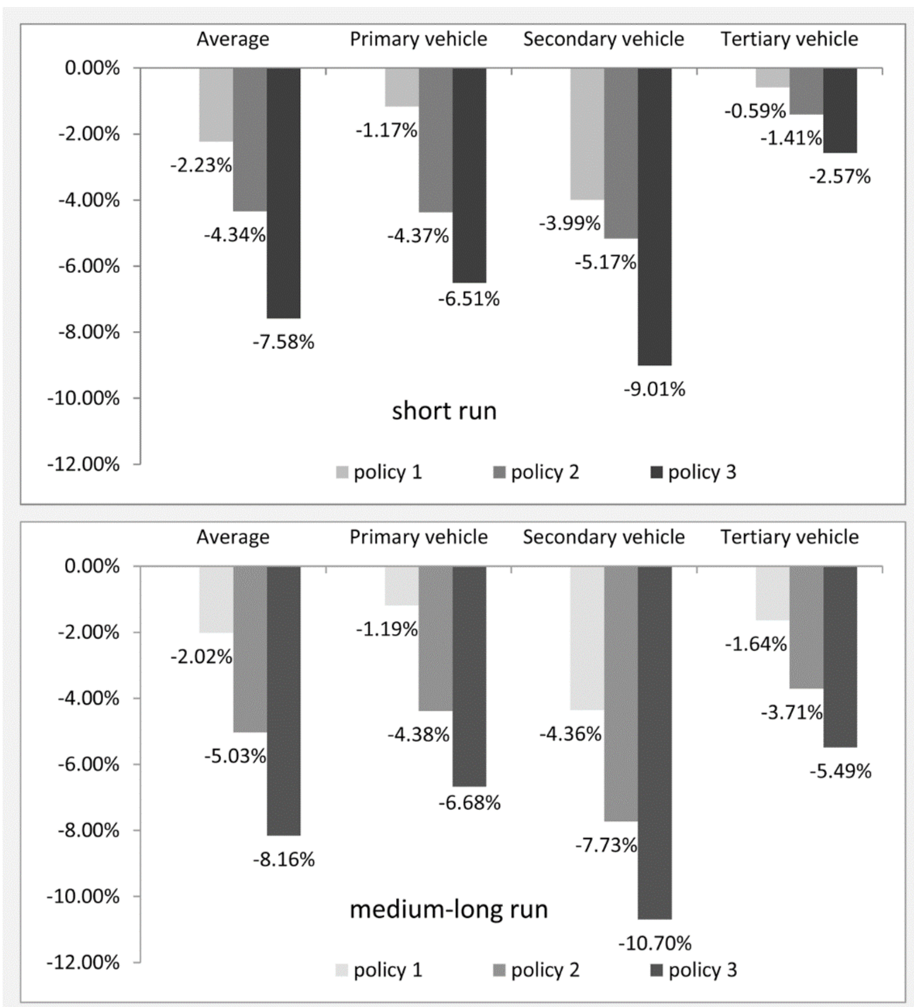


Figure 6. 4 Annual GHG emissions reduction rates under gasoline taxes in the short and medium-long runs

Figure 6.4 shows annual GHG emission reduction rates under the three gasoline taxation plans. We can observe that gasoline taxes are effective in reducing GHG emissions: the average reduction rates under “policy 3” are 7.58% and 8.16% for the short run and the medium-long run, respectively. In addition, gasoline taxes have much higher impact on emission reductions for households’ secondary vehicles; this pattern is more obvious in the medium-long run.

- *Sensitivity Analysis for Ownership Tax*

Ownership tax is an annual fee for households who held one or more conventional gasoline vehicles. I proposed three plans requiring an annual ownership fee of $1000, $2000, and $3000. Ownership tax is designed to reduce vehicle GHG emissions by inducing households to decrease vehicle quantity or to switch to more fuel-efficient vehicle type. Figure 6.5 shows the annual VMT reduction rates under the three proposed taxation plans in different time periods.

Either in the short run or in the medium-long run described by Figure 6.5, ownership taxes are not effective in reducing vehicle use even if a high annual fee of $3000 is charged. Comparatively, ownership tax has a higher impact on reducing VMT in the short run than in the medium-long run.

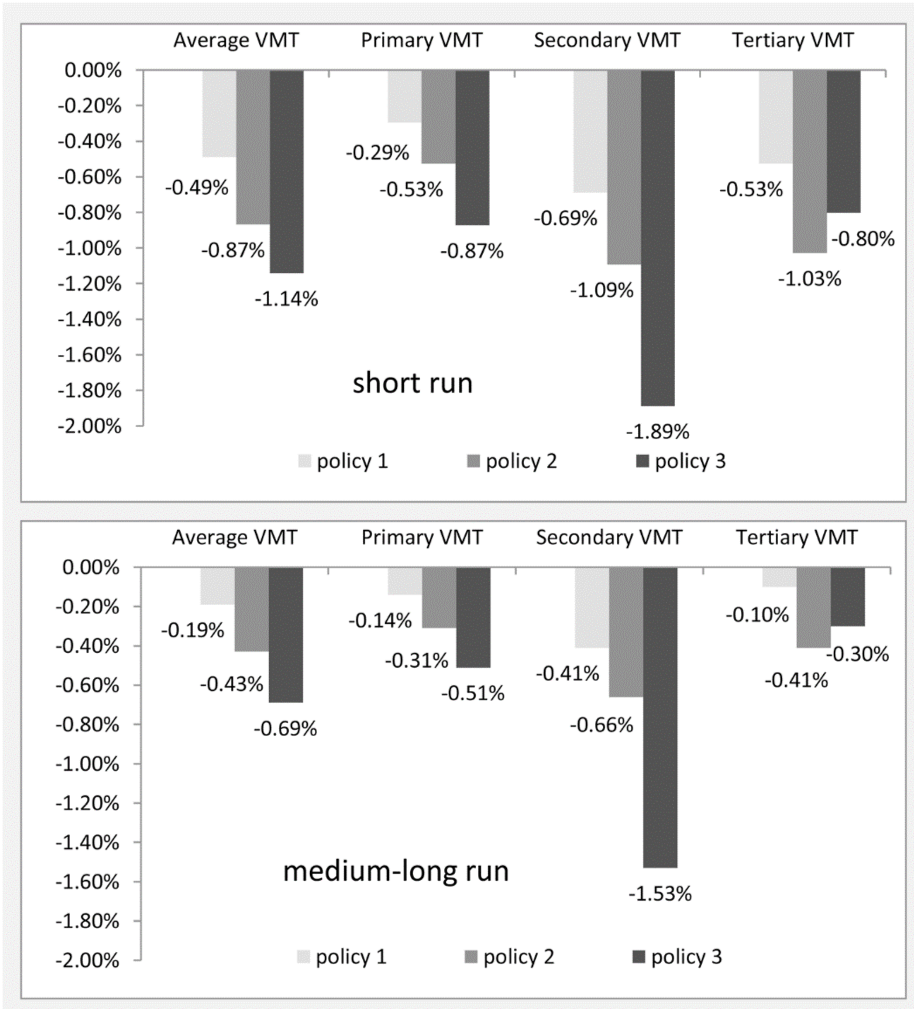


Figure 6. 5 Annual VMT reduction rates under ownership taxes in the short and medium-long runs

Figure 6.6 shows the change rates of households' vehicle quantity under the three ownership taxation plans in the short and medium-long runs. We can observe that this type of tax slightly reduces vehicle quantity in both scenarios. However, although the annual ownership fee is increased to $3000, the reduction rates of vehicle quantity are smaller than 1%. Similar with the impact on VMT reductions, the ownership tax has a higher impact on vehicle quantity reductions in a shorter time period.

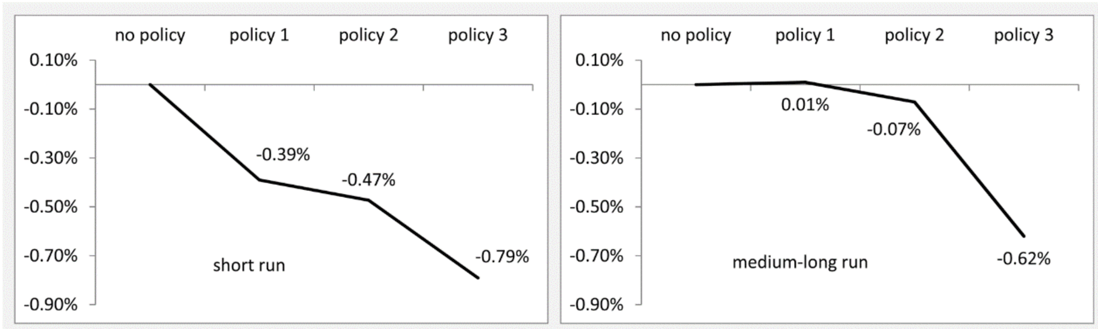


Figure 6. 6 Car ownership change rates under ownership taxes in the short and medium-long runs

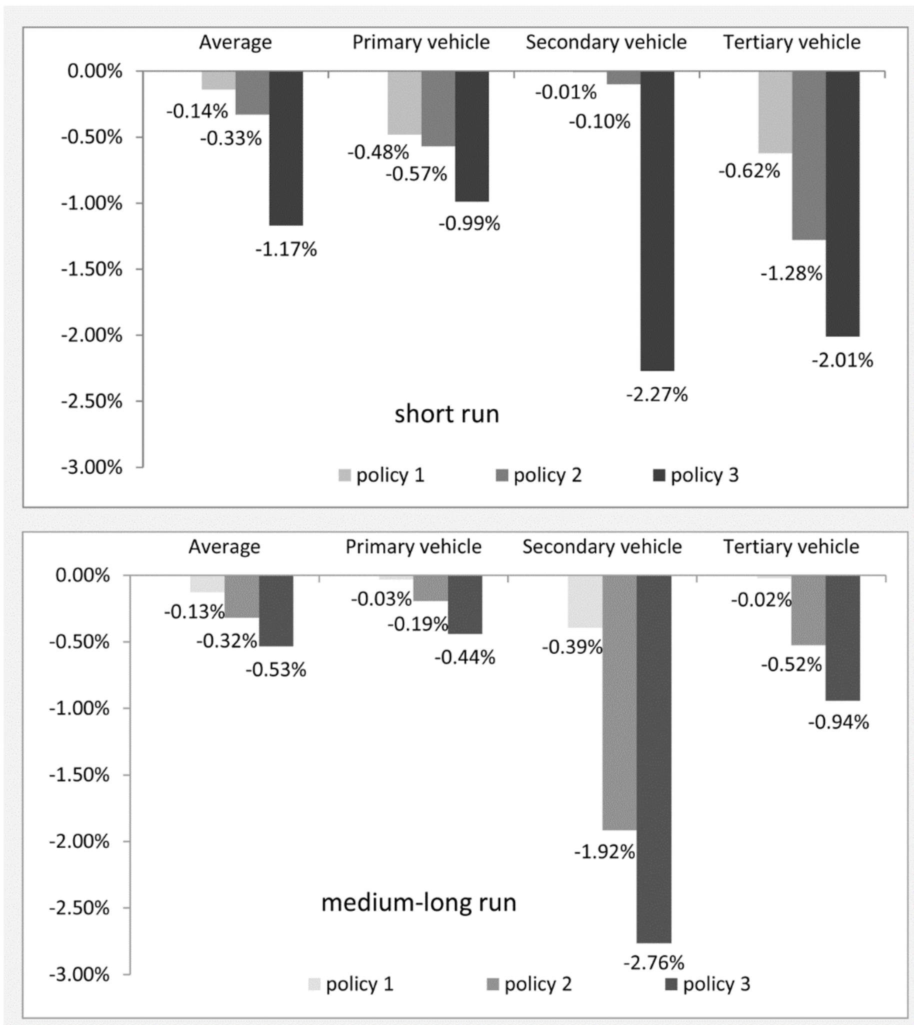


Figure 6. 7 Total GHG emissions reduction rates under ownership taxes in the short and medium-long runs

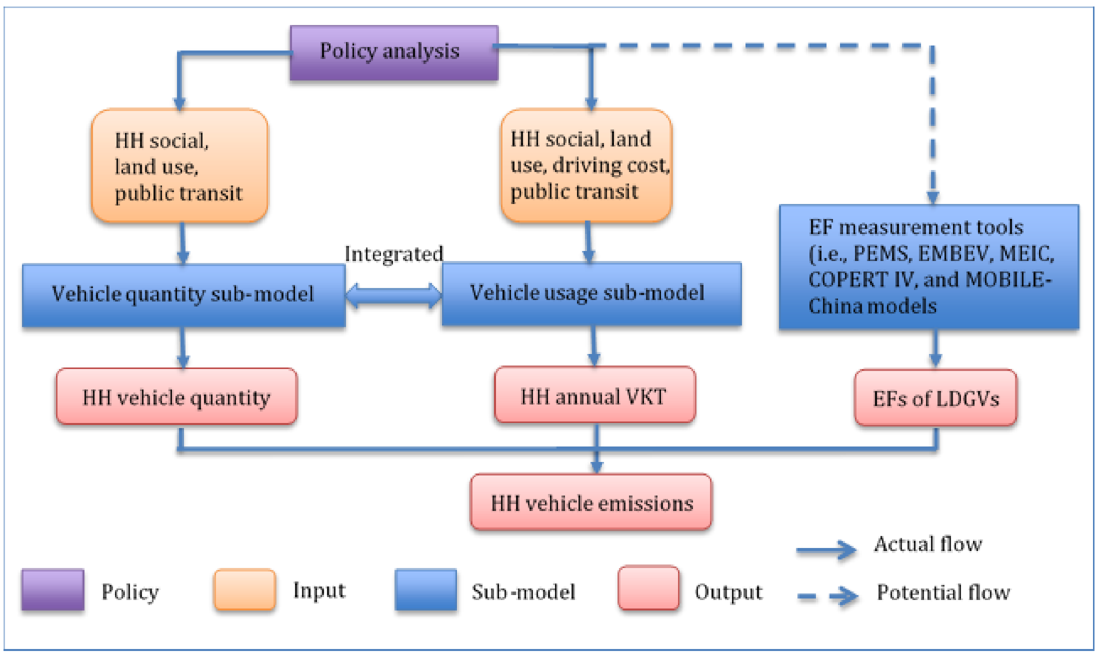
Figure 6.7 shows annual GHG emission reduction rates under the three ownership taxation plans for the two time periods. We can observe that ownership taxes are not effective to reduce GHG emissions, and they have a higher impact on emission reductions in the short run. By implementing annual ownership fee of $3000, the average GHG emission reduction rates are 1.17% in the short run and 0.53% in the medium-long run.

From a policy perspective, the results provide important implications for determining strategies to reduce emissions of private cars. These can be summarized as follows:

- Moderate gasoline tax will effectively lead to an emission reduction by reducing vehicle use. The impact increases with time.
- High ownership tax will lead to a small emission reduction. The impact decreases with time. This type of tax is not effective to reduce vehicle quantity, and it is also not effective to encourage households to choose greener vehicles.
- Despite the financial effect on a household level of a 20% gasoline tax is likely to be less than that of a $3000 ownership fee, the gasoline tax shows a much higher impact on emission reductions especially in a longer run. The finding is consistent with previous studies (Hayashi et al., 2001; Liu and Cirillo, 2016).

6.7 Application 2: Vehicle Ownership, Use, and Emission in Beijing

In this section, a comprehensive framework is proposed to estimate household-level vehicular emissions in Beijing; pollutants considered include carbon monoxide (CO), hydro-carbons (HC), NOx, carbon dioxide (CO2), PM2.5 and PM10. Specifically, the proposed integrated discrete-continuous choice model in previous sections has been transferred to measure vehicle ownership and use; while MEIC (multi-resolution emissions inventory for China) and COPERT IV models (Huo et al., 2015; Wang et al., 2011) have been employed to estimate the aggregated average emission factor (EF) of light-duty gasoline vehicles (LDGVs) in Beijing, China. A flow chart of the modeling structure is given in Figure 6.8.



Note: "HH" means "household", "VKT" means "vehicle kilometer traveled", "EF" means "emission effect", and "LDGV" means "light-duty gasoline vehicle".

Figure 6. 8 A flow chart of proposed modeling framework

The modeling framework integrates three sub-models (in blue): vehicle quantity sub-model, vehicle usage sub-model, and a motor emission simulator to estimate EFs of LDGVs. Data used as the input (in orange) are households' socioeconomics, land use and public transit information, car holding and traveling information, and driving cost. Households' vehicle emissions have been calculated based on outputs (in red) of the three sub-models. The modelling framework is effective to test different policy scenarios and to evaluate their impact on vehicle ownership, use and emission reductions. The modeling framework is estimated on the BHTS data in Section 3.5.

6.7.1 Model Estimation and Validation Results

- *Estimation Results of Integrated Discrete-Continuous Choice Model*

I apply the integrated discrete-continuous choice model to jointly estimate vehicle quantity and vehicle kilometers traveled (VKT) for households' primary and secondary vehicles in Beijing. Primary vehicle is defined as the one used the most by a household. The model is estimated on the sample of 8,540 households living within the 5th Ring road in Beijing. The number of households holding zero, one and two or more vehicles are 4760, 3237 and 523, respectively. Table 6.3 reports the estimation results of the integrated model.

Table 6. 3 Integrated Discrete-Continuous Choice Model: Estimation Results

Variable	Alternative	Coefficient	S.D.	p-value
Constant	1 car	-2.553	0.213	<0.001
	2 cars	-3.404	0.185	<0.001
Number of workers	1 car	0.248	0.030	<0.001
	2 cars	0.326	0.036	<0.001
Annual income (￥10,000)	1 car	0.422	0.025	<0.001
	2 cars	0.516	0.020	<0.001
Gender (Male)	1 car	0.893	0.043	<0.001
	2 cars	0.957	0.042	<0.001
Education level	1 car	0.172	0.013	<0.001
	2 cars	0.170	0.016	<0.001
Private car	1 car	0.439	0.060	<0.001
	2 cars	0.506	0.063	<0.001
Rented house	1 car	-0.673	0.070	<0.001
	2 cars	-0.730	0.077	<0.001
Public transit card	1 car	-1.100	0.082	<0.001

	2 cars	-1.245	0.080	<0.001
Live in traditional district	1 car	-0.167	0.054	0.002
	2 cars	-0.140	0.058	0.015
Live half year or longer	1 car	-0.664	0.175	<0.001
	2 cars	-0.922	0.206	<0.001
Density of bus stop	1 car	-0.074	0.032	0.022
	2 cars	-0.097	0.034	0.004
Density of metro station	1 car	-0.125	0.051	0.014
	2 cars	-0.094	0.054	0.084
Constant	Regression for primary vehicle	1.869	0.169	<0.001
Has bicycle		-0.093	0.028	0.001
Annual income (￥10,000)		0.100	0.015	<0.001
Household head age		-0.002	0.001	0.253
Live half year or longer		-0.069	0.147	0.638
Public transit card		-0.331	0.041	<0.001
Fuel cost per kilometer		-0.283	0.014	<0.001
Density of bus stop		-0.032	0.020	0.120
Density of metro station		-0.074	0.033	0.025
Constant	Regression for secondary vehicle	2.364	0.783	0.003
Has bicycle		-0.008	0.052	0.872
Annual income (￥10,000)		0.011	0.017	0.515
Household head age		-0.010	0.002	<0.001
Live half year or longer		-0.319	0.782	0.683
Public transit card		-0.177	0.071	0.013
Fuel cost per kilometer		-0.131	0.017	<0.001
Density of bus stop		-0.038	0.042	0.364
Density of metro station		-0.055	0.064	0.391
Log-likelihood at zero	-12811.11			
Log-likelihood at convergence	-10306.42			
Number of observations	8540			
R square	0.196			

Note: the model uses bootstrapping re-sampling method to calculate standard deviations.

Assuming "owning zero car" as the base alternative, the estimation results regarding the discrete choice of vehicle quantity are interpreted as follows:

The positive and significant coefficients of households' annual income suggest a positive correlation between income and car ownership as expected. The magnitude of income coefficients indicates households with more vehicles are more sensitive to income, which is probably due to the fact that buying and maintaining cars affect the resources left for other living expenses.

Household demographics are important factors to determine the number of cars held. The positive coefficients for the number of workers indicate that households tend to own more cars if more workers are in the households. This variable has a higher impact on households with more vehicles. The positive coefficients of household head gender suggest that male household heads are more likely to hold one or more cars. The positive coefficients of education level suggest that household heads with a bachelor or higher degree in Beijing tend to hold one or more cars. This could be explained that highly educated household heads usually have higher income and can afford to buy cars.

Households' living condition is also essential to decide the number of cars held. The negative coefficients of renting a living place indicate that households who rent a place tend to have fewer cars. This is reasonable because households who rent are mostly young or low-income, and probably they are not able to afford to buy and to maintain a car. The negative coefficients of public transit card suggest that households possessing a discounted public transit card tend to own fewer vehicles, which is reasonable because they are more likely to use public transit.

Besides, household residential location and public transit accessibility also influence the decision on the quantity of private cars. Households who live in city center (four traditional districts in Beijing: Xicheng, Dongcheng, Xuanwu and Chongwen districts) tend to have fewer cars. This probably because they are closer to public transit/bus stops or because the lack of parking places constitutes a problem for those living in the city center. The negative coefficients of density of bus stop and metro station indicate that households living in an area with higher bus stop or metro station density tend to hold fewer cars given the easy access to public transit. Besides, households with a stable living place (live half year or longer in the same place) are more likely to hold one or more vehicles.

The estimation results regarding the continuous choice of households' primary and secondary annual VKT can be interpreted as follows:

The negative coefficients of "has bicycle" indicate that households owning bicycle or motorcycle tend to drive less. This probably because that they are more likely to use bicycles or motorcycles for short trips. The positive coefficients of annual income indicate that households with higher income tend to drive more, especially with their primary car. Moreover, households with a young head or with a stable living place tend to drive more, while households with a discount public transit card tend to drive less because they are more likely to use public transportation modes. The negative coefficients of fuel cost indicate that households tend to drive less under higher fuel cost as expected. Further, the magnitude of the coefficients illustrates that the usage of primary car is more sensitive to fuel price. Besides, the negative coefficients of bus stop and metro station density suggest that households living in an area with higher density of bus stops or metro stations tend to drive less. It is important to note that some coefficients from the regression of secondary vehicles are not significant due to the small sample size available for this segment of the population.

- *Cross-Sample Validation*

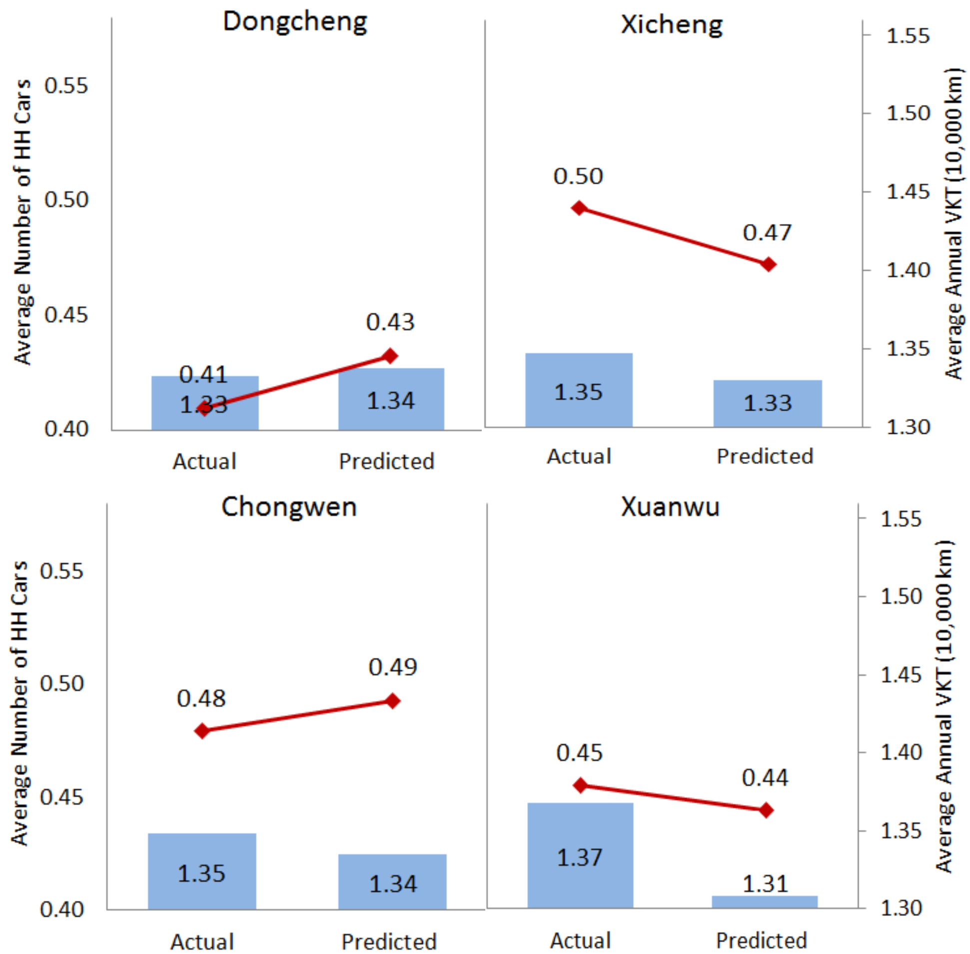
In order to validate the model, I apply the estimation results on an out-of-sample dataset to predict the share of households owning zero, one and two or more cars, and the annual VKT for households' primary and secondary cars. Table 6.4 reports the actual car ownership and VKT, the predicted car ownership and VKT, and their differences.

Table 6. 4 Integrated Discrete-Continuous Choice Model: Model Validation

	Actual	Predicted	Difference
0-car household	55.74%	55.77%	0.03%
1-car household	37.90%	38.14%	0.24%
2-car household	6.36%	6.09%	-0.27%
Average car ownership	**0.51**	**0.50**	**-0.01**
VKT for primary car	1.42	1.33	-6.05%
VKT for secondary car	1.14	1.37	19.96%
Average annual VKT	**1.38**	**1.34**	**-3.44%**

The model is able to accurately reproduce actual choices, but it slightly underestimates car ownership and annual VKT on average. In terms of the discrete choice on the number of cars, the model slightly overestimates the shares of zero-car and one-car households by 0.03% and 0.24% respectively, and underestimates the share of two-car households by 0.27%.

A further validation is necessary to exclude the possibility that on average the model has a good performance but it fails to predict the actual situations in smaller areas. It would also provide an evidence that the model not only can be transferred to large cities but also can be used by small regional governments. Consequently, I compare the actual and predicted average car ownership and VKT in the eight different districts considered. Among these eight districts, Dongcheng, Xicheng, Chongwen and Xuanwu are located in the city center; they are characterized by smaller areas (25.34, 31.62, 16.52 and 18.91 km2) and higher population density (21783.7, 21031.0, 18099.3 and 29243.8 person/km2) (Beijing Municipal Bureau of Land and Resources, 2007) when compared to more peripheral districts such as Fengtan and Shijingshan.



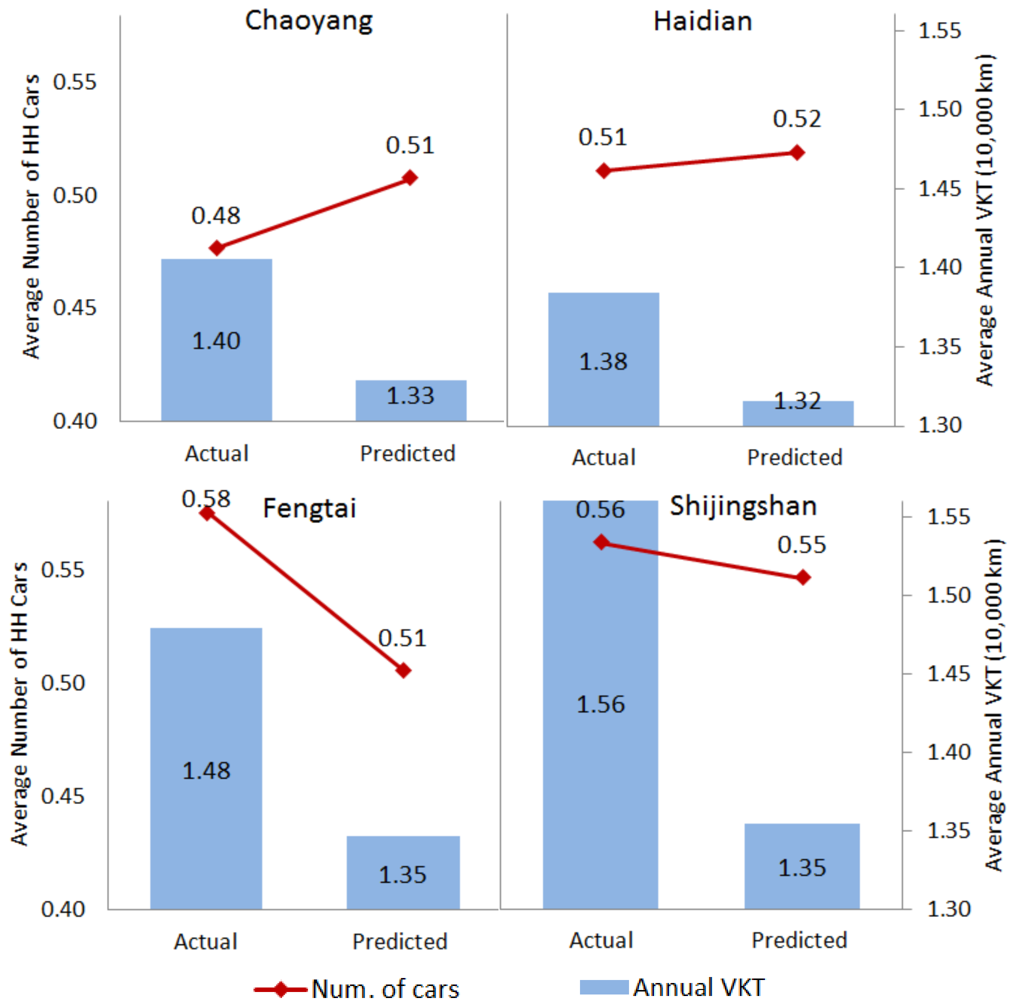


Figure 6. 9 Applications in different districts in Beijing

Generally, the model is able to replicate the actual values across different districts in our study area. In particular, the model shows a very good performance for six out of the eight districts analyzed; precisely the integrated model is able to reproduce both vehicle ownership rates and VKT. The model underestimates households' annual VKT for Fengtai and Shijingshan districts, which are less dense and far from the city center. This could be explained by the fact that Fengtai and Shijingshan districts have more two-car households, and the model does not predict well the continuous choices for two-car households, given the small number of observations in this category.

6.7.2 Vehicular Emissions

A motor emission simulator such as EMBEV, COPERT IV and MOBILE can be used to calculate average EFs. With the knowledge of households' number of cars, annual driving distance of each car and EFs of LDGVs, we can calculate household-level emission rates of CO, HC, NOx, CO2, PM2.5 and PM10 of private vehicles. The emission (Q_{hint}) of pollutant h from the i^{th} car held by household n in area p in year t can be calculated as follows:

$$Q_{hinpt} = EF_{hpt}\left(\frac{grams}{kilometer}\right) \times VKT_{int}(kilometers) \times scale\ factor \qquad (6.7)$$

where EF_{hpt} represents the vehicle emission factor of pollutant h in area p in year t; and VKT_{int} represents vehicle kilometers traveled of the i^{th} car held by household n in year t; a scale factor is used to transfer unit.

The integrated discrete-continuous choice model is estimated on the 2010 BHTS data to predict the number of cars held by households and the annual driving distance of each car (refer to Table 6.3). The average vehicle EFs of different pollutants are derived from MEIC and COPERT IV models (Huo et al., 2015; Wang et al., 2011), based on the emission standards implemented in Beijing. Table 6.5 shows the specific timetable for the application of vehicle emission standards in China, particularly in Beijing, and the European Union (EU).

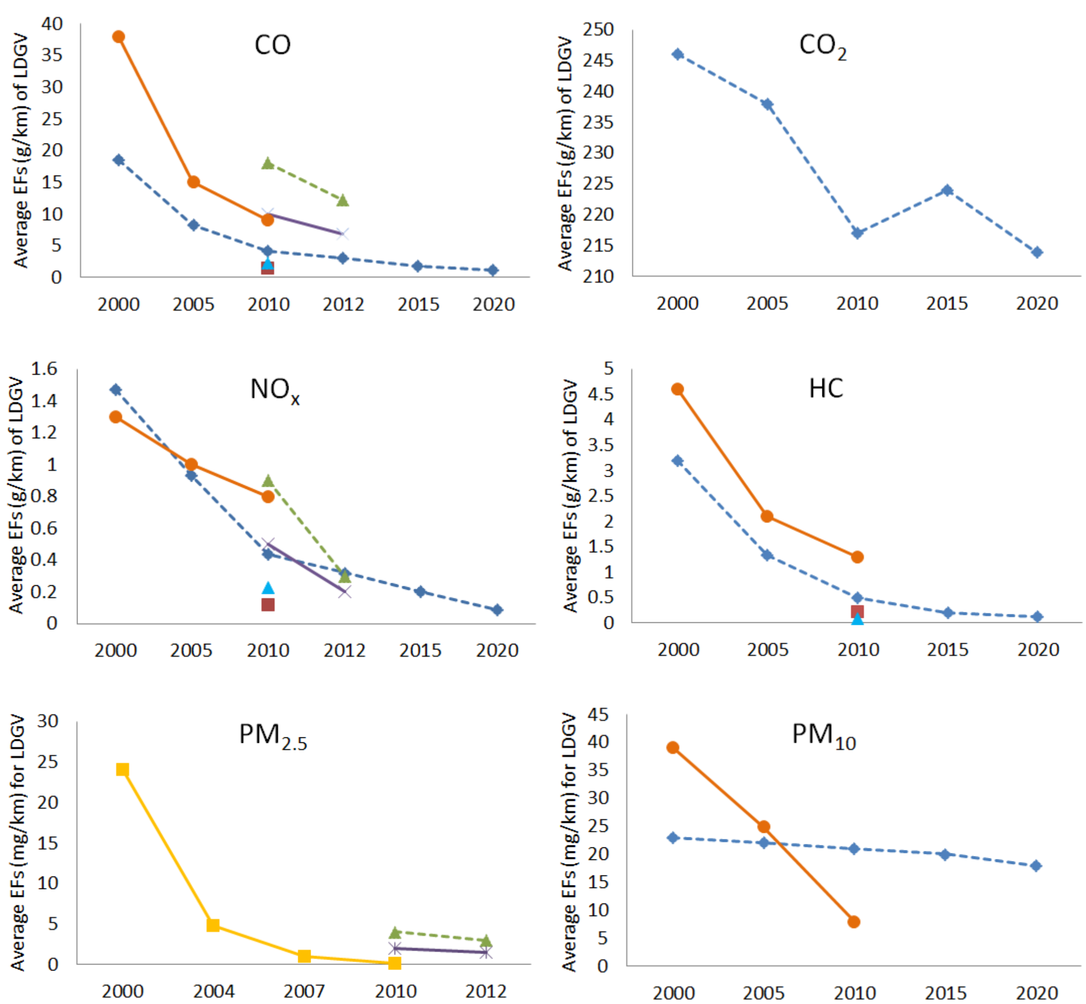
Table 6. 5 Timetable of Vehicle Emission Standards in China and EU

Emission standards	China		EU	Difference between China and EU in year
	Nationwide	Beijing	Nationwide	
Pre Euro 1	1990	1990	1973	17
Euro 1	2000	1999	1992	8

Euro 2	2004	2004	1996	8
Euro 3	2008	2005	2000	8
Euro 4	2011	2008	2005	6
Euro 5	2016	2013	2009	7

Sources: Wang et al., 2011; DieselNet, 2016

It is assumed that the Chinese government have been fully implemented the emission standards according to the schedule in Table 6.5. Euro 3, 4, and 5 was introduced to Beijing in 2005, 2008 and 2013, respectively. Under the emission standards in Beijing and China, I compare the average EFs of LDGVs measured by some recent studies in China with a special focus on Beijing; the values are presented in Figure 6.10.



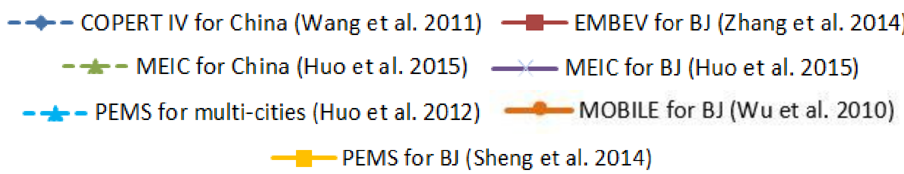


Figure 6. 10 A comparison of the average EFs for LDGVs

Despite the fact that the average EFs may be influenced by variations in calendar year, local features (i.e., city-level or nation-level, fuel quality, temperature and road conditions) and measurement techniques (i.e., PEMS, MEIC, EMBEV, COPERT and MOBILE-China models) (Zhang et al., 2014), all measurement results present a clear decreasing trend in EFs with the improvement of vehicle emission standards in China (i.e., from pre-Euro 1 to Euro 4). Furthermore, in terms of CO, HC and NOx, the EFs of LDGVs in Beijing estimated with MOBILE-China model (Wu et al., 2010) are substantially higher compared to the estimates with EMBEV model (Zhang et al., 2014). To estimate household-level vehicle emissions, we employ the average EFs of private cars at nation-level (China) and city-level (Beijing) from these previous studies (Wang et al., 2011; Huo et al., 2015; Zhang et al., 2014; Shen et al., 2014; Wu et al., 2010); the values obtained and used for our analysis are presented in Table 6.6.

Table 6. 6 The Average EFs of Private Cars in China and Beijing in 2010

(unit: g/km)

g/km	CO	HC	NO_x	CO_2	$PM_{2.5}$	PM_{10}
China	4.13	0.50	0.44	217	0.004*	0.021
Beijing	1.45	0.23	0.12	217*	0.002*	0.008

*Note: * means approximation is applied to the value*

Given the predictions of households' vehicle quantity, use and the average EFs of CO, HC, NOx, CO2, PM2.5 and PM10 in Beijing, I calculate the average annual emissions of households' primary and secondary vehicles based on equation (6.7); the values obtained are shown in Figure 6.11.

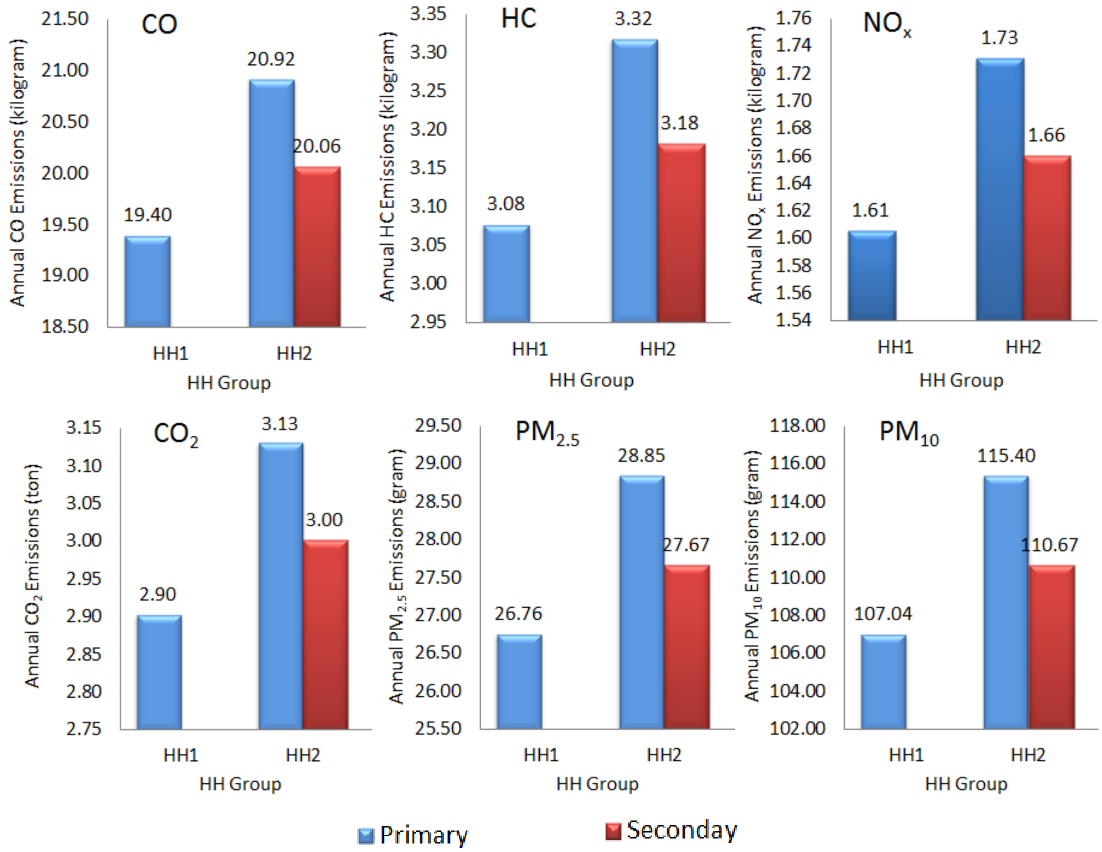


Figure 6. 11 Average annual vehicular emissions

Figure 6.11 shows the average annual emissions of CO, HC, NOx, CO2, PM2.5 and PM10 for each household vehicle. One-car households account for a majority of the population who have cars; the estimated average annual emissions of CO, HC, NOx, CO2, PM2.5 and PM10 for one-car households are 19.40 kilograms, 3.08 kilograms, 1.61 kilograms, 2.90 tons, 26.76 grams and 107.04 grams, respectively. For two-car households, we can observe that primary cars emit more than secondary cars because they are more frequently used in the family. On average, two-car households produce significantly more emissions per vehicle than one-car households, which indicates that two-car households have a higher demand for cars and use them more frequently.

Knowing the average annual vehicular emissions and the number of one-car and two-car households in Beijing in 2010, the total vehicular emissions across the eighteen districts of Beijing can be calculated. Based on 2010 Census data, the total number of households in Beijing is 6,680,552, among which 45% do not have any car and 55% have at least one car. Based on China National Bureau of Statistics (2009), there are approximately 4 million registered private cars at the beginning of 2010. Accordingly, the share (number) of households holding zero, one and two or more cars in Beijing is approximately 45% (3,006,248), 50% (3,348,608) and 5% (325,696). The total emissions of private cars in Beijing in 2010 are shown in Table 6.7; the estimated total emissions of CO, HC, NO_x, CO_2, $PM_{2.5}$ and PM_{10} are 78.31 gigagrams (Gg), 12.42 Gg, 6.48 Gg, 11.72 teragrams (Tg), 108.01 tons and 432.06 tons, respectively.

Table 6. 7 Total Emissions of Private Cars in Beijing in 2010

Beijing 2010	CO (Gg)	HC (Gg)	NO_x (Gg)	CO_2 (Tg)	$PM_{2.5}$ (ton)	PM_{10} (ton)
Total Emissions	78.31	12.42	6.48	11.72	108.01	432.06

6.7.3 Sensitivity Analysis and Policy Implications

The estimation results of the modeling framework in Table 6.3 have been applied to test different policy scenarios: the variables of interest are family income, bicycle ownership, discount public transit card ownership, density of bus stop, availability of metro station and fuel cost. In more details, for each variable, six scenarios are proposed as follows:

- Family income

  Increase or decrease family income by: +20%, +10%, +5%, -5%, -10%, -20%

- Bicycle ownership

  Increase or decrease the percentage of households who have bicycle, electric bicycle or motorcycle by: +20%, +10%, +5%, -5%, -10%, -20%

- Discount public transit card ownership

  Increase or decrease the percentage of households who have discount public transit card by: +8%, +6%, +4%, -4%, -6%, -8%

- Density of bus stops

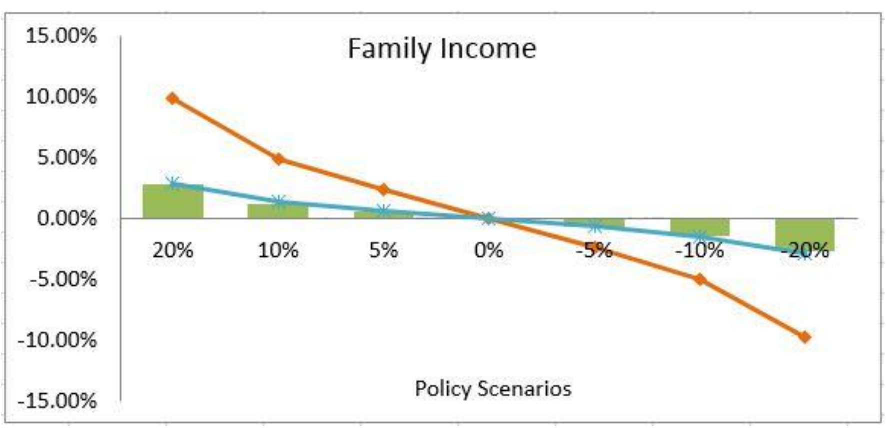
  Increase or decrease bus stop density by: +50%, +25%, +10%, -10%, -25%, +50%

- Availability of metro stations

  Increase or decrease the percentage of households who have access to metro station by: +50%, +25%, +10%, -10%, -25%, +50%

- Fuel cost

  Increase or decrease gasoline price by: +20%, +10%, +5%, -5%, -10%, -20%




Bicycle Ownership
1.50%
1.00%
0.50%
0.00%
-0.50%
-1.00%
-1.50%
20%
10%
5%
0%
-5%
-10%
-20%
Policy Scenarios
Discount PT Card
6.00%
4.00%
2.00%
0.00%
-2.00%
-4.00%
-6.00%
8%
6%
4%
0%
-4%
-6%
-8%
Policy Scenarios
Density of Bus Stop
2.50%
2.00%
1.50%
1.00%
0.50%
0.00%
-0.50%
-1.00%
-1.50%
-2.00%
-2.50%
50%
25%
10%
0%
-10%
-25%
-50%
Policy Analysis
Availbility of Metro Station
0.60%
0.40%
0.20%
0.00%
-0.20%
-0.40%
-0.60%
50%
25%
10%
0%
-10%
-25%
-50%
Policy Scenarios
TotalEmissions
NumOfCar
TotalVKT

Figure 6. 12 Sensitivity analysis

Results in Figure 6.12 shows how the changes in these variables influence households' decisions on the number of cars owned, annual VKT, and the corresponding vehicular emissions at an aggregate level. Overall, variables such as family income, ownership of a discount public transit card and gasoline cost have significant impacts on households' vehicle-related decisions.

Family income is one of the most influential factors on households' vehicle quantity and use decisions, in particular, its changes have large effect on the number of cars held by households. For example, a 20% increase of family income leads to a 9.89% increase in the average number of cars, and a 10% increase leads to a 4.92% increase in the average number of cars. The elasticity of car ownership with respect to family income is approximately 0.5. On the other hand, changes in family income have less effect on households' annual VKT and the vehicular emissions. For instances, a 20% increase of family income leads to a 2.92% and 2.74% increase in the annual VKT and emissions, respectively. Generally, the increase/decrease of family income will encourage/discourage people to have more cars, but will have small influence on their use and emission patterns.

The availability of bicycle or motorcycle in a household has a relatively small impact on their vehicle-related decisions. Specifically, the negative impact on the number of cars is negligible, while the negative impact on the annual VKT and emissions is relatively significant. For example, a 20% increase in the percentage of households who have bicycle or motorcycle leads to a 0.94% decrease in the annual

VKT and emissions. This indicates that owning bicycle or motorcycle has higher impact on households' car use pattern than on the number of cars owned.

To measure the impact from changes of public transit services on households' car ownership and use behavior and eventually on vehicular emissions, three related variables are considered: ownership of a discount public transit card, density of bus stops and availability of metro stations. Among them, owning a discount public transit card affects the most households' car ownership and use decisions. An 8% increase in the percentage of households who have a discount public transit card leads to a 4.81%, 2.74% and 2.54% decrease in households' number of cars, annual VKT and vehicular emissions, respectively. On the other hand, changes in the density of bus stops or the availability of metro stations only slightly influence households' decisions. In particular, a 50% change in the percentage of TAZs that have at least one metro station leads to less than 0.6% changes in the number of cars, annual VKT and vehicular emissions. Changes in the density of bus stops produce a slight greater impact; a 50% increase in the density of bus stops leads to a 1.84%, 0.85% and 0.80% decrease of households' number of cars, annual VKT and emissions, respectively. To summarize, improving public transit services will reduce the number and usage of passenger cars, as well as the emissions; among public transportation related policies, lower public transit fares is found to be an effective way to reduce emissions.

Another effective way to reduce vehicle use and on-road emissions is to regulate fuel price. Although changes in gasoline price have little influence on car ownership, they make a significant impact on households' annual VKT. For instance, a 10% increase in gasoline price leads to a 1.70% decrease in households' driving distance per year. To summarize, the model system is sensitive to policy and can provide valuable references for decision makers.

6.8 Chapter Conclusions

From an economic perspective, the proposed integrated discrete-continuous choice model provides a novel approach for the analysis of discrete and continuous decisions. The model is able to include a large number of alternatives in the discrete choice set, and allows unrestricted correlations of the unobserved factors between the discrete and continuous parts. More specifically, it is able to capture the interdependency of discrete choices such as vehicle holding and type, and continuous choice such as vehicle usage at the household level, by using a full unrestricted variance-covariance matrix. Besides, the model accommodates flexible specifications and no budget constraint in the mileage traveled, which can be applied for policy analysis. For model estimation, an approximation method (Genz, 1992) for multivariate density function is employed to shorten the convergence time.

From an application perspective, the proposed model is general to predict household vehicle ownership, use, and emissions in different metropolitan areas, including developed societies and developing societies. In addition, the model is able to capture the diversity of vehicle market with the consideration of both conventional vehicles and "green" vehicles.

However, the model only captures the indirect correlation between vehicle type and vehicle usage by a "two-step" estimation approach; the estimated coefficients could be insufficient in this case. Besides, the model formulation is static and only provides short-run predictions and short-run policy implications. Moreover, the model does not consider state dependency, future expectation, and market evolution.

Chapter 7: Methodology Part 4: Sequential Discrete-Continuous Choice Model

7.1 Introduction

In recent literature, Glerum et al. (2013) and Gillingham et al. (2015) proposed two different dynamic modeling frameworks to estimate time-series discrete-continuous choices in the context of car ownership and use. Their models not only account for households' joint decisions on vehicle ownership, type, and usage, but also consider dynamics in households' decisions and vehicle market. Specifically, Glerum's model jointly estimates vehicle transaction type, annual distance driven, and fuel type of each household car; the discrete choices are estimated by an optimal stopping formulation, while the continuous choice is optimized with a constant elasticity of substitution (CES) utility. Gillingham's model estimates household vehicle ownership, type, and driving distance in Denmark; a "nested logit" structure is used for discrete choices, while the utility of driving is modeled as a 2^{nd}-order polynomial function of annual kilometers traveled. However, their models have strong limitations on the number of cars held by households, and they are not able to measure the correlation between discrete and continuous choices. For model estimation, they have high computation cost and are difficult to reach a convergence.

This Chapter proposes a sequential discrete-continuous choice model to overcome these limitations. The model system is applied to jointly estimate household vehicle ownership and use over time, with the consideration of forward-looking agents in a finite time horizon. In particular, a recursive probit model is formulated to estimate a sequence of household vehicle holding decisions, while a regression is used to estimate a sequence of household vehicle usage decisions. The inherent Gaussian distributed error components enable the integration between the two parts. The time-dependent correlation is captured with a full unrestricted covariance matrix of the error components.

The proposed model is validated on simulated data sets of car ownership and use choices. Different simulation scenarios are compared to determine appropriate sample size for estimation, including appropriate number of households, length of study time, and households' look-forward time periods. Interesting findings are presented in the following sections. By comparing the true and predicted car ownership market shares and annual vehicle miles driven over time, it is reasonable to summarize that the proposed model is capable to reproduce the evolving trends of households' discrete and continuous demands in a real market.

The following sections present the dynamic formulation to jointly model discrete and continuous choices over time. First, the decision variables and explanatory factors in a car ownership problem setting are described. Then, in Section 7.2, a recursive probit model is proposed to capture a sequence of discrete choices made by individuals. Two situations are considered based on the number of alternatives in the discrete choice set: binary case and multivariate case. In Section 7.3, regression model is introduced to capture individuals' continuous decisions over time. After that, Section 7.4 explains the way to integrate the recursive probit and regression models by introducing correlations between their error components. Section 7.5 describes the maximum likelihood technique to estimate parameters.

Finally, different simulation scenarios are evaluated and compared to identity appropriate properties of data for estimation.

The finite horizon model aims to estimate vehicle holding and driving decisions over time. We consider the situation where a household i, within a set $I = \{1, \ldots, I\}$, has to make choices within finite choice sets J_t at time periods $t = 1, \ldots, T_i$, where T_i is the time horizon for household i and J_t can vary over time. In each time period t, household i will obtain an instantaneous utility u_{ijt} if choose alternative $j \in J_t$. The instantaneous utility can be expressed as follows, using bold font for random variables and normal font for their realizations (this is valid through this Chapter):

$$\boldsymbol{u}_{ijt} = f\left(x_{it}, q_{jt}; \beta_{ij}, \alpha_{ij};\ \boldsymbol{\varepsilon}_{ijt}\right) \tag{7.1}$$

where x_{it} are attributes for household i at time t such as gender, age, income, education and residential location; q_{jt} are attributes for alternative j at time t such as diversity of vehicle types in the market; $\boldsymbol{\varepsilon}_{ijt}$ is a random component which is independently and identically normal distributed over households, alternatives, and time periods; β_{ij} and α_{ij} are the corresponding parameters to be estimated. Although the parameters can vary over households and alternatives, their formats are reduced through this Chapter for simplification purposes, i.e., $\beta_{ij} = \beta$ and $\alpha_{ij} = \alpha$.

The model structure is flexible to consider different number of alternatives over time. In this car ownership estimation, two cases are investigated separately including two alternatives and multiple alternatives in the discrete choice set J_t. In particular, Section 7.2.1 proposes the formulation with two alternatives in the discrete choice set J_t: owning no car and owning at least one car. Section 7.2.2 proposes the formulation with multiple alternatives in the discrete choice set J_t: owning zero, one, two or more cars. If households hold at least one car, their continuous decisions on annual vehicle miles driven (VMD) over time will be further investigated. Specifically, the model considers the discrete choices and continuous decisions simultaneously; the sequence of decision variables D_{it} is shown as follows:

$$D_{it} = \left(Y_{it}^{disc}, Y_{it}^{cont}\right) \tag{7.2}$$

where Y_{it}^{disc} represents the discrete choice of vehicle holdings, and Y_{it}^{cont} is the continuous decision on annual VMD for household i at time t. Based on this problem setting, the following dynamic framework appropriately models households' forward-looking behavior and panel effect in a finite time horizon, with the consideration of substitution pattern among discrete alternatives and correlations between discrete and continuous decision variables.

<u>*7.2 Discrete Choice Sub-Models*</u>

7.2.1 Recursive Binomial Probit Model

This section proposes a recursive binomial probit (RBP) model to capture a sequence of vehicle holding decisions made by households over time, accounting for forward-looking agents. In each time period t, two alternatives are considered: holding no car and at least one car. When an alternative is chosen, household i will obtain an instantaneous utility $\boldsymbol{u}_{ijt}$ and an expected downstream (future) utility $V_{it}^{n}(j)$ associated with the choice j, where n is the forward-looking time periods of agents and can be reduced for simplification purposes (i.e., $V_{it}^{n}(j) = V_{it}(j)$). The instantaneous utility $\boldsymbol{u}_{ijt}$ consists of an observable part v_{ijt} and an error component $\boldsymbol{\varepsilon}_{ijt}$; its specific formulation is described as follows:

$$\boldsymbol{u}_{ijt} = v_{ijt} + \boldsymbol{\varepsilon}_{ijt},\ \boldsymbol{\varepsilon}_{ijt} \sim N(0,\ \mu^2) \quad (7.3)$$

$$v_{ijt} = x_{it}^T \beta + q_{jt}^T \alpha \quad (7.4)$$

where μ is a scale factor usually assumed to be 1; $\boldsymbol{u}_{ijt}$, v_{ijt}, $\boldsymbol{\varepsilon}_{ijt}$, x_{it}, q_{jt}, β and α are defined as above. A household is assumed to make decision in the next time period $t+1$ given the decision in the current period t in a stochastic process having the Markov property (Rust, 1987; Aguirregabiria and Mira, 2010). With known decision at time t, the household observes random utility terms $\varepsilon_{ij't+1}$ at time $t+1$, then he chooses the alternative that maximizes the sum of instantaneous utility $u_{ij't+1}$ and expected downstream utility $\mathrm{V}_{it+1}(j')$ at time $t+1$. The value function of expected downstream utility is defined by taking the continuation of this process into account via the Bellman equation (Bellman, 1957) as follows (Fosgerau et al., 2013):

$$V_{it}(j) = E\left[max_{j' \in J_{t+1}}\left(v_{ij't+1} + \delta \mathrm{V}_{it+1}(j') + \varepsilon_{ij't+1}\right)\right] \quad (7.5)$$

where j' is the alternative chosen from the choice set J_{t+1} at time $t+1$, given that j is the chosen alternative at time t; $\delta \in (0, 1]$ is a discount factor, which is assumed to be 1 for this study. Note that the choice set in the next time period may differ based on the current decision. Generally, it is assumed that households have expectations on the future utility and are able to make decisions that maximize their total utility $\boldsymbol{U}_{it}$ at a generic time t.

$$\boldsymbol{U}_{it} = max_{j \in J_t}(v_{ijt} + \delta V_{it}(j) + \boldsymbol{\varepsilon}_{ijt}) \quad (7.6)$$

Then, the probability of choosing alternative j at time t is given by the binary probit model:

$$P\left(Y_{it}^{disc} = j \mid x_{it}, q_{jt}, \beta, \alpha, \Sigma\right) = \int \mathbb{I}\left(v_{ijt} + \delta V_{it}(j) + \boldsymbol{\varepsilon}_{ijt} > v_{ikt} + \delta V_{it}(k) + \boldsymbol{\varepsilon}_{ikt}, \forall k \in J_t \; and \; k \neq j\right) \varphi(\varepsilon_{it}) d\varepsilon_{it} \quad (7.7)$$

where the indicator function $\mathbb{I}(\cdot)$ ensures that choosing alternative j will obtain a larger utility than any other alternative k; $\varphi(\cdot)$ is the probability density function of a standard normal distribution; Σ is the covariance of the error terms. As there are two alternatives, the dimension of the integral is 2. Since only difference in utility matters, the choice probability can be equivalently expressed in the following form, reducing the dimension of the integral to 1:

$$P(Y_{it}^{disc} = j) = \int \mathbb{I}\left(\tilde{v}_{ijt} + \delta\widetilde{V_{it}}(j) + \tilde{\boldsymbol{\varepsilon}}_{ijt} < 0, \forall k \in J_t \ and \ k \neq j\right)\varphi(\tilde{\varepsilon}_{ijt})d\tilde{\varepsilon}_{ijt} \quad (7.8)$$

where $\tilde{v}_{ijt} = v_{ikt} - v_{ijt}$; $\widetilde{V_{it}}(j) = V_{it}(k) - V_{it}(j)$; $\tilde{\varepsilon}_{ijt} = \varepsilon_{ikt} - \varepsilon_{ijt}$. The difference of two normally distributed error terms follows a normal distribution with a new mean and new variance. In this formulation, alternative j is treated as the base. Then, the likelihood of a sequence of vehicle holding choices from household i over time is:

$$P(\sigma_i) = \prod_{t=1}^{T_i - n} P(Y_{it}^{disc}) \quad (7.9)$$

where $\sigma_i = \{Y_{it}^{disc}\}_{t=1}^{T_i}$ represents the sequence of choices made by household i over time; T_i is the time horizon for household i; n is the number of forward-looking time periods.

To obtain the choice probabilities and to estimate the model, the key point is to figure out how to calculate the expected downstream utility. Here I employ a finite horizon scenario tree to approximate the infinite horizon problem expressed by the Bellman equation. This technique, which is a well-founded approximation approach for multi-period expectations, has a better behavioral rooting because households can only project themselves in a limited time horizon (Cirillo et al., 2015).

In each time period, a household has an expectation over a limited number of future time periods, which is characterized by attributes of alternatives changing over time. As a simple illustration, suppose that starting from a generic time t, the household faces two possible alternatives: owning zero car and at least one car; each of the two scenarios will generate another two possible alternatives at time $t+1$, for a total of four scenarios. This process can be illustrated by a scenario tree shown in Figure 7.1.

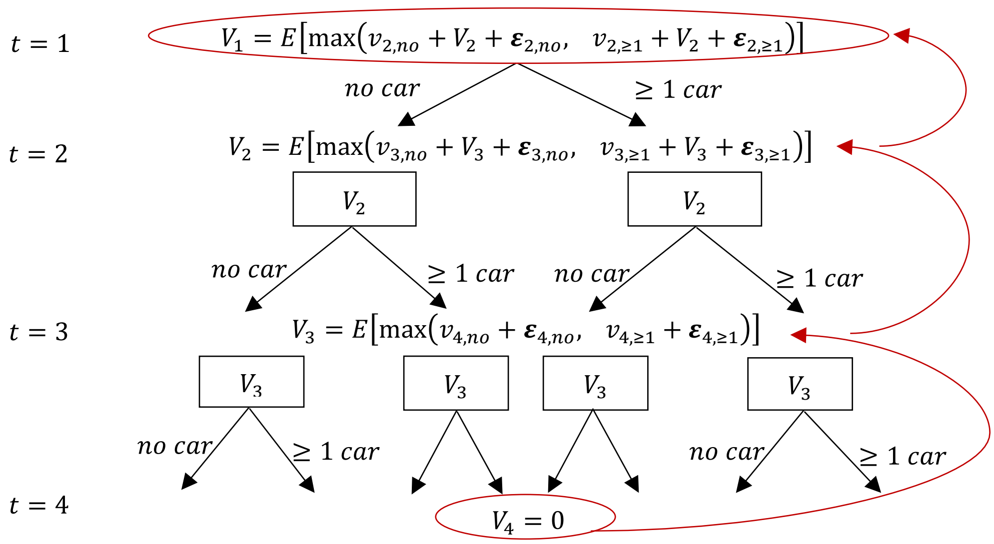


Figure 7. 1 A simple scenario tree

To reduce the number of leaves in the scenario tree, in this simple example, it is assumed that at time 1 the household can anticipate possible alternative attributes for time 2 and 3, but has no knowledge of time 4 or further (i.e., $V_4 = 0$). Therefore, given $V_4 = 0$ we can calculate the value of expected downstream utilities V_t backward from $t = 3$ to $t = 1$. For simplification purposes, households are assumed to have the same number of forward-looking time periods.

The elegance of using scenario-tree technique in the RBP formulation is to avoid building up the dimension of integral in the estimation process. More specifically, the problem of calculating the expected downstream utility ultimately reduces to calculate the expected maximum value of two Gaussian random variables, which has a closed form (Nadarajah and Kotz, 2008). For example, in Figure 7.1, V_1 is the expected maximum value of the two normally distributed random utilities $v_{2,no} + V_2 + \boldsymbol{\varepsilon}_{2,no}$ and $v_{2,\geq 1} + V_2 + \boldsymbol{\varepsilon}_{2,\geq 1}$.

If (X_1, X_2) follows a bivariate Gaussian random vector with mean μ and covariance Σ:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{7.10}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{7.11}$$

where ρ is the correlation coefficient. Then, the expected value of $X = max(X_1, X_2)$ can be calculated as follows (Nadarajah and Kotz, 2008):

$$E(X) = \mu_1 \Phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) + \mu_2 \Phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) + \theta\varphi(\frac{\mu_1 - \mu_2}{\theta}) \tag{7.12}$$

where $\theta = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$; $\Phi(\cdot)$ and $\varphi(\cdot)$ are the cumulative density function and the probability density function of standard normal distribution, respectively.

Similarly, in this problem, the random utilities of the two alternatives j and k at a generic time $t + 1$ follow a bivariate Gaussian distribution:

$$\begin{bmatrix} \boldsymbol{U}_{ijt+1} \\ \boldsymbol{U}_{ikt+1} \end{bmatrix} \sim N\left(\begin{bmatrix} v_{ijt+1} + \delta \mathrm{V}_{it+1}(j) \\ v_{ikt+1} + \delta \mathrm{V}_{it+1}(k) \end{bmatrix}, \begin{bmatrix} \mu^2 & \rho\mu^2 \\ \rho\mu^2 & \mu^2 \end{bmatrix}\right), J_{t+1} = \{j, k\} \tag{7.13}$$

where μ is a scale factor usually assumed to be 1. Thus, the expected downstream utility $V_{it} = E\left[max_{J_{t+1}}\left(\boldsymbol{U}_{ijt+1}, \boldsymbol{U}_{ikt+1}\right)\right]$ at time t can be written as:

$$V_{it} = \left[v_{ijt+1} + \delta \mathrm{V}_{it+1}(j)\right] \Phi\left(\frac{\tilde{v}_{ikt} + \widetilde{\delta V_{it}}(k)}{\theta}\right) + \left[v_{ikt+1} + \delta \mathrm{V}_{it+1}(k)\right] \Phi\left(\frac{\tilde{v}_{ijt} + \widetilde{\delta V_{it}}(j)}{\theta}\right) +$$

$$\theta\varphi(\frac{\tilde{v}_{ikt} + \widetilde{\delta V_{it}}(k)}{\theta}) \tag{7.14}$$

where $\theta = \sqrt{2\ \mu^2 - 2\rho\ \mu^2}$. In this way, the expected downstream utility can be easily calculated without simulation.

7.2.2 Recursive Multinomial Probit Model

This section extends the RBP formulation to capture multiple alternatives in the discrete choice set, namely recursive multinomial probit (RMP) model. All notations in this section are consistent with those in section 7.2.1.

In each time period t, it is assumed that three or more alternatives are considered for vehicle holding decision; for example, holding zero, one, and two or more cars in the case of three alternatives. Similar to the binary case, when an alternative is chosen, household i will obtain an instantaneous utility $\boldsymbol{u}_{ijt}$ and an expected downstream utility $V_{it}(j)$ associated with the choice j. The instantaneous utility $\boldsymbol{u}_{ijt}$ consists of an observable part v_{ijt} and an error component $\boldsymbol{\varepsilon}_{ijt}$; its specific formulation is described by equation (7.3) and (7.4). The expected downstream utility $V_{it}(j)$ has a recursive form via Bellman equation, described by equation (7.5). Consistently, the model bases on the assumption that households have expectations about the future market and are able to make decisions that maximize their total utility at a generic time t. Therefore, the maximum total utility $\boldsymbol{U}_{it}$ can be formulated by equation (7.6). The only difference between the binary and the multinomial case is that the discrete choice set J_t of the multinomial case contains at least three alternatives.

Then, the probability of choosing alternative j at time t is given by the multivariate probit model:

$$P(Y_{it}^{disc} = j) = \iint_{N-1} \mathbb{I}\big(\tilde{v}_{ijt} + \delta\widetilde{V_{it}}(j) + \tilde{\boldsymbol{\varepsilon}}_{ijt} < 0, \forall k \in J_t \ and \ k \neq j\big)\varphi(\tilde{\varepsilon}_{ijt})d\tilde{\varepsilon}_{ijt}$$

(7.15)

where $\tilde{v}_{ijt} = v_{ikt} - v_{ijt}$; $\widetilde{V_{it}}(j) = V_{it}(k) - V_{it}(j)$; $\tilde{\varepsilon}_{ijt} = \varepsilon_{ikt} - \varepsilon_{ijt}$; the indicator function $\mathbb{I}(\cdot)$ ensures that choosing alternative *j* will obtain the largest utility among all alternatives; $\varphi(\cdot)$ is the probability density function of standard normal distribution. Since only differences in utility matter, the choice probability can be expressed as $(N-1)$ – dimensional integrals over the differences between the errors. N is the number of alternatives in the discrete choice set, which is no smaller than 3. In this formulation, *j* is treated as a base alternative. The likelihood of a sequence of vehicle holding choices from household i over time can be described by equation (7.9).

To obtain the likelihood function, the key point is to calculate the expected downstream utility. A finite-horizon scenario tree technique is used to approximate the infinite horizon problem expressed by the Bellman equation (7.5). However, although this technique is well-founded and has a better behavioral rooting, the calculation of expected downstream utility for the multivariate case will build up the dimension of integral in model estimation, which can be explained by a scenario tree in Figure 7.2.

In this simple example, suppose that starting from a generic time t, a household encounters three possible alternatives: owning zero, one, and two or more cars; each of the three scenarios will generate another three possible alternatives at time $t+1$, for a total of nine scenarios. In each time period, the household is assumed to have an expectation over a limited number of future time periods, which is characterized by attributes of alternatives changing over time. For example, at time 1 the household can anticipate possible alternative attributes for time 2 and 3, but has no information of time 4 or further (i.e., $V_4 = 0$). Therefore, given $V_4 = 0$ we can calculate the value of expected downstream utilities V_t backward from time 3 to time 1. For simplification purposes, all households are assumed to have the same forward-looking time periods.

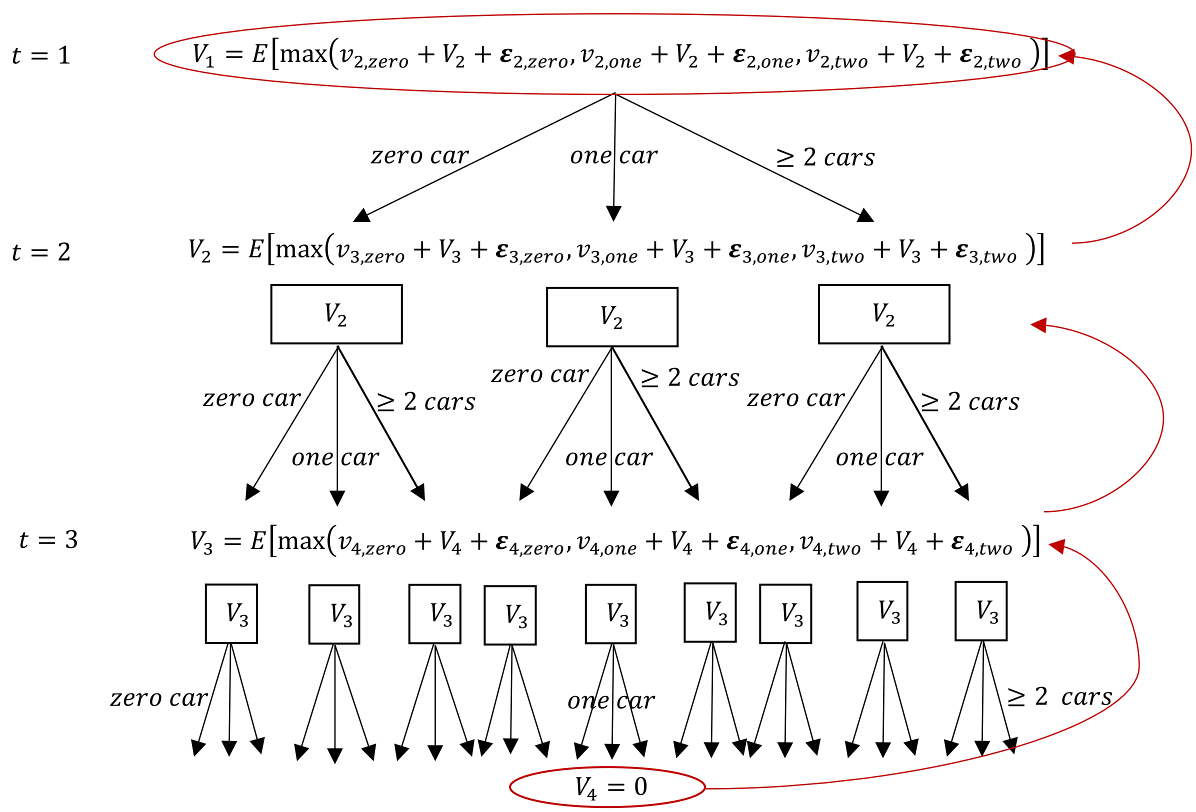


Figure 7. 2 A simple scenario tree for the three-alternative case

Then, the problem of calculating the expected downstream utility for the multivariate case ultimately reduces to calculate the expected maximum value of three or more Gaussian random variables. For example, in Figure 7.2, V_1 is the expected maximum value of the three normally distributed random utilities $v_{2,zero} + V_2 + \boldsymbol{\varepsilon}_{2,zero}$, $v_{2,one} + V_2 + \boldsymbol{\varepsilon}_{2,one}$, and $v_{2,two} + V_2 + \boldsymbol{\varepsilon}_{2,two}$. Unfortunately, to the best of my knowledge, currently there is no closed-form to compute this value so simulation is needed to calculate the expected downstream utility V_t. This limitation can possibly be overcome if a closed-form formulation for the expected maximum value of three or more Gaussian random variables is available in the future.

7.2.3 Logsum Approximation

To reduce the computation cost of RMP model estimation, an approximation method is proposed to replace the simulation process of the expected maximum value. In particular, a logsum formulation is used to approximate the expected maximum value of three or more Gaussian random variables. The formulations of the two methods, simulation and logsum, are explained here. Their calculated values are compared with Q-Q plots.

- *Method of Simulation*

Let $s = (s_1, s_2, \dots, s_J)$ follow a multivariate normal distribution with mean m and covariance n:

$$m = \begin{bmatrix} m_1 \\ m_2 \\ \dots \\ m_J \end{bmatrix} \tag{7.16}$$

$$n = \begin{bmatrix} n_1^2 & n_{12} & \dots & n_{1J} \\ n_{21} & n_2^2 & \dots & n_{2J} \\ \dots & \dots & \dots & \dots \\ n_{J1} & n_{J2} & \dots & n_J^2 \end{bmatrix} \tag{7.17}$$

In the car ownership problem, s can be considered as random utilities and J is the number of alternatives in the vehicle holding choice set. In the simulation experiment, the mean of the multivariate normal distribution is assumed to be randomly generated from a uniform distribution with range (R_m^-, R_m^+). The elements of the covariance matrix are also randomly generated from some uniform distributions with range (R_n^-, R_n^+). The simulation process has two stages. In the first stage, the mean and covariance are simulated for A times based on the given range (R_m^-, R_m^+) and (R_n^-, R_n^+). In the second stage, I use each pair of simulated mean and covariance matrix to generate multivariate normally distributed draws for B times.

Then, the expected maximum value $\hat{S}$ given certain mean $m^{(a)}$ and covariance $n^{(a)}$ can be expressed as follows:

$$\hat{S}^{(a)} = \frac{1}{B}\sum_{b=1}^{B} max_{s_j}\{s_1^{(b)}, s_2^{(b)}, \dots, s_J^{(b)} \mid m^{(a)}, n^{(a)}\} \tag{7.18}$$

where $s_1^{(b)}, s_2^{(b)}, \dots, s_J^{(b)}$ are generated draws from a multivariate normal distribution with mean $m^{(a)}$ and covariance matrix $n^{(a)}$; $m^{(a)}$ and $n^{(a)}$ are simulated from uniform distributions. Given a certain range of mean values, $(A =)500$ pairs of mean and covariance are simulated to explore as many situations as possible within this range. For each simulated pair of mean and covariance matrix, the expected maximum value can be expressed as the average maximum value of $(B =)1000$ Monte Carlo simulations.

- *Method of Logsum*

Using the same notations as above, the logsum can be expressed as follows:

$$L = ln \sum_{j=1}^{J} \exp(m_j) \tag{7.19}$$

Similarly, elements of the mean $(m_1, m_2, \dots, m_J)$ are simulated from a uniform distribution with a given range (R_m^-, R_m^+); A is the number of simulations which is 500 in this study. Then, the logsum for each sequence of simulated draws can be calculated as follows:

$$L^{(a)} = ln \sum_{j=1}^{J} \exp(m_j^{(a)}) \tag{7.20}$$

- *Simulation Experiments: Comparison of the Two Methods with Q-Q Plots*

Different simulation scenarios are investigated to compare the two methods; variables of interested are as follows:

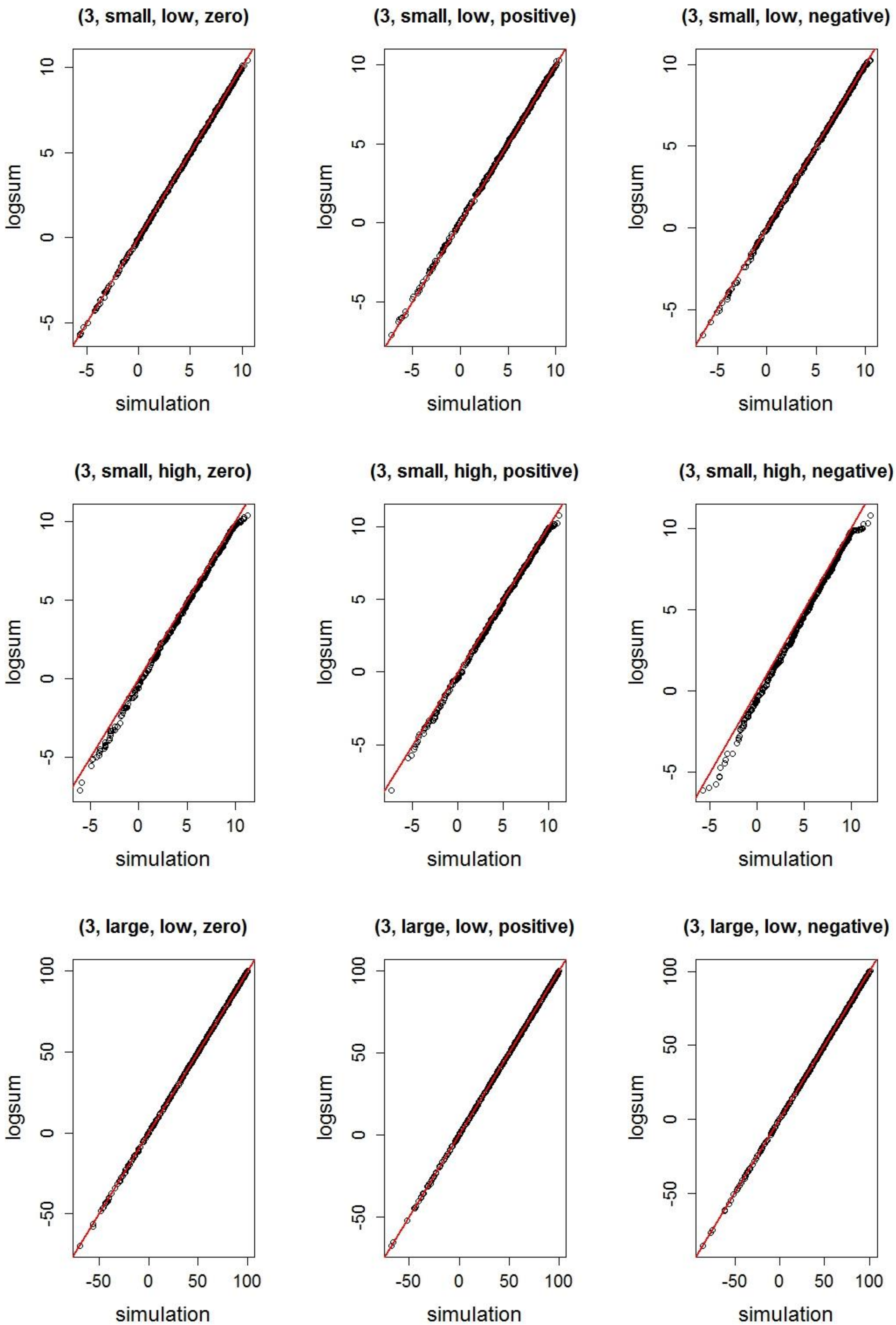
- ✓ The dimension of multivariate normal distribution: 3, 4, 5
- ✓ The range of mean: (-10, 10), (-100, 100)
- ✓ The range of variance: 10% of the range of mean, 50% of the range of mean
- ✓ The setting of covariance: zero, positive, negative

It should be noted that the dimension of multivariate normal distribution is equivalent to the number of discrete alternatives in car ownership problem setting. Table 7.1 summarizes all possible simulation scenarios. For each scenario, the expected maximum value of the multivariate normal distribution is estimated with the *simulation method* and the *logsum method*; their values are compared with the Q-Q plots in Figure 7.3 – 7.5.

Table 7. 1 Summary of Simulation Scenarios: Multivariate Normal Distribution

Scenario ID	Dimension	Range of mean	Range of variance	Covariance
1	3	(-10, 10) small	(1, 3) low	Zero
2		(-10, 10) small	(1, 3) low	Positive
3		(-10, 10) small	(1, 3) low	Negative
4		(-10, 10) small	(1, 11) high	Zero
5		(-10, 10) small	(1, 11) high	Positive
6		(-10, 10) small	(1, 11) high	Negative
7		(-100, 100) large	(1, 21) low	Zero
8		(-100, 100) large	(1, 21) low	Positive
9		(-100, 100) large	(1, 21) low	Negative

10		(-100, 100) large	(1, 101) high	Zero
11		(-100, 100) large	(1, 101) high	Positive
12		(-100, 100) large	(1, 101) high	Negative
13	4	(-10, 10) small	(1, 3) low	Zero
14		(-10, 10) small	(1, 3) low	Positive
15		(-10, 10) small	(1, 3) low	Negative
16		(-10, 10) small	(1, 11) high	Zero
17		(-10, 10) small	(1, 11) high	Positive
18		(-10, 10) small	(1, 11) high	Negative
19		(-100, 100) large	(1, 21) low	Zero
20		(-100, 100) large	(1, 21) low	Positive
21		(-100, 100) large	(1, 21) low	Negative
22		(-100, 100) large	(1, 101) high	Zero
23		(-100, 100) large	(1, 101) high	Positive
24		(-100, 100) large	(1, 101) high	Negative
25	5	(-10, 10) small	(1, 3) low	Zero
26		(-10, 10) small	(1, 3) low	Positive
27		(-10, 10) small	(1, 3) low	Negative
28		(-10, 10) small	(1, 11) high	Zero
29		(-10, 10) small	(1, 11) high	Positive
30		(-10, 10) small	(1, 11) high	Negative
31		(-100, 100) large	(1, 21) low	Zero
32		(-100, 100) large	(1, 21) low	Positive
33		(-100, 100) large	(1, 21) low	Negative
34		(-100, 100) large	(1, 101) high	Zero
35		(-100, 100) large	(1, 101) high	Positive
36		(-100, 100) large	(1, 101) high	Negative


(3, small, low, zero)
(3, small, low, positive)
(3, small, low, negative)
(3, small, high, zero)
(3, small, high, positive)
(3, small, high, negative)
(3, large, low, zero)
(3, large, low, positive)
(3, large, low, negative)
logsum
simulation

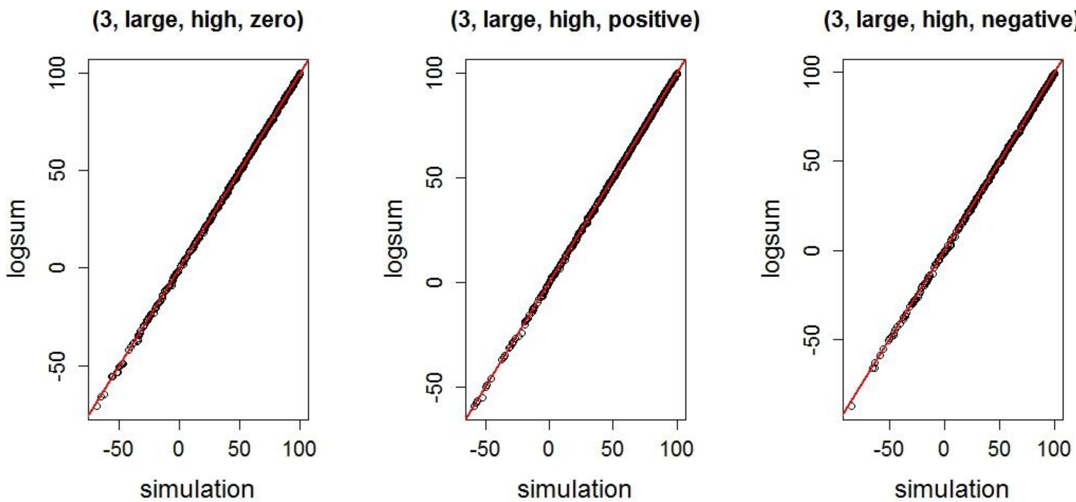
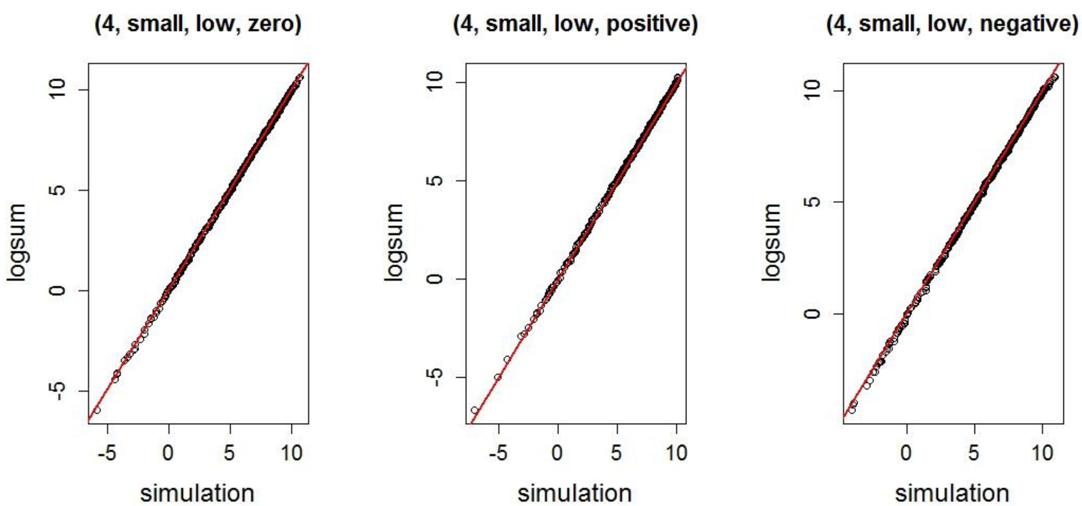


Figure 7. 3 Q-Q plots for scenario 1 – 12 (dimension = 3)



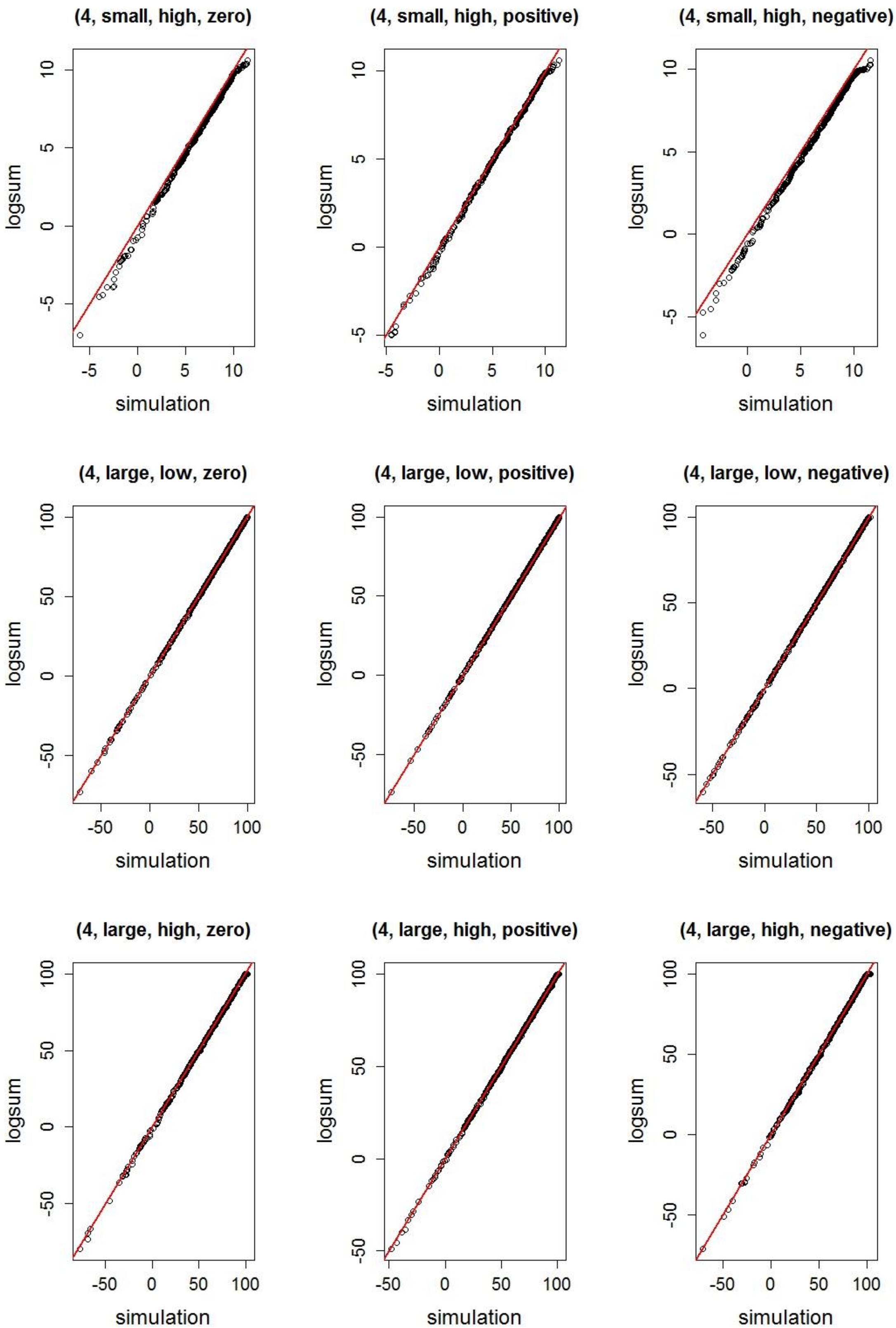
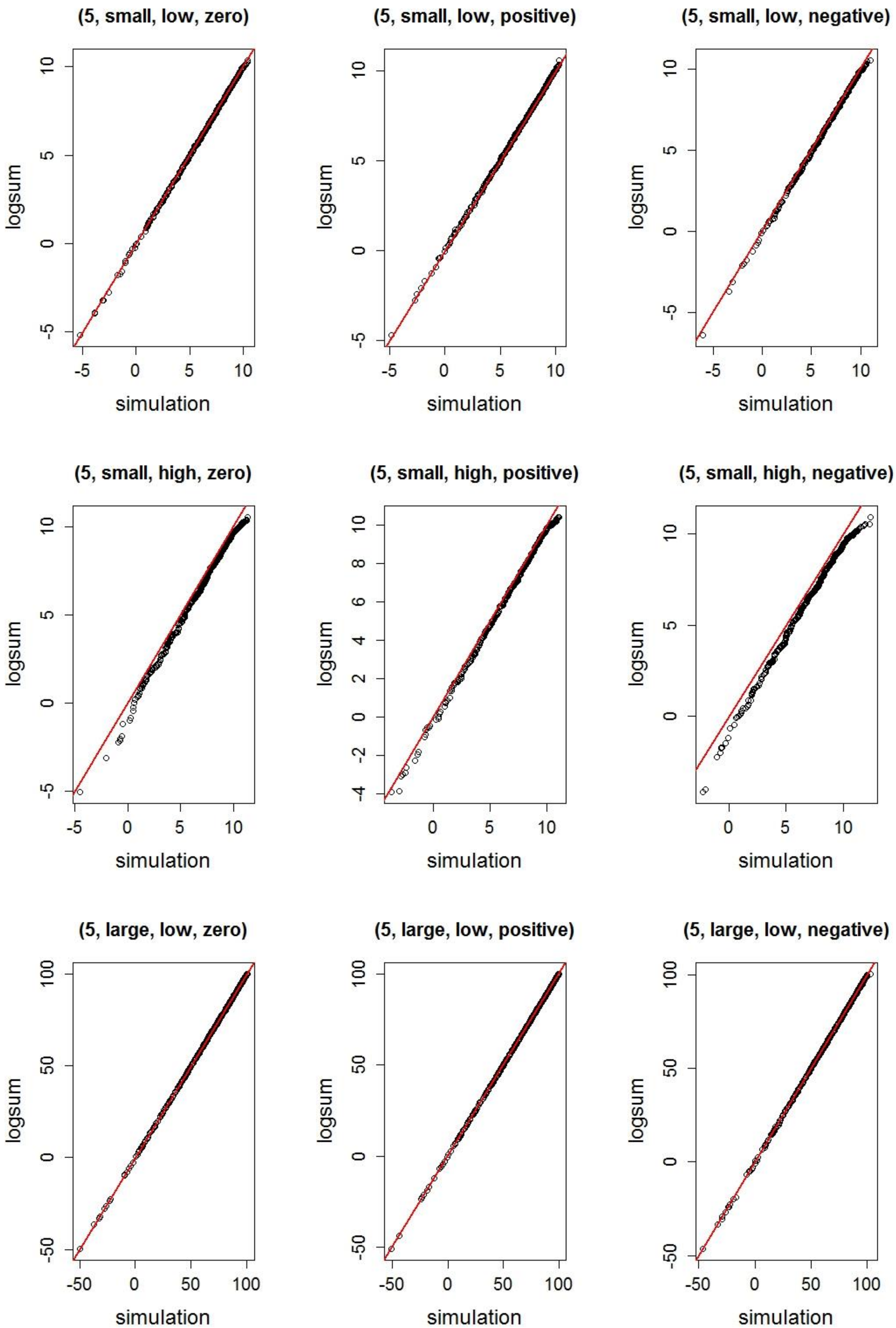


Figure 7. 4 Q-Q plots for scenario 13 – 24 (dimension = 4)


(5, small, low, zero)
(5, small, low, positive)
(5, small, low, negative)
(5, small, high, zero)
(5, small, high, positive)
(5, small, high, negative)
(5, large, low, zero)
(5, large, low, positive)
(5, large, low, negative)
logsum
simulation

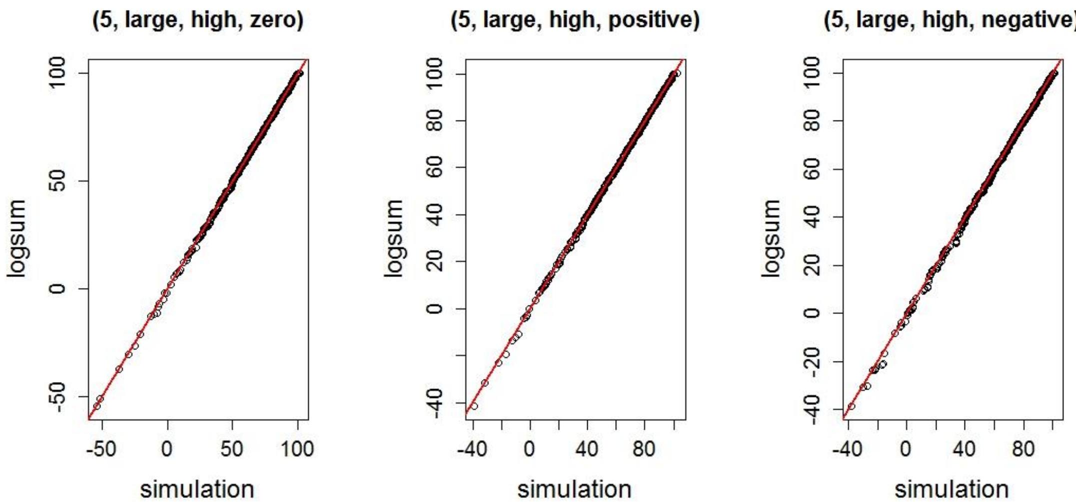


Figure 7. 5 Q-Q plots for scenario 25 – 36 (dimension = 5)

By observing the Q-Q plots in Figure 7.3 - 7.5, it is reasonable to summarize that in most cases the method of logsum perfectly approximates the expected maximum value of three or more Gaussian random variables. The accuracy of the approximation mainly depends on the range of variance, the value of covariance, and the number of variables (dimension). Specifically, the method of logsum creates higher bias in approximation when the range of variance is larger; the situation become worse when the range of mean is smaller or dimension is higher. Another important observation is that the method of logsum performs better approximation when the covariance is zero or positive; the worst case is the combination of high dimension, small range of mean, large variance, and negative covariance (i.e., scenario 30).

7.3 Continuous Choice Sub-Model: Regression

Regression models are employed to capture the time-series continuous decisions on annual VMD for households owning at least one car. In particular, one regression is used to estimate households' annual VMD at each time period. Although the error components of the regressions over time can be correlated, they are assumed to be independent for simplicity. In each regression model, the dependent variable $\boldsymbol{y}_{it}^{cont}$ is assumed to be a linear combination of a vector of predictors x_{it}^{cont} and an error term $\boldsymbol{\varepsilon}_{it}^{cont}$:

$$\boldsymbol{y}_{it}^{cont} = (x_{it}^{cont})^T \beta^{cont} + \boldsymbol{\varepsilon}_{it}^{cont},\ \boldsymbol{\varepsilon}_{it}^{cont} \sim N(0, \tau^2) \tag{7.21}$$

where β^{cont} is a vector of parameters to be estimated; τ is a scale factor. In order to integrate the continuous part with the RBP model, the regressions are solved by the maximum likelihood estimator instead of the ordinary least square estimator (Liu et al. 2014). The probability of observing certain miles y_{it}^{cont} at time t equals to the normal density that is centered at $(x_{it}^{cont})^T \beta^{cont}$ and has variance τ^2:

$$P(y_{it}^{cont} | x_{it}^{cont}, \beta^{cont}, \tau^2) = \varphi(y_{it}^{cont} | (x_{it}^{cont})^T \beta^{cont}, \tau^2) \tag{7.22}$$

Then, the likelihood of a sequence of continuous decisions on annual VMD from household i over time is:

$$P(\sigma_i^{cont}) = \prod_{t=1}^{T_i} P(Y_{it}^{cont}) \tag{7.23}$$

where $\sigma_i^{cont} = \{Y_{it}^{cont}\}_{t=1}^{T_i}$ represents the sequence of continuous choices made by household i from time 1 to T_i.

7.4 Integration of Discrete and Continuous Choice Model

To jointly estimate the discrete choices Y_{it}^{disc} and the continuous decisions Y_{it}^{cont}over time, it is essential to capture the correlation between them. In particular, we allow the error components of the regressions and the recursive probit model to be correlated by introducing a flexible covariance in each time period t. It's important to note that the differences of the error terms from the recursive probit model are used to guarantee the estimated parameters are identified. Thus, the integrated discrete-continuous error components ($\tilde{\boldsymbol{\varepsilon}}_{ijt}, \boldsymbol{\varepsilon}_{it}^{cont}$) is assumed to follow a new multivariate normal (MVN) distribution. In this section, subscripts i and j are omitted for simplicity.

$$(\tilde{\boldsymbol{\varepsilon}}_t, \boldsymbol{\varepsilon}_t^{cont}) \sim MVN(0, \Sigma_{N,t}) \tag{7.24}$$

where $\Sigma_{N,t}$ is a $N \times N$ covariance matrix at time t; the dimension of the MVN distribution equals: N (number of discrete alternatives) – 1 (for normalization) + 1 (the dimension of regression). For example, in the case of binary discrete choices, the dimension equals 2.

Then, the joint probability of observing Y_t^{disc} and Y_t^{cont} can be derived as the product of the marginal probability of observing the continuous choice Y_t^{cont} and the probability of observing the discrete choice Y_t^{disc} conditional on Y_t^{cont} (Liu et al., 2014):

$$P(D_{it}) = P\left(Y_{it}^{disc}, Y_t^{cont}\right) = P(Y_t^{cont})P\left(Y_{it}^{disc}\middle|Y_t^{cont}\right) = \varphi(\boldsymbol{\varepsilon}_t^{cont})\varphi(\tilde{\boldsymbol{\varepsilon}}_t|\boldsymbol{\varepsilon}_t^{cont}) \tag{7.25}$$

From equation (7.23), we know the marginal probability is given by a normal density function. The conditional probability can also be derived from a normal density function with a new mean and new variance.

In multinomial normal distribution, if $\begin{bmatrix}\boldsymbol{A}\\ \boldsymbol{B}\end{bmatrix}$ follow a multivariate normal distribution with mean $\boldsymbol{\mu} = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$ and variance $\boldsymbol{\Sigma} = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{22} & \Sigma_{21}\end{bmatrix}$, then $(\boldsymbol{A}|\boldsymbol{B} = B_1)$ follows a multivariate normal distribution with new mean and new variance as follows:

$$\boldsymbol{\mu}_{\boldsymbol{A}|\boldsymbol{B}} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(B_1 - \mu_2) \tag{7.26}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{A}|\boldsymbol{B}} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{7.27}$$

Similarly, in our problem:

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}}_t \\ \boldsymbol{\varepsilon}_t^{cont} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{disc} & \Sigma_{disc,reg} \\ \Sigma_{reg,disc} & \tau^2 \end{bmatrix}\right) \tag{7.28}$$

Thus, the conditional term $(\tilde{\boldsymbol{\varepsilon}}_t|\boldsymbol{\varepsilon}_t^{cont} = err)$ follows a multivariate normal distribution with mean and variance as follows:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{\varepsilon}}_t|\boldsymbol{\varepsilon}_t^{cont}} = 0 + \frac{\Sigma_{disc,reg}}{\tau^2}(err - 0) \tag{7.29}$$

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\varepsilon}}_t|\boldsymbol{\varepsilon}_t^{cont}} = \Sigma_{disc} - \frac{\Sigma_{disc,reg}\Sigma_{reg,disc}}{\tau^2} \tag{7.30}$$

To further improve the flexibility of the model, different covariance matrices of the integrated error terms are allowed for different time periods. Therefore, the dimension of the covariance matrix of the errors is expanded to $NT \times NT$, where T is the number of total time periods. For example, in the case of binary discrete choices, the dimension is $2T \times 2T$ and the covariance matrix therefore takes the following form:

$$\Omega = \begin{bmatrix} \Sigma_{2,t=1} & 0 & \cdots & \cdots & 0 \\ 0 & \Sigma_{2,t=2} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & \Sigma_{2,t=T} \end{bmatrix} \tag{7.31}$$

Here, $\Sigma_{2,t}$ is defined as above. For identification purposes, the first diagonal element of $\Sigma_{2,t}$ is fixed to 2. So, only two parameters are estimated at each time t: one describes the correlation between the discrete and continuous choices and the other is the variance of errors for the regression, with a total of $2T$ parameters to be estimated. The structure of covariance matrix can be easily extended for the case of multivariate discrete choices by increasing the dimension. In this study, the covariance between different time periods is assumed to be zero. However, this assumption can be relaxed in future research.

7.5 Estimation Process

The maximum likelihood technique is employed to estimate the dynamic integrated discrete-continuous choice model; the parameters to be estimated in the car ownership problem setting are summarized as follows:

- β, a vector of parameters related to households' characteristics and land use attributes that influence vehicle holding (discrete) decisions;
- α, a vector of parameters related to alternative-specific attributes that influence vehicle holding (discrete) decisions;
- β^{cont}, a vector of parameters related to households' characteristics, land use, and market economic indicators that influence vehicle use (continuous) decisions;
- Ω, a normalized covariance matrix of the integrated error components over time;
- δ, a discount factor, setting to 1 for simplicity.

The likelihood function to estimate these parameters is proposed as follows:

$$L(\beta, \alpha, \beta^{cont}, \Omega, \delta) = \prod_{i=1}^{I} \prod_{t=1}^{T_i} P_{it}(D_{it} | \, \beta, \alpha, \beta^{cont}, \Omega, \delta) \qquad (7.32)$$

The probabilities are derived with simulation because the discrete part $\tilde{\boldsymbol{\varepsilon}}_t | \boldsymbol{\varepsilon}_t^{cont}$ has no closed form. In this study, simulations are executed using 1000 pseudo Monte Carlo draws to ensure the efficiency of the estimation.

7.6 Experiment with Simulated Data – Bivariate Discrete Choice

This section simulates households' choices on car ownership and use over time to validate the proposed dynamic discrete-continuous choice model. Households are assumed to provide discrete choices on car holding and continuous choices on annual VMD. In particular, two alternatives are available for the discrete part: owning zero car and owning at least one car. The discrete choice for each time period is generated based on probabilities of different alternatives obtained from the following utility functions:

$$U_0 = \beta_1 X_1 + \beta_2 X_2 + \delta V_0 + \varepsilon_0 \tag{7.33}$$

$$U_{\geq 1} = \beta_1 X_3 + \beta_3 X_4 + \beta_4 X_5 + \beta_5 X_6 + \delta V_{\geq 1} + \varepsilon_{\geq 1} \tag{7.34}$$

where the discount factor δ is assumed to be 1. Households are assumed to be rational and make decisions to maximize their utility. In addition, households are assumed to have expectations about the future alternatives in the market.

Meanwhile, the continuous choice for each time period is generated by a regression as follows:

$$Y = \beta_6 X_7 + \beta_7 X_8 + \varepsilon_{cont} \tag{7.35}$$

It is important to note that the unobserved error components of discrete and continuous parts are simulated with predetermined correlations varying over time.

In the following simulation experiment, 13 scenarios are proposed considering different numbers of households, length of study time, and look-forward time periods to test the accuracy of estimated parameters under different situations. All scenarios have the same model specification, which considers eight predictors: $X_1, X_2, \ldots, X_8$; they are assumed to be independent from each other. In the car ownership problem setting, these variables have been generated with specific meaning; the corresponding criteria are described as follows:

- Personal and household characteristics, i.e. age, gender, education level, income, number of family members, workers, adults, and children. These variables usually have positive sign and can either be categorical or continuous. Some variables are static (i.e. gender) while others change over time (i.e. age, education). In the example, X_5 is a vector of constant which possibly represents static characteristics such as gender, X_6 follows a truncated normal distribution possibly representing a dynamic continuous variable such as income.
- Land use variables and accessibility to alternative travel modes, i.e. residential location, residential density, distance to nearest public transit station, coverage of public transit, access to non-motorized infrastructure. In the example, X_7 follows a normal distribution with the mean changing over time, which possibly represents residential density (X_5 and X_6 can also belong to this category).
- Vehicle characteristics and diversity of vehicle types in the market. These variables are alternative specific; the corresponding coefficient can either be generic (identical for all alternatives) or specific (different among alternatives). In the example, X_1, X_2, X_3, and X_4, which follow uniform distributions, could possibly belong to this category; X_1 and X_3 share a generic coefficient β_1, while X_2 and X_4 have alternative-specific coefficients β_2 and β_3.

- Fuel price and driving cost, i.e. gasoline price. These variables usually evolve over time. In this example, X_8 follows a truncated normal distribution with mean changing over time, which possibly represents driving cost per mile.

7.6.1 Model Estimation

Given the model specification, the proposed model is estimated on simulated datasets using a self-developed R package. To determine appropriate numbers of households, length of study time, and look-forward time periods, Table 7.2 categorizes 13 scenarios into four groups for comparison purpose. For example, the synthetic sample in the "base" scenario is composed of 600 households; each of them is assumed to provide responses over 15 time periods and their look-forward time is 3.

Table 7. 2 Summary of Simulation Scenarios

Group No.	Description	No. of households	Length of study time (T)	Look-forward time periods (n)	T-n	Scenario No.
1	Comparison with No. of households	**200**	15	3	12	1
		400	15	3	12	2
		600	15	3	12	base
		800	15	3	12	3
2	Comparison with (T-n); Fix n	600	5	3	**2**	4
		600	10	3	**7**	5
		600	15	3	**12**	base
		600	20	3	**17**	6
3	Comparison with (T-n); Fix T	600	15	1	**14**	7
		600	15	3	**12**	base
		600	15	5	**10**	8
		600	15	7	**8**	9
4	Comparison with T; Fix (T-n)	600	**8**	**1**	7	10
		600	**10**	**3**	7	5
		600	**12**	**5**	7	11
		600	**14**	**7**	7	12

Table 7.3 – 7.6 summarize the estimation results for each group separately, and compare the true and estimated parameters. It should be noted that each scenario contains 10 simulated datasets, and the reported results in Table 7.3 – 7.6 are the average values based on the 10 simulations.

Overall, the estimated parameters are approaching the true values for both discrete and continuous parts of the model in all 13 scenarios. The values of $\boldsymbol{R^2}$ and adjusted $\boldsymbol{R^2}$ show that the log-likelihood has been highly improved by using the proposed model for estimation. Further, the values of root-mean square error (RMSE) are plotted in Figure 7.6 – 7.9, which provide essential evidence to determine the appropriate numbers of households and time periods (including study time periods, look-forward time periods, and their difference).

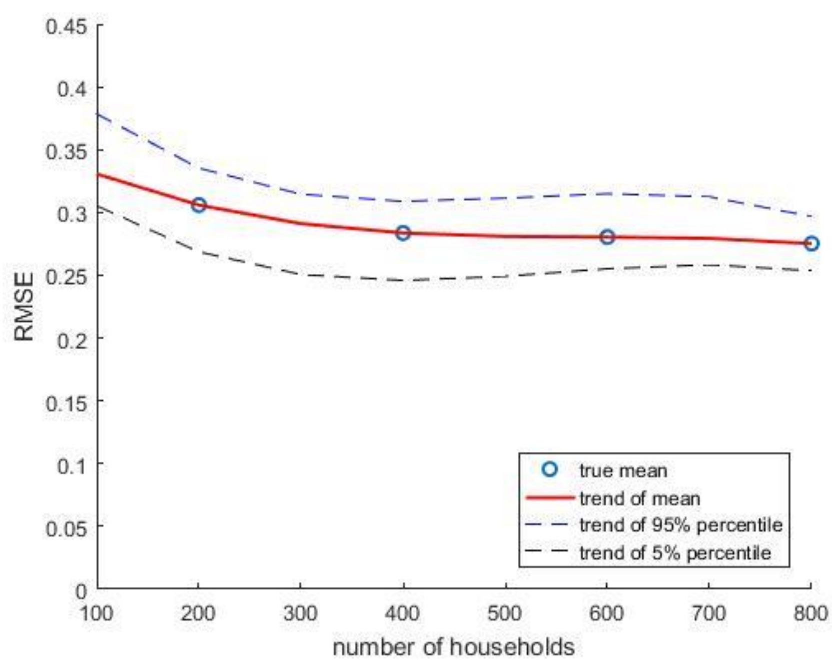
Table 7.3 compares the estimation results for scenarios with different number of households (in group 1). Fixing the numbers of study time periods (T) and look-forward time periods (n), the influence of household number on the RMSE of parameters are evaluated. The RMSE decreases as household number increases; in other words, increasing the number of households improves the accuracy of the estimated parameters.

Table 7. 3 Result Comparison: Simulation Scenarios with Different Number of Households

Coefficients	zero car	at least one car	True $\boldsymbol{\beta}$	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 1	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 2	Estimated $\boldsymbol{\beta}$ (t-statistics) Base	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 3
$\boldsymbol{\beta_1}$	X	X	1.00	0.95 (10.5)	1.00 (15.5)	0.97 (18.4)	1.00 (21.9)
$\boldsymbol{\beta_2}$	X		-0.50	-0.49 (-7.0)	-0.50 (-10.2)	-0.48 (-11.8)	-0.49 (-13.9)

β_3	X	-2.00	-1.78 (-6.7)	-2.02 (-10.6)	-1.92 (-12.3)	-1.98 (-14.6)
β_4	X	-3.00	-2.96 (-25.3)	-2.94 (-36.3)	-2.95 (-44.6)	-2.97 (-51.7)
β_5	X	1.00	0.93 (12.4)	0.97 (18.6)	0.98 (22.6)	0.99 (26.4)
β_6		1.00	1.00 (52.7)	1.00 (75.3)	1.00 (92.4)	1.00 (106.1)
β_7		0.50	0.50 (34.1)	0.50 (48.4)	0.50 (59.5)	0.50 (69.0)
Households no.			200	400	600	800
Study Time T			15	15	15	15
Look-forward n			3	3	3	3
T-n			12	12	12	12
Null LL			-39445.6	-78513.9	-118259.0	-157536.0
Final LL			-4584.5	-9182.5	-13754.8	-18382.3
R^2			0.884	0.883	0.884	0.883
Adjusted R^2			0.863	0.874	0.878	0.879
RMSD ($\widehat{\beta}$)			0.306	0.284	0.280	0.275

Note: for each scenario, we report the average values based on 10 simulations.



Note: the three (dash) lines are plotted using spline interpolation.

Figure 7. 6 Approximate RMSE with respect to number of households

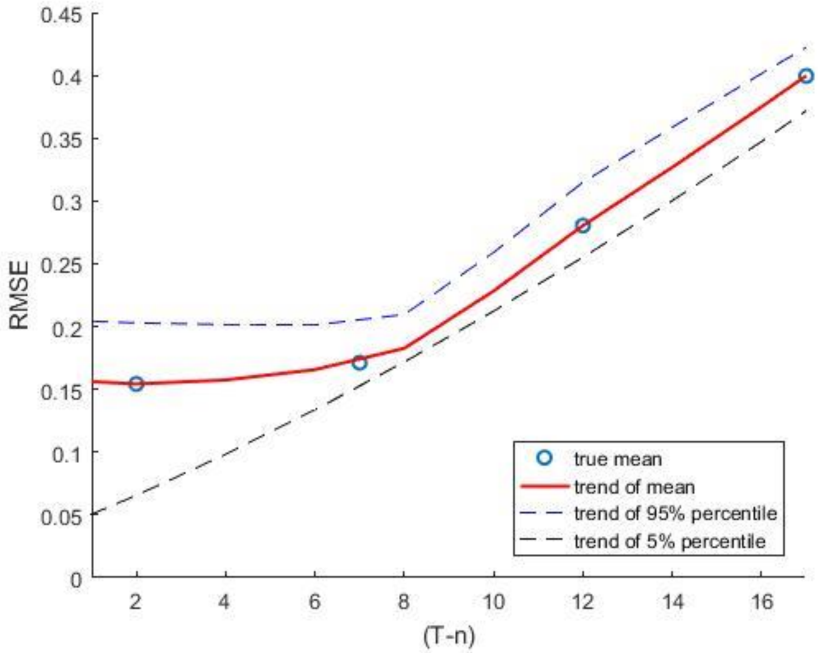
Figure 7.6 plots the mean, 95% percentile, and 5% percentile of the RMSE with respect to household number. The vertical distance between the two dash lines describes the variability of RMSE. As household number increases, the decreasing rate of RMSE is approaching to zero especially when household number is larger than 600. In Figure 7.6, there is no obvious change in the variability of RMSE. Thus, in order to have an appropriate household number and to save computational cost, 600 households are simulated for the scenarios in group 2-4.

By fixing look-forward time periods (n), Table 7.4 compares the estimation results for scenarios with different study time periods (T). In group 2, another changing attribute is the time difference (T-n), which illustrates the true time periods for estimation. After evaluating the influence of time difference on the RMSE of parameters, we can observe that to some extent the RMSE increases as time difference increases.

Table 7. 4 Simulation Scenarios with Different (T-n): Fix Look-forward Time Periods

Coefficients	zero car	at least one car	True $\boldsymbol{\beta}$	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 4	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 5	Estimated $\boldsymbol{\beta}$ (t-statistics) Base	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 6
$\boldsymbol{\beta_1}$	X	X	1.00	1.10 (8.4)	1.00 (14.6)	0.97 (18.4)	1.02 (22.7)
$\boldsymbol{\beta_2}$	X		-0.50	-0.46 (-4.6)	-0.50 (-9.4)	-0.48 (-11.8)	-0.51 (-14.9)
$\boldsymbol{\beta_3}$		X	-2.00	-2.03 (-5.3)	-2.00 (-9.8)	-1.92 (-12.3)	-2.05 (-15.4)
$\boldsymbol{\beta_4}$		X	-3.00	-3.11 (-18.3)	-2.98 (-34.1)	-2.95 (-44.6)	-2.94 (-52.7)
$\boldsymbol{\beta_5}$		X	1.00	1.02 (9.4)	1.01 (17.7)	0.98 (22.6)	0.99 (26.9)
$\boldsymbol{\beta_6}$			1.00	1.01 (41.5)	1.00 (72.3)	1.00 (92.4)	1.00 (105.1)
$\boldsymbol{\beta_7}$			0.50	0.50 (26.6)	0.50 (47.2)	0.50 (59.5)	0.50 (68.0)
Households no.				600	600	600	600
Study Time T				5	10	15	20
Look-forward n				3	3	3	3
T-n				2	7	12	17
Null LL				-18576.3	-66688.0	-118259.0	-172205.0
Final LL				-2164.6	-7879.3	-13754.8	-19927.6
$\boldsymbol{R^2}$				0.883	0.882	0.884	0.884
Adjusted $\boldsymbol{R^2}$				0.881	0.878	0.878	0.876
RMSE ($\widehat{\boldsymbol{\beta}}$)				0.154	0.171	0.280	0.400

Note: for each scenario, we report the average values based on 10 simulations.



Note: the three (dash) lines are plotted using spline interpolation.

Figure 7. 7 Approximate RMSE with respect to (T-n): fix look-forward time periods

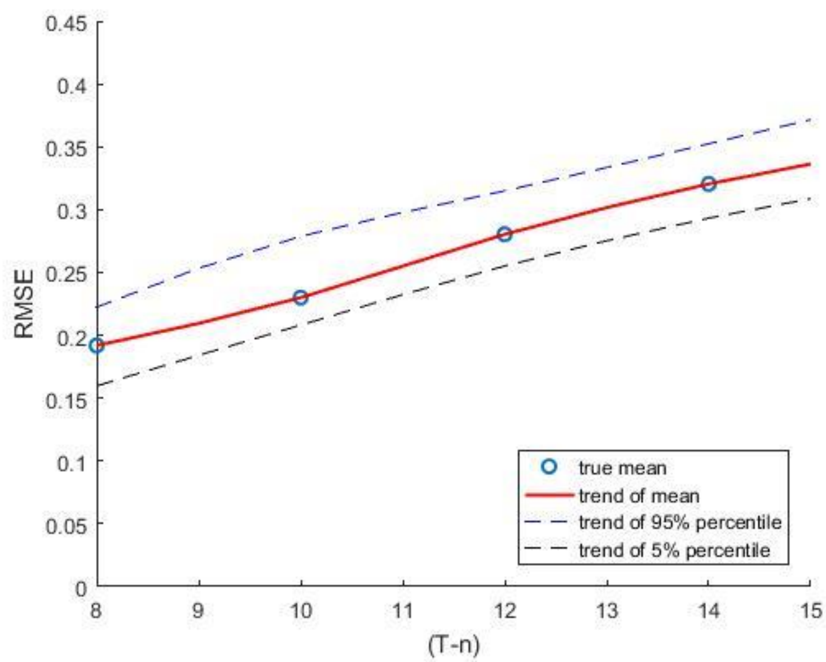
Similarly, Figure 7.7 plots the mean, 95% percentile, and 5% percentile of RMSE with respect to time difference. The vertical distance between the two dash lines describes the variability of RMSE. We can observe that when (T-n) is smaller than 7 (approximately), the value of RMSE slightly increases and its variability decreases as the time difference increases. When (T-n) is larger than 7 (approximately), the value of RMSE increases and its variability is stable as the time difference increases. Thus, 7 is selected as an appropriate time difference to guarantee small RMSE and small variability for the scenarios in group 4.

By fixing study time periods (T), Table 7.5 compares the estimation results for scenarios with different look-forward time periods (n). Similar with group 2, scenarios in group 3 also focus on the change of the time difference (T-n), which corresponds to the true time periods for estimation.

Table 7. 5 Simulation Scenarios with Different (T-n): Fix Study Time Periods

Coefficients	zero car	at least one car	True $\boldsymbol{\beta}$	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 7	Estimated $\boldsymbol{\beta}$ (t-statistics) Base	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 8	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 9
$\boldsymbol{\beta_1}$	X	X	1.00	0.97 (19.8)	0.97 (18.4)	0.98 (17.2)	1.01 (15.8)
$\boldsymbol{\beta_2}$	X		-0.50	-0.50 (-13.5)	-0.48 (-11.8)	-0.49 (-11.0)	-0.50 (-10.1)
$\boldsymbol{\beta_3}$		X	-2.00	-1.95 (-13.4)	-1.92 (-12.3)	-1.99 (-11.7)	-2.03 (-10.6)
$\boldsymbol{\beta_4}$		X	-3.00	-2.94 (-48.3)	-2.95 (-44.6)	-2.91 (-40.6)	-3.03 (-36.4)
$\boldsymbol{\beta_5}$		X	1.00	0.97 (24.3)	0.98 (22.6)	0.96 (20.2)	1.01 (19.1)
$\boldsymbol{\beta_6}$			1.00	1.00 (98.2)	1.00 (92.4)	1.00 (85.3)	0.99 (77.4)
$\boldsymbol{\beta_7}$			0.50	0.50 (63.0)	0.50 (59.5)	0.50 (55.7)	0.50 (50.8)
Households no.				600	600	600	600
Study Time T				15	15	15	15
Look-forward n				1	3	5	7
T-n				14	12	10	8
Null LL				-138606.0	-118259.0	-96879.3	-77011.3
Final LL				-16185.8	-13754.8	-11361.8	-9001.5
$\boldsymbol{R^2}$				0.883	0.884	0.883	0.883
Adjusted $\boldsymbol{R^2}$				0.876	0.878	0.877	0.879
RMSD ($\widehat{\boldsymbol{\beta}}$)				0.320	0.280	0.230	0.191

Note: for each scenario, we report the average values based on 10 simulations.



Note: the three (dash) lines are plotted using spline interpolation.

Figure 7. 8 Approximate RMSE with respect to (T-n): fix study time periods

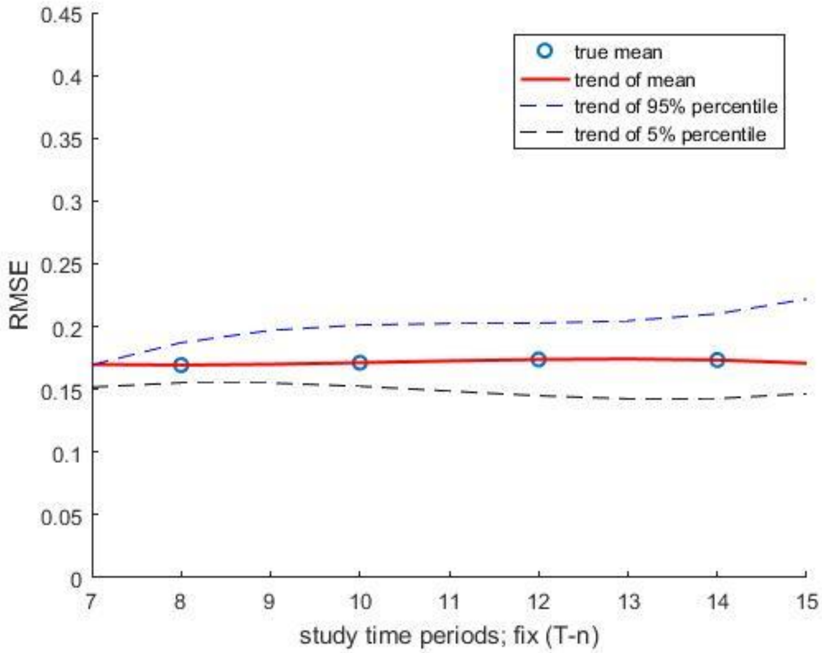
The increasing trend of RMSE in Figure 7.8 is consistent with the trend in Figure 7.7; the only difference is the study range of time difference (T-n) on the horizontal axle. The comparison between Figure 7.7 and 7.8 indicates that the value of RMSE to a large extent depends on the time difference (T-n) instead of study time periods (T) or look-forward time periods (n).

To further prove this finding, Tables 7.6 compares the estimation results for scenarios with fixed time difference (T-n) and flexible study time periods (T). As expected, the value of RMSE almost keeps the same as the number of study time periods increases. In addition, Figure 7.9 shows that variability of RMSE increases as the number of study time periods increases.

Table 7. 6 Simulation Scenarios with Different Study Time Periods: Fix (T-n)

Coefficients	zero car	at least one car	True $\boldsymbol{\beta}$	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 4	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 5	Estimated $\boldsymbol{\beta}$ (t-statistics) Base	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 6
$\boldsymbol{\beta_1}$	X	X	1.00	0.96 (14.2)	1.00 (14.6)	0.99 (14.5)	0.97 (14.4)
$\boldsymbol{\beta_2}$	X		-0.50	-0.49 (-9.3)	-0.50 (-9.4)	-0.50 (-9.4)	-0.46 (-8.8)
$\boldsymbol{\beta_3}$		X	-2.00	-1.90 (-9.4)	-2.00 (-9.8)	-1.96 (-9.7)	-1.89 (-9.4)
$\boldsymbol{\beta_4}$		X	-3.00	-3.01 (-34.3)	-2.98 (-34.1)	-2.97 (-34.3)	-2.95 (-34.3)
$\boldsymbol{\beta_5}$		X	1.00	-1.01 (17.8)	1.01 (17.7)	0.97 (17.5)	0.98 (17.5)
$\boldsymbol{\beta_6}$			1.00	1.00 (73.4)	1.00 (72.3)	0.99 (73.2)	1.00 (73.3)
$\boldsymbol{\beta_7}$			0.50	0.50 (48.1)	0.50 (47.2)	0.51 (48.6)	0.50 (47.9)
Households no.				600	600	600	600
Study Time T				8	10	12	14
Look-forward n				1	3	5	7
T-n				7	7	7	7
Null LL				-66944.9	-66688.0	-67240.4	-66944.3
Final LL				-7846.4	-7879.3	-7833.1	-7846.2
$\boldsymbol{R^2}$				0.883	0.882	0.883	0.883
Adjusted $\boldsymbol{R^2}$				0.879	0.878	0.879	0.879
RMSE ($\widehat{\boldsymbol{\beta}}$)				0.169	0.171	0.174	0.173

Note: for each scenario, we report the average values based on 10 simulations.



Note: the three (dash) lines are plotted using spline interpolation.

Figure 7. 9 Approximate RMSE with respect to study time periods: fix (T-n)

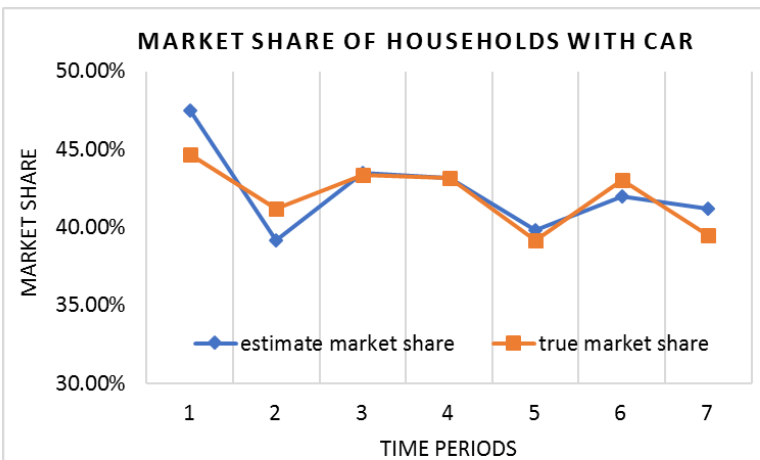
Based on the findings of the simulation experiment, I choose the best case and the worst case from the 13 proposed scenarios that reduce the value and variability of RMSE of the estimated parameters; they are listed as follows:

- Best Case - Scenario 10: a sample of 600 households with 1 look-forward time periods over 8 study time periods
- Worst Case - Scenario 6: a sample of 600 households with 3 look-forward time periods over 20 study time periods

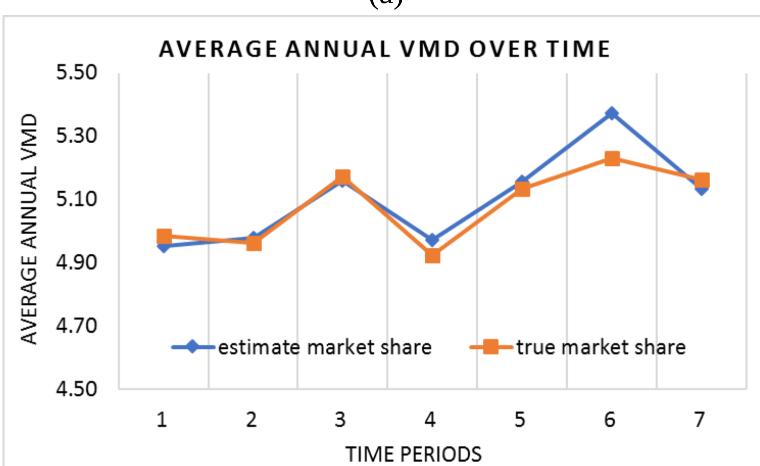
The findings from the simulation experiment can help researchers to better understand and use the proposed dynamic discrete-continuous choice model, and to have a thoughtful idea for data selection.

7.6.2 Model Prediction

To measure the predictive power of the proposed model, I apply the estimated coefficients to forecast households' car ownership and use choices over time. The model prediction results are shown for the best scenario (Scenario 10) and the worst scenario (Scenario 6) for comparison purpose. Figure 7.10 (a) compares the predicted and actual shares of households owning at least one car over 7 time periods (study time of 8 minus look-forward time of 1); while Figure 7.10 (b) compares the predicted and actual average annual VMD over 7 time periods. Similarly, Figure 7.11 compares the predict and actual trends of market share and annual VMD over 17 time periods (study time of 20 minus look-forward time of 3).

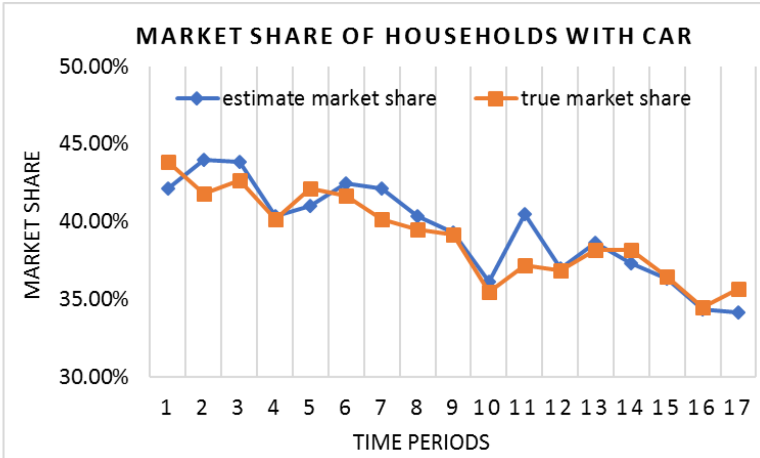


(a)

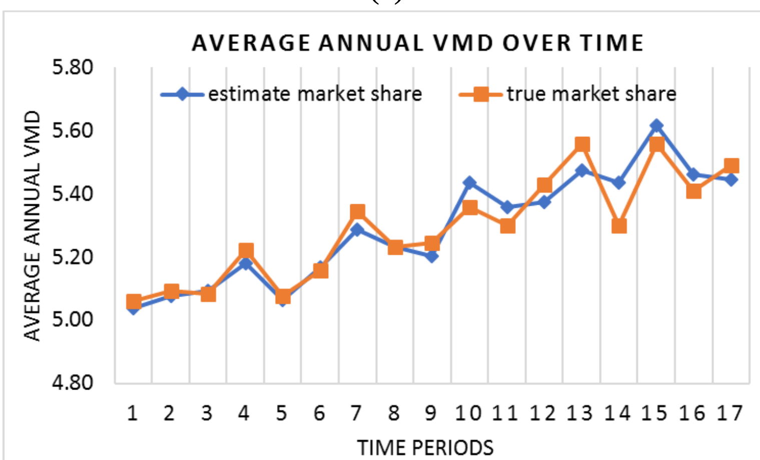


(b)

Figure 7. 10 Model prediction of scenario 10: comparison of true and estimate values



(a)



(b)

Figure 7. 11 Model prediction of scenario 6: comparison of true and estimate values

By observing the results in Figure 7.10 – 7.11, it is reasonable to conclude that the proposed model is capable to reproduce the market share of households with car and the annual VMD over time, by capturing fluctuations, peaks, and valleys of their evolving trends.

7.7 Experiment with Simulated Data – Multivariate Discrete Choice

To validate the proposed dynamic discrete-continuous choice model with multiple discrete choices, simulated data is generated based on the descriptive statistics of MVSPS data in Section 3.1. The simulated data includes household socioeconomics, land use variables, vehicle information, and driving cost. Household discrete choices on vehicle holding and continuous choices on annual VMD are then generated based on utility maximization theory. In particular, three alternatives are available for the discrete part: owning zero or one car, owning two cars, and owning three or more cars. The utility functions of the discrete choice for each time period are:

$$U_{\leq 1} = \beta_{typ} X_{typ,\leq 1} + \delta V_{\leq 1} + \varepsilon_{\leq 1} \tag{7.36}$$

$$U_2 = \beta_{typ} X_{typ,2} + \beta_{asc,2} X_{asc,2} + \beta_{sex,2} X_{sex,2} + \beta_{edu,2} X_{edu,2} + \beta_{inc,2} X_{inc,2} + \beta_{kid,2} X_{kid,2} + \beta_{res,2} X_{res,2} + \delta V_2 + \varepsilon_2 \tag{7.37}$$

$$U_{\geq 3} = \beta_{typ} X_{typ,\geq 3} + \beta_{asc,\geq 3} X_{asc,\geq 3} + \beta_{sex,\geq 3} X_{sex,\geq 3} + \beta_{edu,\geq 3} X_{edu,\geq 3} + \beta_{inc,\geq 3} X_{inc,\geq 3} + \beta_{kid,\geq 3} X_{kid,\geq 3} + \beta_{res,\geq 3} X_{res,\geq 3} + \delta V_{\geq 3} + \varepsilon_{\geq 3} \tag{7.38}$$

where the discount factor δ is assumed to be 1; and V is the expected downstream utility. The variables of interest are diversity of vehicle types (X_{typ}), household head gender (X_{sex}), education level of household head (X_{edu}), family income (X_{inc}), number of kids (X_{kid}), and residential density (X_{res}). Households are assumed to be rational and make decisions to maximize their utility.

Meanwhile, the continuous choice for each time period is generated by a regression as follows:

$$Y = \beta_{asc,r} X_{asc,r} + \beta_{sex,r} X_{sex,r} + \beta_{age,r} X_{age,r} + \beta_{inc,r} X_{inc,r} + \beta_{res,r} X_{res,r} + \beta_{gas,r} X_{gas,r} + \varepsilon_r \tag{7.19}$$

In the regression, two additional variables are considered; they are household head age (X_{age}) and driving cost per mile (X_{gas}). It is important to note that the unobserved error components of discrete and continuous parts are simulated with predetermined correlations varying over time.

The model considers nine predictors (Xs); they are assumed to be independent from each other. Table 7.7 summarizes the distributions of these variables of interest, which are simulated in accordance with household and vehicle characteristics of MVSPS data.

Table 7. 7 Distributions of Variables in Simulated Data

Variable Name	Variable Type	Distribution for Simulation	Changes from t to t+1
Diversity of vehicle types (X_{typ})	Continuous	Uniform distribution	Mean increases by 0.05 Range keeps the same
Alternative specific constant (X_{asc})	Constant	1	None
Household head gender (X_{sex})	Dummy	Bernoulli distribution	None
Household head age (X_{age})	Categorical	Categorical distribution	Add 0.5
Household head education level (X_{edu})	Categorical	Categorical distribution	None
Family income (X_{inc})	Categorical	Categorical distribution	None
Number of kids (X_{kid})	Integer	Categorical distribution	None
ln(Residential density (X_{res}))	Continuous	Truncated normal distribution	None
Driving cost (X_{gas})	Continuous	Uniform distribution	Mean increases by 5% Range keeps the same

7.7.1 Model Estimation

Given the model specification, two datasets are simulated including vehicle holding and usage choices of 456 households over a short run and a medium-long run. The short run contains a hypothetical five-year period, while the medium-long run contains a hypothetical nine-year period. In particular, households are assumed to make decisions every half year, with a total of 10 time periods for the short run and 18 time periods for the medium-long run. An R package is developed by the author to estimate the model. To determine the appropriate number of look-forward time periods, Table 7.8 proposes eight estimation scenarios for the short run and five estimation scenarios for the medium-long run for comparison purpose.

Table 7. 8 Summary of Estimation Scenarios

Group	Number of study time (T)	Look-forward time periods (n)	T-n	Estimation Scenario
Short Run	10	1	9	1
		2	8	2
		3	7	3
		4	6	4
		5	5	5
		6	4	6
		7	3	7
		8	2	8
Medium-Long Run	18	2	16	9
		4	14	10
		6	12	11
		8	10	12
		10	8	13

Table 7.9 - 7.10 summarize the estimation results for each scenario, and compare the true and estimated parameters. It should be noted that each scenario contains 10 simulated datasets for estimation, and the reported results are the average values based on the 10 estimations. We can increase the number of estimations in each scenario to guarantee a non-bias result in future research.

We can observe that most estimated parameters are approaching to the true values in all 13 scenarios. The values of $\boldsymbol{R^2}$ ranges from 0.311 to 0.337 in the short run and ranges from 0.291 to 0.311 in the medium-long run, which show that the log-likelihood has been improved by using the proposed model for estimation especially in the short run. Figure 7.12 – 7.13 plot the trend line of RMSE of the estimated coefficients and provide important evidence to determine the appropriate look-forward time periods.

Table 7.9 compares the estimation results for eight scenarios in the short run with different look-forward time periods. Besides the alternative specific constants in the discrete part, other estimated coefficients are close to the true values. The value of RMSE of the estimated coefficients decreases as the number of look-forward time periods increases. However, when the number of look-forward time periods is very large and is close to the total study time, some estimated coefficients become insignificant possibly due to the fact that the number of observations for estimation is small.

Table 7. 9 Result Comparison: Estimation Scenarios with Different Look-Forward Time Periods in the Short Run

Coefficients	≤ 1 car	2 cars	≥ 3 cars	True $\boldsymbol{\beta}$	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 1	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 2	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 3	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 4
$\boldsymbol{\beta_{typ}}$	X	X	X	0.50	0.52 (14.2)	0.52 (13.4)	0.51 (12.6)	0.52 (11.8)
$\boldsymbol{\beta_{asc,2}}$		X		-1.00	-0.78 (-4.6)	-0.81 (-4.5)	-0.90 (-5.0)	-0.89 (-4.4)
$\boldsymbol{\beta_{asc,3}}$			X	-2.00	-1.82 (-7.2)	-1.85 (-6.9)	-1.94 (-6.8)	-1.90 (-6.3)
$\boldsymbol{\beta_{sex,2}}$		X		-0.40	-0.40 (-6.4)	-0.47 (-7.1)	-0.45 (-6.4)	-0.37 (-4.8)
$\boldsymbol{\beta_{sex,3}}$			X	-0.30	-0.31 (-4.1)	-0.35 (-4.3)	-0.34 (-4.0)	-0.23 (-2.4)
$\boldsymbol{\beta_{edu,2}}$		X		-0.10	-0.14 (-5.6)	-0.13 (-4.9)	-0.11 (-4.3)	-0.13 (-4.5)
$\boldsymbol{\beta_{edu,3}}$			X	-0.20	-0.24 (-8.4)	-0.24 (-7.8)	-0.23 (-7.1)	-0.26 (-7.1)
$\boldsymbol{\beta_{inc,2}}$		X		0.30	0.30 (10.4)	0.31 (9.9)	0.31 (9.9)	0.30 (8.5)
$\boldsymbol{\beta_{inc,3}}$			X	0.50	0.51 (12.1)	0.52 (11.6)	0.53 (10.6)	0.52 (9.7)
$\boldsymbol{\beta_{kid,2}}$		X		0.20	0.18 (4.5)	0.16 (3.8)	0.20 (4.4)	0.23 (4.5)
$\boldsymbol{\beta_{kid,3}}$			X	0.30	0.29 (6.2)	0.27 (5.6)	0.30 (5.5)	0.32 (5.2)
$\boldsymbol{\beta_{res,2}}$		X		-0.10	-0.11 (-6.3)	-0.11 (-5.5)	-0.11 (-5.7)	-0.11 (-5.0)
$\boldsymbol{\beta_{res,3}}$			X	-0.20	-0.23 (-10)	-0.23 (-9.3)	-0.23 (-8.9)	-0.24 (-8.1)
$\boldsymbol{\beta_{asc,r}}$	X	X	X	3.00	2.96 (35.3)	3.01 (34.8)	3.03 (33.2)	2.97 (30.0)
$\boldsymbol{\beta_{sex,r}}$	X	X	X	-0.30	-0.31 (-8.8)	-0.32 (-8.7)	-0.33 (-8.4)	-0.27 (-6.4)
$\boldsymbol{\beta_{age,r}}$	X	X	X	-0.20	-0.19 (-15.7)	-0.20 (-15.0)	-0.20 (-14.4)	-0.20 (-13.5)

$\boldsymbol{\beta_{inc,r}}$	X	X	X	0.10	0.11 (9.2)	0.10 (8.3)	0.10 (7.6)	0.10 (6.9)
$\boldsymbol{\beta_{res,r}}$	X	X	X	-0.10	-0.10 (-11.7)	-0.10 (-11.8)	-0.10 (-11.5)	-0.10 (-9.9)
$\boldsymbol{\beta_{gas,r}}$	X	X	X	-6.00	-6.01 (-19.9)	-5.93 (-18.7)	-5.92 (-17.4)	-5.93 (-16.2)
Households					456	456	456	456
Study time					10	10	10	10
Look-forward					1	2	3	4
T-n					9	8	7	6
Null LL					-14666.8	-13040.8	-11418.6	-9744.1
Final LL					-10108.5	-8920.8	-7787.3	-6646.4
$\boldsymbol{R^2}$					0.311	0.316	0.318	0.318
Adjusted $\boldsymbol{R^2}$					0.300	0.305	0.306	0.306
RMSD ($\hat{\boldsymbol{\beta}}$)					1.605	1.525	1.554	1.078

Note: for each scenario, we report the average values based on 10 simulations.

Coefficients	$\leq \mathbf{1}$ car	2 cars	$\geq \mathbf{3}$ cars	True $\boldsymbol{\beta}$	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 5	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 6	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 7	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 8
$\boldsymbol{\beta_{typ}}$	X	X	X	0.50	0.54 (10.9)	0.53 (9.5)	0.51 (8.1)	0.51 (6.6)
$\boldsymbol{\beta_{asc,2}}$		X		-1.00	-0.83 (-3.9)	-0.89 (-3.6)	-0.71 (-2.4)	-0.68 (-1.9)*
$\boldsymbol{\beta_{asc,3}}$			X	-2.00	-2.07 (-5.8)	-1.94 (-4.9)	-1.72 (-4.0)	-1.58 (-3.1)
$\boldsymbol{\beta_{sex,2}}$		X		-0.40	-0.40 (-4.8)	-0.36 (-3.9)	-0.37 (-3.4)	-0.42 (-3.0)
$\boldsymbol{\beta_{sex,3}}$			X	-0.30	-0.40 (-3.5)	-0.29 (-2.4)	-0.29 (-2.1)	-0.42 (-2.3)
$\boldsymbol{\beta_{edu,2}}$		X		-0.10	-0.13 (-4.0)	-0.14 (-3.6)	-0.14 (-3.2)	-0.17 (-3.2)
$\boldsymbol{\beta_{edu,3}}$			X	-0.20	-0.25 (-6.1)	-0.28 (-5.6)	-0.26 (-5.1)	-0.30 (-4.4)
$\boldsymbol{\beta_{inc,2}}$		X		0.30	0.30 (8.1)	0.30 (6.8)	0.31 (6.3)	0.31 (5.4)
$\boldsymbol{\beta_{inc,3}}$			X	0.50	0.56 (8.7)	0.53 (7.6)	0.52 (7.0)	0.53 (5.6)
$\boldsymbol{\beta_{kid,2}}$		X		0.20	0.20 (3.8)	0.24 (3.7)	0.15 (2.1)	0.20 (2.3)
$\boldsymbol{\beta_{kid,3}}$			X	0.30	0.33 (4.8)	0.31 (4.1)	0.29 (3.5)	0.26 (2.5)
$\boldsymbol{\beta_{res,2}}$		X		-0.10	-0.12 (-5.3)	-0.12 (-4.6)	-0.12 (-3.9)	-0.12 (-3.4)
$\boldsymbol{\beta_{res,3}}$			X	-0.20	-0.24 (-7.2)	-0.22 (-6.2)	-0.23 (-5.7)	-0.24 (-4.6)
$\boldsymbol{\beta_{asc,r}}$	X	X	X	3.00	2.96 (28.1)	2.96 (25.1)	2.97 (21.8)	2.97 (18.2)
$\boldsymbol{\beta_{sex,r}}$	X	X	X	-0.30	-0.28 (-6.4)	-0.32 (-6.4)	-0.28 (-5.0)	-0.26 (-3.8)
$\boldsymbol{\beta_{age,r}}$	X	X	X	-0.20	-0.21 (-12.8)	-0.20 (-11.2)	-0.20 (-9.5)	-0.19 (-7.6)
$\boldsymbol{\beta_{inc,r}}$	X	X	X	0.10	0.10 (6.7)	0.10 (6.0)	0.10 (5.3)	0.10 (4.4)
$\boldsymbol{\beta_{res,r}}$	X	X	X	-0.10	-0.09 (-8.5)	-0.09 (-7.6)	-0.10 (-7.3)	-0.10 (-5.8)
$\boldsymbol{\beta_{gas,r}}$	X	X	X	-6.00	-5.87 (-14.8)	-5.85 (-13.0)	-6.04 (-11.8)	-6.28 (-10.0)
Households					456	456	456	456
Study time					10	10	10	10
Look-forward					5	6	7	8
T-n					5	4	3	2
Null LL					-8109.6	-6511.7	-4878.3	-3284.2

Final LL	-5493.7	-4386.7	-3269.2	-2177.1
R^2	0.323	0.326	0.330	0.337
Adjusted R^2	0.310	0.312	0.313	0.316
RMSD ($\hat{\beta}$)	1.082	1.006	0.743	0.673

*Note: for each scenario, we report the average values based on 10 simulations; * indicates the estimated coefficient is not significant at 95% significant level.*

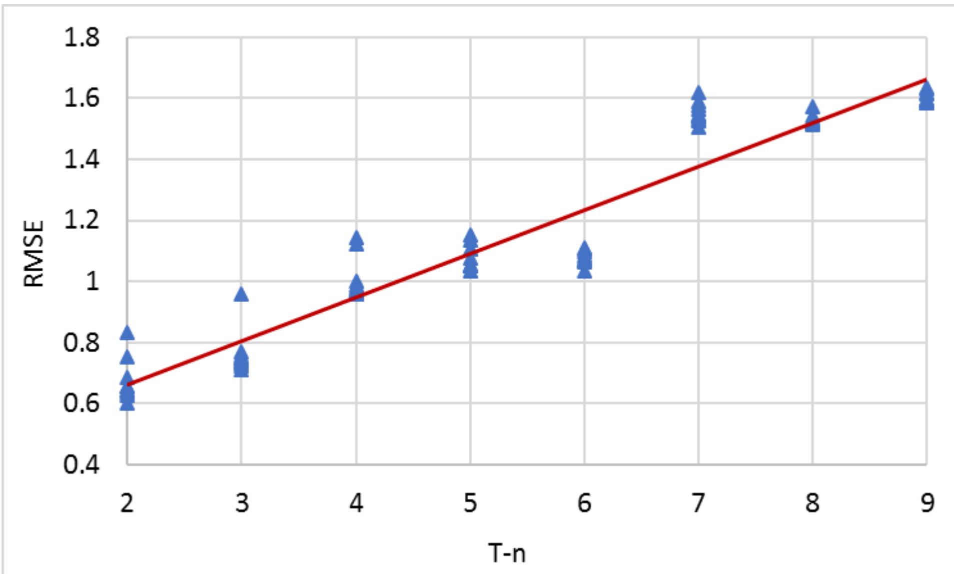


Figure 7. 12 Trend line of RMSE with respect to time difference (T-n) in a short run

Figure 7.12 plots the trend line of RMSE of the estimated coefficients with respect to the difference between total study time and look-forward time. In the short run, the total study time is 10 and the look-forward time ranges from 1 to 8. In figure 7.12, one blue triangle represents the RMSE for one estimation, with a total of 80 estimations (8 scenarios times 10 estimations for each scenario). We can observe that the value of RMSE increases as the time difference increases. In other words, given enough observations for estimation, the value of RMSE decreases as the number of look-forward time periods increases.

Table 7.10 compares the estimation results for five scenarios in the medium-long run with different look-forward time periods. Most estimated coefficients are close to the true values in all five scenarios. Similar with the case in the short run, the value of RMSE of the estimated coefficients decreases as the number of look-forward time periods increases.

Table 7. 10 Result Comparison: Estimation Scenarios with Different Look-Forward Time Periods in the Medium-Long Run

Coefficients	$\leq \mathbf{1}$ car	2 cars	$\geq \mathbf{3}$ cars	True $\boldsymbol{\beta}$	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 9	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 10	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 11	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 12	Estimated $\boldsymbol{\beta}$ (t-statistics) Scenario 13
$\boldsymbol{\beta_{typ}}$	X	X	X	0.50	0.50 (19.0)	0.51 (16.8)	0.51 (16.8)	0.51 (14.6)	0.50 (13.4)
$\boldsymbol{\beta_{asc,2}}$		X		-1.00	-0.82 (-6.9)	-0.75 (-5.5)	-0.73 (-5.8)	-0.77 (-4.9)	-0.89 (-5.4)
$\boldsymbol{\beta_{asc,3}}$			X	-2.00	-1.88 (-10.1)	-1.89 (-8.8)	-1.89 (-8.7)	-1.97 (-7.6)	-1.95 (-7.5)
$\boldsymbol{\beta_{sex,2}}$		X		-0.40	-0.40 (-8.6)	-0.44 (-8.8)	-0.41 (-7.6)	-0.43 (-7.2)	-0.43 (-6.4)
$\boldsymbol{\beta_{sex,3}}$			X	-0.30	-0.32 (-5.8)	-0.30 (-4.7)	-0.32 (-5.3)	-0.32 (-4.3)	-0.31 (-4.2)
$\boldsymbol{\beta_{edu,2}}$		X		-0.10	-0.11 (-6.4)	-0.12 (-5.9)	-0.13 (-6.1)	-0.12 (-5.3)	-0.11 (-4.6)
$\boldsymbol{\beta_{edu,3}}$			X	-0.20	-0.23 (-11.1)	-0.22 (-9.3)	-0.23 (-14.1)	-0.24 (-8.4)	-0.21 (-7.4)
$\boldsymbol{\beta_{inc,2}}$		X		0.30	0.28 (14.0)	0.28 (11.9)	0.28 (12.4)	0.29 (10.2)	0.30 (15.1)
$\boldsymbol{\beta_{inc,3}}$			X	0.50	0.50 (16.3)	0.49 (13.8)	0.50 (35.2)	0.53 (12.2)	0.50 (11.3)
$\boldsymbol{\beta_{kid,2}}$		X		0.20	0.19 (6.5)	0.19 (5.3)	0.19 (5.8)	0.17 (4.4)	0.20 (4.7)
$\boldsymbol{\beta_{kid,3}}$			X	0.30	0.32 (9.0)	0.31 (7.5)	0.32 (8.5)	0.31 (6.7)	0.31 (6.3)
$\boldsymbol{\beta_{res,2}}$		X		-0.10	-0.11 (-9.1)	-0.12 (-8.1)	-0.12 (-8.6)	-0.11 (-6.7)	-0.11 (-6.9)
$\boldsymbol{\beta_{res,3}}$			X	-0.20	-0.22 (-13.4)	-0.23 (-11.6)	-0.22 (-14.6)	-0.22 (-10.2)	-0.21 (-9.4)
$\boldsymbol{\beta_{asc,r}}$	X	X	X	3.00	3.02 (46.2)	3.03 (43.3)	2.98 (40.1)	3.00 (37.6)	3.00 (34.1)
$\boldsymbol{\beta_{sex,r}}$	X	X	X	-0.30	-0.32 (-11.7)	-0.32 (-11.1)	-0.30 (-9.6)	-0.32 (-9.6)	-0.33 (-8.9)
$\boldsymbol{\beta_{age,r}}$	X	X	X	-0.20	-0.21 (-21.4)	-0.21 (-20.0)	-0.20 (-18.1)	-0.21 (-17.3)	-0.20 (-15.2)
$\boldsymbol{\beta_{inc,r}}$	X	X	X	0.10	0.10 (11.3)	0.10 (10.8)	0.10 (10.4)	0.11 (10.3)	0.11 (8.8)
$\boldsymbol{\beta_{res,r}}$	X	X	X	-0.10	-0.10 (-15.0)	-0.10 (-15.1)	-0.10 (-13.4)	-0.10 (-12.9)	-0.10 (-11.0)
$\boldsymbol{\beta_{gas,r}}$	X	X	X	-6.00	-5.99 (-26.8)	-5.96 (-24.7)	-5.97 (-22.9)	6.07 (-21.2)	-6.16 (-19.1)
Households					456	456	456	456	456
Study time					18	18	18	18	18
Look-forward					2	4	6	8	10
T-n					16	14	12	10	8
Null LL					-25933.5	-22777.4	-19407.1	-16205.3	-13015.9
Final LL					-18375.9	-15939.9	-13602.0	-11241.7	-8950.5
$\boldsymbol{R^2}$					0.291	0.300	0.299	0.306	0.311
Adjusted $\boldsymbol{R^2}$					0.282	0.290	0.289	0.296	0.301
RMSD ($\widehat{\boldsymbol{\beta}}$)					1.527	1.557	1.480	1.472	1.474

Note: for each scenario, we report the average values based on 10 simulations.

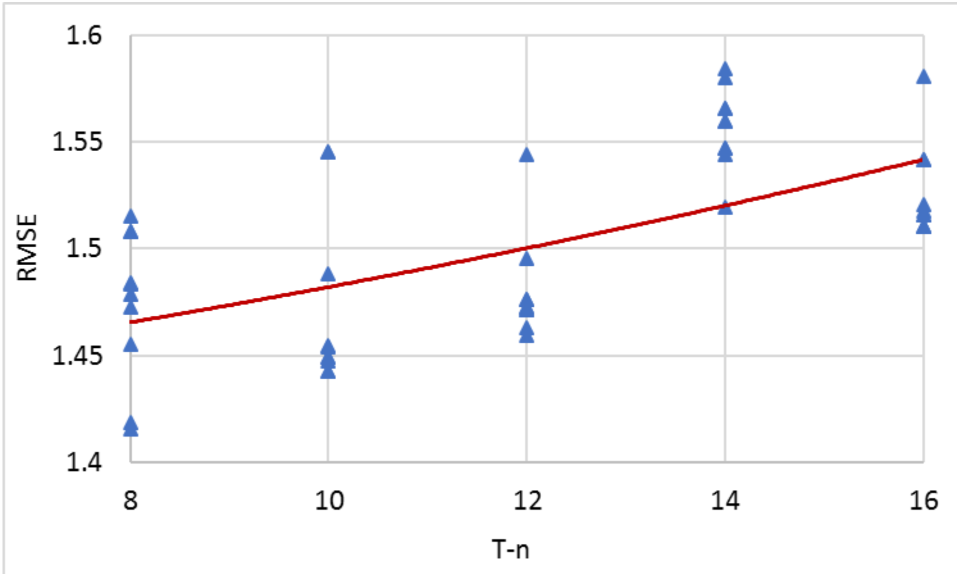


Figure 7. 13 Trend line of RMSE with respect to time difference (T-n) in a medium-long run

Figure 7.13 plots the trend line of RMSE of the estimated coefficients with respect to the difference between total study time and look-forward time. In the medium-long run, the total study time is 18 and the look-forward time ranges from 2 to 10. In figure 7.13, one blue triangle represents the RMSE for one estimation, with a total of 50 estimations (5 scenarios times 10 estimations for each scenario). We can observe that the value of RMSE increases as the time difference increases. The variability of RMSE keeps stable when the time difference increases from 8 to 14. More estimations can be considered for each scenario in order to determine the appropriate look-forward time that identifies the balance between the value and the variability of RMSE.

Based on the findings of the simulation results, the best scenario for the short run and for the medium-long run are chosen to reduce the value and variability of RMSE of the estimated parameters. The two best scenarios are listed as follows:

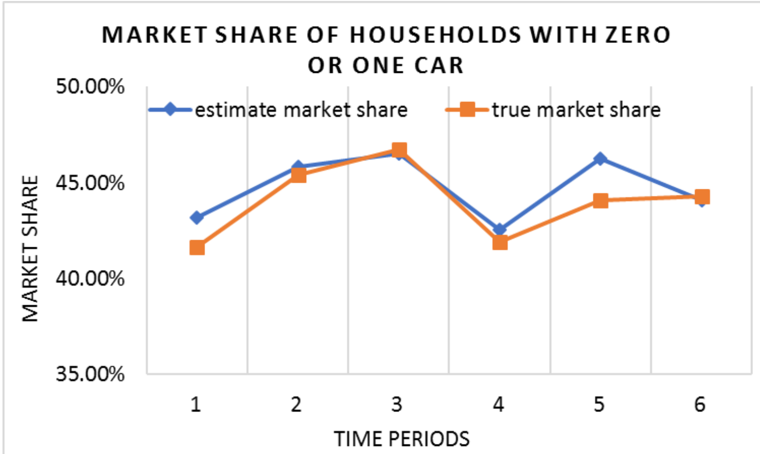
- Scenario 4 in the short run: a sample of 456 households with 4 look-forward time periods over 10 study time periods
- Scenario 11 in the medium-long run: a sample of 456 households with 6 look-forward time periods over 18 study time periods

These two scenarios are employed for model prediction in the following section.

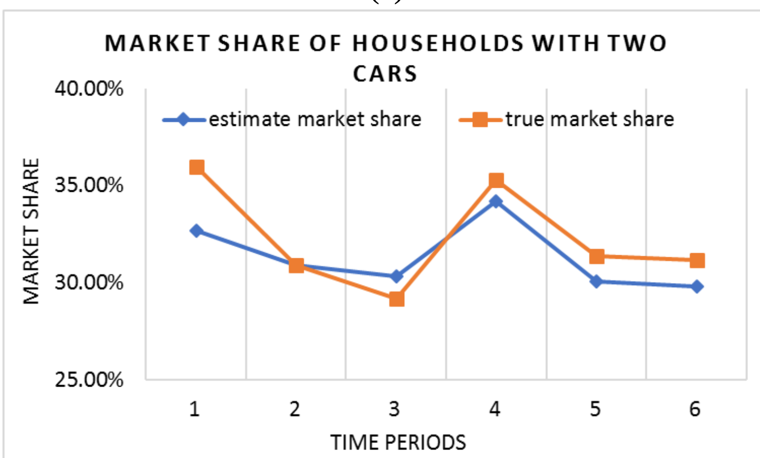
7.7.2 Model Prediction

To measure the predictive power of the sequential discrete-continuous choice model (with multiple discrete alternatives), the estimated coefficients are applied to forecast the shares of households holding different number of cars and the average annual VMD over time. Two scenarios are considered for model prediction: one (scenario 4) is for the short run and the other (scenario 11) is for the medium-long rum. In particular, the case of short run contains 6 time periods (study time of 10 minus look-forward time of 4), while the case of medium-long run contains 12 time periods (study time of 18 minus look-forward time of 6). The prediction results are shown in Figure 7.14 – 7.15. In both cases, figures (a) – (c) compare the predicted and actual percentages of households holding zero or one car, two cars, and three or more cars respectively; figure (d) compares the predicted and actual average annual VMD over time.

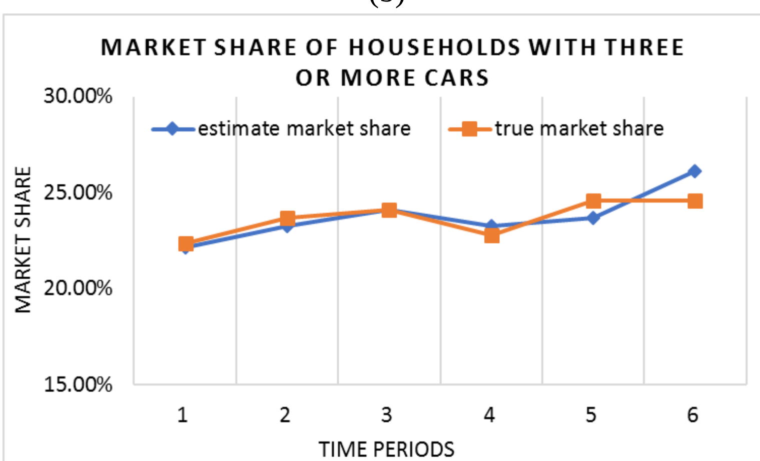
We can observe that the proposed model is capable to reproduce the shares of households holding different number of cars and the average annual VMD over time, especially in the short run. The predicted trends capture the fluctuations, peaks, and valleys of the actual trends in market shares and annual VMD.


MARKET SHARE OF HOUSEHOLDS WITH ZERO OR ONE CAR
50.00%
45.00%
40.00%
35.00%
MARKET SHARE
estimate market share
true market share
1
2
3
4
5
6
TIME PERIODS

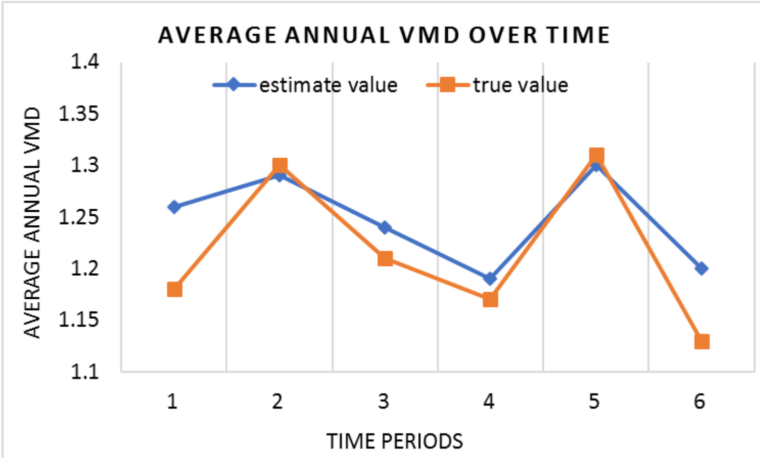
(a)


MARKET SHARE OF HOUSEHOLDS WITH TWO CARS
40.00%
35.00%
30.00%
25.00%
MARKET SHARE
estimate market share
true market share
1
2
3
4
5
6
TIME PERIODS

(b)

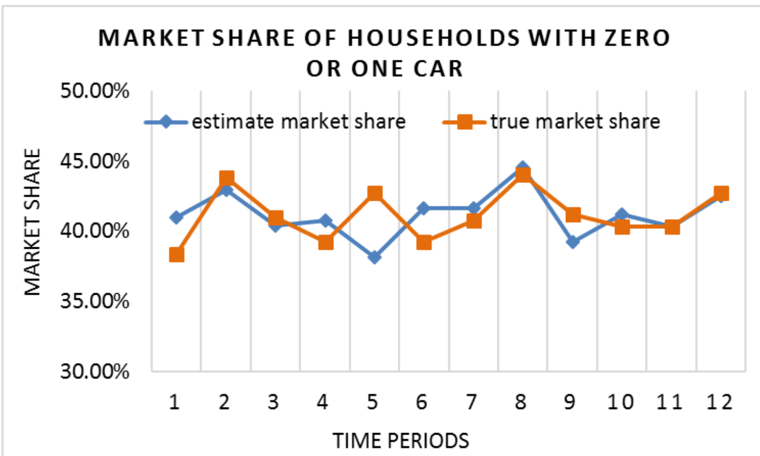

MARKET SHARE OF HOUSEHOLDS WITH THREE OR MORE CARS
30.00%
25.00%
20.00%
15.00%
MARKET SHARE
estimate market share
true market share
1
2
3
4
5
6
TIME PERIODS

(c)

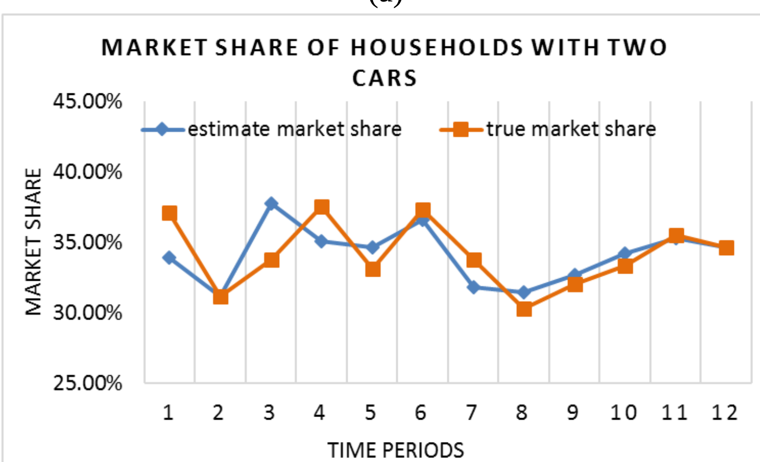


(d)

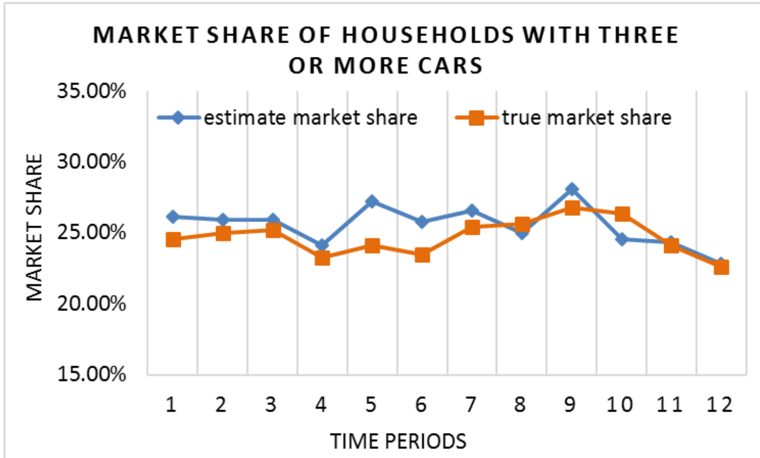
Figure 7. 14 Model prediction of scenario 4: comparison of true and estimate values



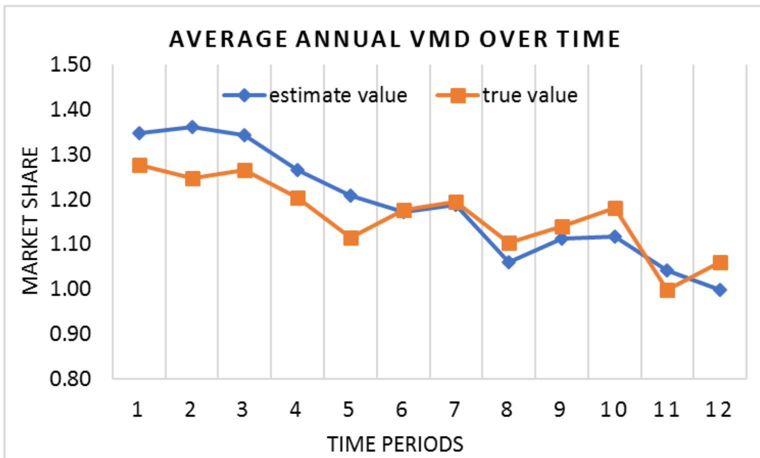
(a)



(b)



(c)



(d)

Figure 7. 15 Model prediction of scenario 11: comparison of true and estimate values

7.8 Chapter Conclusions

This Chapter describes a dynamic integrated modeling framework that accounts for a sequence of discrete and continuous decisions made by forward-looking agents in a finite time horizon. In the car ownership problem setting, households make decisions on the number of cars to hold and the annual VMD for owned cars. A recursive probit model is formulated to estimate the sequence of vehicle holding decisions over time. The inherent Gaussian distributed error component of the recursive probit model enables its integration with regressions to simultaneously capture households' annual VMD over time. In particular, the time-dependent correlation between the discrete and continuous parts is captured by a sequence of full unrestricted variance-covariance matrices of the unobserved components. The estimation of the model is based on the properties of multivariate normal distribution and the finite-horizon scenario tree technique.

The main contributions of the proposed sequential discrete-continuous choice model can be summarized as follows:

- In the model setting, decision makers can have different starting conditions: each household can own zero, one, or multiple vehicles, and no restriction on their annual VMD.
- Building on the recursive logit model which captures a sequence of discrete choices over time, the paper proposes a recursive probit formulation with Gaussian distributed error terms to further allow: (a) unrestricted substitution pattern between alternatives by considering correlated random components; and (b) integration with regressions which capable of modeling continuous choices.
- The inherent utility is a linear combination of an instantaneous utility, a downstream utility, and an error term, which includes information both on current alternatives and individuals' expectations about future alternatives. The downstream utility is expressed and calculated in a recursive manner; its estimation process does not increase the dimension of integral.
- The proposed model is able to capture the interdependency between discrete choice (i.e. car ownership) and continuous choice (i.e. VMD) over time by

introducing different covariance matrices of the integrated error terms for different time periods. The integration of recursive probit and regression models takes advantage of the property of conditional normal distributions.

- The parameters in both discrete and continuous parts are simultaneously estimated with the maximum likelihood technique; the dynamic discrete decision process is solved by generating a finite-horizon scenario tree at each iteration. The joint probability of discrete and continuous decisions is expressed as the product of marginal probability of annual VMD and conditional probability of car holding decision given annual VMD.

The modeling framework has been applied to simulated datasets of car ownership and use choices. The results show that the estimated coefficients approach to the true values, and the log-likelihood value has been highly improved by using the proposed model for estimation. Besides, it has a strong prediction power to recover changes in both discrete and continuous decisions over time.

In fact, there are wide applications on joint discrete and continuous choices in different areas such as car ownership and use (Liu et al., 2014), activity type and duration (Cirillo et al., 2015-2), energy appliance type and demand (Vaage. 2000). I hope that this work will generate innovations in demand modeling and will be extended to other problems with discrete-continuous choices and dynamic nature.

Chapter 8: Conclusions and Future Research

8.1 Research Conclusions and Contributions

Vehicles with new technologies and alternative fuels are gaining consumer's interest and trust; their market shares are increasing around the world. These technologies include plug-in electric vehicles, long-lasting batteries, electric drive technologies, and efficient combustion engines. They gradually diversify today's vehicle market and influence people's preference on vehicle ownership, type, and usage. In this circumstance, modeling vehicle ownership and usage in the era of advanced vehicle technology becomes important for decision makers to achieve a balance between the objectives of transportation, energy consumption, emission, and economy.

Considering the evolution of vehicle technology and changes in economy, a number of dynamic discrete choice models have been developed in recent decades and applied to the car ownership problem. These studies have addressed a number of interesting modeling issues, such as initial conditions of agents, state dependency or inertia effect, agent's forward-looking behavior, taste variation, substitution pattern among alternatives, and type of data for estimation. In practice, many of these studies ignore the correlation between vehicle usage and vehicle ownership behavior. There are two recent studies that aim at estimating simultaneously household vehicle ownership and usage decisions over time. However, their modeling frameworks have several limitations including (a) a restriction on the number of cars held by households (i.e., at most two cars), (b) a fixed total mileage budget and not sensitive to policy, (c) no measurement of correlations between car ownership and usage decisions, and (d) a two-stage estimation process that leads to insufficient estimated parameters.

Studies in this dissertation overcome the above limitations; a multi-facet approach is taken to develop a mature methodology to forecast the changes in household vehicle ownership, type choice, and usage behavior in a dynamic marketplace. Specifically, the dissertation continues the modeling efforts by:

- Studying the impact of new vehicle technology on car ownership decisions by incorporating indicators of vehicle type in modeling car ownership and usage behaviors
- Jointly modeling the market evolution with a stochastic diffusion process in a car ownership problem
- Joint estimating vehicle ownership and usage choices over time, with no restriction on the number of cars held by households and their driving distance
- Developing a comprehensive framework for the estimation of vehicle ownership, type choice, usage behavior, and vehicular emission, by combining car ownership models with motor emission simulators
- Transferring an advanced car ownership model, calibrated in an urban area in a developed country, to an urban area in a developing country in the presence of survey data with selection bias
- Adopting a dynamic programming framework to facilitate the estimation process of dynamic car ownership models

In particular, the investigation process includes four studies progressively; the modeling frameworks, applications, and results of the four studies are presented in

Chapter 4 – 7. The proposed four models are appropriate for different situations considering research purpose and available data; one model is not necessarily superior compared to the others. The contributions and findings for each study are concluded as follows.

Study 1 in Chapter 4 proposes a stated preference approach to measure vehicle type preference and time-dependent market elasticity of conventional and green vehicles. Specifically, the study uses a stated preference survey to collect respondent's vehicle purchase choices in a dynamic vehicle market. Mixed Multinomial Logit models with panel effect are employed to model consumers' preferences, elasticity values, and willingness-to-pays (WTPs) for different vehicle characteristics based on the stated preference survey data. The main findings are summarized as follows:

- The stated preference survey approach is able to capture respondent's trade-offs between vehicles with different technologies over time, and to mimic dynamics and provide insights in the real vehicle market.
- Mixed logit models calibrated on time-dependent observations deliver results that are consistent with general economic expectations. Results attest that consumers are more sensitive to the purchasing price of new-technology vehicles such as hybrid and electric cars, and young people are more likely to choose these cars.
- Conventional gasoline vehicles are price inelastic while hybrid and electric vehicles are price elastic. Besides, the long-run (9 years) market elasticity for gasoline vehicle is greater than the short-run (5 years) elasticity by a factor of

1.5, which indicates that consumers are reluctant to switch their preferred vehicle in a shorter time period.

- Market elasticity with respect to electricity price is much higher than that to gasoline price, which indicates that potential buyers of electricity-powered vehicles are more sensitive to the energy price.
- Based on WTP analysis, increasing vehicle size is an important factor to encourage people to buy electric cars

Study 2 in Chapter 5 proposes a generalized dynamic discrete choice model to estimate consumer's purchase behavior and future preference on vehicle types in a finite time horizon. The modeling framework allows forward-looking agents to optimize their utilities over time; two options are available at each time: keeping the current vehicle or buying a new vehicle among the options available in the market. Different model forms are proposed to consider the purchase pattern of different durable goods in the market. These dynamic models are used to predict market penetration of gasoline, hybrid and electric cars, and to evaluate the impact of changes in fuel price and car characteristics on vehicle type preferences. The main findings are summarized as follows:

- A vector autoregressive process, integrated with the dynamic framework, is able to mimic market evolution by capturing the changes of multiple correlated market indicators. The estimation requires historical time-series data points, and the quality of estimation depends on the number of data points.

- The proposed dynamic model with market evolution is superior to predict consumer's preference on difference vehicle types and market penetration of new vehicle technology over time. Model validation results show that the dynamic model is particularly appropriate to recover peaks and rapid changes in consumer demand over time.
- Based on sensitivity analysis, consumers in Maryland are more interested in purchasing gasoline and hybrid car. The market share of electric cars gradually increases from 4% to 7% between Year 2014 and 2022; the market share of electric cars highly depends on electricity price, vehicle purchasing price, fuel economy, and recharging range.

Study 3 in Chapter 6 proposes an integrated discrete-continuous choice model that jointly estimates household decisions on vehicle ownership, type preference, and usage pattern. The model combines with motor emission simulators such as MOVES to estimate household-level vehicle emissions. The entire model has been applied to estimate vehicle ownership, usage behaviors, and emission patterns in Maryland, US and in Beijing, China. The former is estimated on a joint stated preference data and a revealed preference data to evaluate that how the appearance of new vehicle technology influences car ownership and usage behavior in Maryland. The latter is estimated on a revealed preference data to test the feasibility of transferring the advanced model from a developed society to a developing society. Different green policies are evaluated on reducing the dependency on vehicle usage and emissions. The main findings are summarized as follows:

- Joint use of stated preference data and revealed preference data is feasible for car ownership analysis if the two datasets have the same target population and the population shares the same demands for vehicle and travel.
- The vehicle type logit sub-model is appropriate to capture consumer's time-dependent preference on green vehicles and trade-offs between different car characteristics. Its logsum serves as an important indicator of vehicle market diversity.
- Model results show that the diversity of vehicle types, including conventional gasoline cars and green cars, has a positive impact on vehicle ownership and usage. In particular, the appearance of green vehicles has a greater influence on vehicle quantity and usage decisions for a longer time period.
- The entire model is appropriate to estimate household-level vehicle greenhouse gas (GHG) emissions. The average annual GHG emissions for household primary car in Maryland are 5.17 tons in the short run (2014-2017) and 5.11 tons in the medium-long run (2014-2022), consistent with values reported by US EPA.
- Moderate gasoline tax will effectively lead to emission reductions by reducing vehicle use. This impact is found to increase over time. Ownership tax will lead to small emission reductions. According to the results, a $3,000 annual ownership fee is less effective than a 20% increase in gasoline price in Maryland.
- It is feasible to transfer the entire framework to Beijing with the support of Beijing Household Travel Survey data. A stratified random sampling

procedure is helpful to reduce selection bias and improve data representativeness.

- The model is flexible to consider a wide variety of attributes including household socioeconomics, household composition, land use, public transit service, bicycle/motorcycle ownership and fuel cost. Model estimates are coherent with general expectations.
- The model is able to estimate different types of vehicle emissions such as CO, HC, NOx, CO2, PM2.5, and PM10 by capturing number of car by households, vehicle type, annual driving distance, and emission factors of different vehicle types.
- Based on sensitivity analysis, car ownership in Beijing is sensitive to family income, cost of public transit, and accessibility to bus stop; car usage is sensitive to family income, bicycle/motorcycle ownership, cost of public transit, and driving cost. The elasticity of car ownership rate with respect to family income is approximately 0.5; and the elasticity of annual driving distance with respect to gasoline price is approximately -0.1.

Study 4 in Chapter 7 develops a sequential discrete-continuous choice model to estimate a sequence of decisions on household car ownership and usage over time. In particular, a recursive probit model is formulated to estimate a sequence of vehicle holding decisions, while a regression is used to estimate a sequence of vehicle usage decisions over time. The inherent Gaussian distributed error component of the recursive probit model enables its integration with regressions. Correlation between the discrete and continuous parts, varying over time, is captured with a full unrestricted variance-covariance matrix of the unobserved error components. The sequential discrete-continuous choice model has been validated on simulated datasets of car ownership and usage choices, and is able to reproduce the evolving trends of households' discrete and continuous demands in a real market. The main findings are summarized as follows:

- The proposed model extends the theory of integrated discrete-continuous choice analysis on a temporal basis. To the best of my knowledge, the sequential discrete-continuous choice model is the first to measure the dynamic inter-dependency between discrete choice (i.e. vehicle holdings) and continuous choice (i.e. vehicle usage) over time in the car ownership problem.
- The proposed model in embedded into a dynamic programming framework which facilitates the estimation process and improves the model efficiency.
- Results from simulation experiments suggest that the accuracy of estimated parameters depends on the number of households and the time difference between total study time and agent's forward-looking time. In particular, the accuracy increases as the number of households increases and the time difference decreases.
- The proposed sequential discrete-continuous choice model will serve as an efficient tool to help governments and decision makers to evaluate time-dependent policies and pricing schemes that promote new vehicle technologies and reduce dependency on cars and emissions.

8.2 Future Research

Future works will be directed towards the improvement of the sequential discrete-continuous choice model and its application with real data sets.

First, given the correlation between household decisions on vehicle holding, type, and usage, it would be valuable to consider vehicle type decision in the sequential discrete-continuous choice model. Similar to the integrated discrete-continuous choice model, a multinomial logit model could possibly be used to estimate household vehicle type choice for each time period, and its logsum can be treated as an explanatory variable in the vehicle holding sub-model. In this way, the model will jointly consider the diversity of vehicle types in market over time. In the perspective of automobile producers, this model extension would be valuable for providing dynamic information about consumers' demands on vehicles and travel, their vehicle type preference, willingness-to-pay for improvements of vehicle characteristics, and market elasticities with respect to vehicle sale price and fuel price. The information allows automobile industry to produce an appropriate quantity of each vehicle type and helps to maximize their profit.

Second, although the model has a dynamic structure, it ignores changes in vehicle market. Similar to the case of dynamic discrete choice model, a stochastic diffusion process could be integrated into the framework to jointly capture market evolution. Therefore, the extension of the model is able to predict car ownership, type choice and usage in different economic status (i.e., recessions).

Another limitation of the sequential discrete-continuous choice model is that all estimated coefficients are assumed to be constant over different households.

Therefore, random parameter approach could be integrated into the framework to capture the taste variation among households. Given this information, different incentives and marketing strategies can be applied to encourage people to switch to greener vehicles.

In addition, the error components between the discrete and continuous parts are assumed to be multivariate normal distributed. Although the correlation is captured with an unrestricted covariance matrix, it can be further improved with a more flexible correlation structure. For example, copula-based models allow the combination of any univariate marginal distributions even from different distributional family (Danaher and Smith, 2011; Sun et al., 2017).

Besides, given the high requirements for panel data, the sequential discrete-continuous choice model only estimates on simulated datasets in this study. In the future, the model should be calibrated and applied to a revealed preference panel data containing vehicle ownership and usage decisions over time. For example, household travel survey data and vehicle registration data in France and Netherland. The estimation results on real datasets would provide valuable insights for policy implications in different countries, such as the influences from car ownership tax and gasoline tax in the United States, the impacts of vehicle import tariff and usage restrictions in large cities in China, and influences from carbon taxes in some European countries.

Last but not least, more interesting variables should be considered by the model, such as location of refueling/recharging stations, work location, social network, awareness of emission, public transit services, and policy indicators. It is expected that a richer set of independent variables will improve the ability of the model to capture travel behavior and will provide more valuable insights for policy implications in a short, medium, and long run.

Bibliography

Achtnicht, M., 2012. German car buyers' willingness to pay to reduce CO2 emissions. *Climate Change* 113, 679–697.

Adler, T., Wargelin, L., Kostynivehicles, 2003. A large-scale stated preference experiment. In: 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland, 10–15 August 2003.

Aguirregabiria, V., Mira, P., 2010. Dynamic discrete choice structural models: A survey. J. *Econometrics* 156(1):38–67.

Aizaki, H., 2012. Basic functions for supporting an implementation of choice experiments in R. *Journal of Statistical Software*, 50, pp. 1-24.

Atasoy, B., Glerum, A., Hurtubia, R., Bierlaire, M., 2010. Demand for public transport services: Integrating qualitative and quantitative methods. 10th Swiss Transport Research Conference, Monte Verita, Switzerland, 1–3 Sept 2010.

Axsen, J., Kurani, K.S., 2013. Hybrid, plug-in hybrid, or electric-What do car buyers want? *Energy Policy*, 61, pp. 532-543.

Batley, R.P., Knight, M.J., Toner, J.P., 2004. A mixed logit model of UK household demand for alternative-fuel vehicles. Rivista Internazionale di Economia dei Transporti – International Journal of Transport Economics 3, 55–77.

Bayer, P., McMillan, R., Murphy, A., Timmins, C., 2016. A Dynamic Model of Demand for Houses and Neighborhoods. *Econometrica*. 2016 May 1;84(3):893-942.

Ben-Akiva, M., Abou-Zeid, M., 2007. Hybrid choice models: from static to dynamic. Paper presented at the proceedings of the Oslo Workshop on Valuation Methods in Transport Planning, Oslo, Norway, 19–20 March 2007.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium, *Econometrica*, 63, pp. 841–890.

Bertsekas, D.P., 2005. Dynamic Programming and Optimal Control, 3rd ed., Vol. I (Athena Scientific, Belmont, MA).

Bhat, C.R., Pulugurta, V., 1998. A comparison of two alternative behavioural choice mechanisms for household auto ownership decisions. *Transportation Res. Part B* 32(1):61–75.

Bhat, C., 2000. Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modeling. *Transportation Science*, 34(2), pp. 228-238.

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

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

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

Bolduc, D., Boucher, N., Alvarez-Daziano, R., 2007. Hybrid choice modeling of new technologies for car choice in Canada. *Transportation Research Record* 2082, 63–71.

Brownstone, D., Bunch, D.S., Train K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B*, 34(5), pp. 315-338.

Bunch, D. S., Bradley, M., Golob, T. F., Kitamura, R., Occhiuzzo, G.P., 1993. Demand for clean-fuel vehicles in California: A discrete-choice stated preference pilot project. T*ransportation Research Part A*, *27*(3), pp. 237-253.

Carranza, J.E., 2010. Product innovation and adoption in market equilibrium: The case of digital cameras. Internat. J. Indust. Organ. 28(6):604–618.

Cherchi, E., Cirillo, C., 2010. Validation and forecasts in models estimated from multiday travel survey. *Transportation Research Record: Journal of the Transportation Research Board*. 2010 Dec 1(2175):57-64.

China National Bureau of Statistics, 2009. Available at: http://www.livefrombeijing.com/2009/11/beijings-vehicle-population/.

Cirillo, C., Diana, M. and Dong, H., 2015-2, July. Time spent on social media and its impact on travel behaviours and activity participation: Results from a comparative study based on association analysis and discrete choice models. In IATBR 2015-WINDSOR.

Cirillo, C., Liu, Y. and Maness, M., 2017. A time-dependent stated preference approach to measuring vehicle type preferences and market elasticity of conventional and green vehicles. Transportation Research Part A: Policy and Practice, 100, pp.294-310.

Cirillo, C., Xu, R., 2011. Dynamic Discrete Choice Models for Transportation, *Transport Reviews*, 31:4, 473-494, DOI: 10.1080/01441647.2010.533393.

Cirillo C., Xu R., Bastin F., 2015. A Dynamic Formulation for Car Ownership Modeling. Transportation Science, May 2015.

Dagsvik, J.K., Wennemo, T., Wetterwald, D.G., Aaberge, R., 2002. Potential demand for alternative fuel vehicles. *Transportation Research Part B* 36, 361–384.

Danaher, P.J. and Smith, M.S., 2011. Modeling multivariate distributions using copulas: applications in marketing. Marketing Science, 30(1), pp.4-21.

Dargay, J.M., Vythoulkas, P.C., 1999. Estimation of a dynamic car ownership model: a pseudopanel approach, *Journal of Transport economics and Policy*, 33(3), pp. 287–302.

Daziano, R.A., Waygood, E.O.D., Patterson, Z. and Kohlová, M.B., 2017. Increasing the influence of CO2 emissions information on car purchase. Journal of Cleaner Production, 164, pp.861-871.

de Jong, G. C., 1989a. Simulating car cost changes using an indirect utility model of car ownership and car use, PTRC SAM, Brighton.

de Jong, G. C., 1989b. Some joint models of car ownership and car use; Ph.D. thesis. Faculty of Economic Science and Econometrics, University of Amsterdam.

de Jong, G. C., 1991. An indirect utility model of car ownership and car use. *European Economic Review*, 34(5):971-985.

de Jong, G.C., 1996. Disaggregate model system of vehicle holding duration, type choice and use, *Transportation Research Part B: Methodological*, 30(4), pp. 263–276.

Department of Energy - Vehicle Technology office, 2015. Available at: http://energy.gov/eere/vehicles/vehicle-technologies-office-moving-america-forward-clean-vehicles.

DieselNet, 2016. Available at: https://www.dieselnet.com/standards/cn/ld.php.

Dubé, J.P., Fox, J.T., Su, C.L., 2012. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, September, 2012 1;80(5):2231-67.

EPA, 2013. Greenhouse Gas Emissions from a Typical Passenger Vehicle. Office of Transportation and Air Quality. EPA-420-F-13-033a.

Ewing, G.O., Sarigollu, E., 1998. Car fuel-type choice under travel demand management and economic incentives. *Transp. Res.* 3D, 429–444.

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

Farber S., Paez A., 2009. My car, my friends, and me: a preliminary analysis of auto-mobility and social activity participation. Journal of Transport Geography, Volume 17, Issue 3, May 2009, Pages 216–225, DOI:10.1016/j.jtrangeo.2008.07.008.

Feng Y., Fullerton D., Gan L., 2005. Vehicle Choices, Miles Driven and Pollution Policies. Working Paper 11553, http://www.nber.org/papers/w11553. 2005.

Flamm B., 2009. The impacts of environmental knowledge and attitudes on vehicle ownership and use. Transportation Research Part D: Transport and Environment, Volume 14, Issue 4, June 2009, Pages 272–279, DOI:10.1016/j.trd.2009.02.003.

Flamm B., Agrawal A., 2012. Constraints to green vehicle ownership: A focus group study. Transportation Research Part D: Transport and Environment, Volume 17, Issue 2, March 2012, Pages 108–115, DOI:10.1016/j.trd.2011.09.013.

Fontaras, G., Pistikopoulos, P., Samaras, Z., 2008. Experimental evaluation of hybrid vehicle fuel economy and pollutant emissions over real-world simulation driving cycles. *Atmospheric environment* 42.18 (2008): 4023-4035.

Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*. 2013 Oct 31;56:70-80.

Genz, A., 1992. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics Journal of Computational and Graphical Statistics*, 1(2):141-149.

Gilbert, C.C.S., 1992. A duration model of automobile ownership. *Transportation Res. Part B* 26(2):97–111.

Gillingham, K., Iskhakov, F., Munk-Nielsen, A., Rust, J., Schjerning, B., 2015. A Dynamic Model of Vehicle Ownership, Type Choice, and Usage. Unpublished working paper; 2015 Sep 2.

Glerum A., Frejinger E., Karlstrom A., Hugosson M.B., Bierlaire M., 2013. Modeling car ownership and usage: a dynamic discrete-continuous choice modeling approach. 3rd International Choice Modeling Conference.

Golob, T.F., Bunch, D.S., Brownstone, D., 1997. A vehicle use forecasting model based on revealed and stated vehicle type choice and utilisation data. *Journal of Transport Economics and Policy*, 31(1), pp. 69-92.

Golob, T.F., Kitamura, R., Bradley, M., Bunch, D.S., 1993. Predicting the market penetration of electric and clean-fuel vehicles. *The Science of the Total Environment,* 134(1-3), pp. 371-381.

Goodwin, P. (1992) A Review of New Demand Elasticities with Special References to Short and Long Run Effects of Price Changes. *Journal of Transport Economics and Policy*, Vol. 26, No. 2 (May, 1992), pp. 155-169.

Goodwin, P., Dargay, J., Hanly, M., 2011. Elasticities of road traffic and fuel consumption with respect to price and income: a review. *Transport reviews*, 24(3), pp. 275-292.

Gowrisankaran, G., Rysman, M., 2012. Dynamics of consumer demand for new durable goods. J. Political Econom. 120(6):1173–1219.

Gu, Y., Deakin, E., Long, Y., 2015. The Effects of Driving Restrictions on Travel Demand Evidence from Beijing. Working paper available at: http://yizhengu.weebly.com/uploads/3/7/3/5/37359187/drivingrestrictions_nov2015.pdf.

Hackbarth, A., Madlener, R., 2013. Consumer preferences for alternative fuel vehicles: a discrete choice analysis. *Transp. Res.* 25D, 5–17.

Hanly, M., Dargay, J., 2000. Car ownership in Great Britain—a panel data analysis. Paper presented at the Transportation Research Board 79th Annual Meeting, Washington, DC, January 2000.

Hayashi Y., Kato H., Teodoro R., 2001. A model system for the assessment of the effects of car and fuel green taxes on CO2 emission. Transportation Research Part D: Transport and Environment, 6(2):123–139.

Hensher, D. A., Barnard, P., Smith, N., and Milthorpe, F., 1992. Dimensions of automobile demand; a longitudinal study of automobile ownership and use. North-Holland, Amsterdam.

Hensher, D.A., Greene, W.H., 2001. Choosing between conventional, electric and LPG/CNG vehicles in single-vehicle households. *Travel Behaviour Research*, The Leading Edge (Pergamon Press, Oxford, UK), 725–749.

Hensher, D.A., 1994. Stated preference analysis of travel choices: the state of practice. *Transportation*, 21(2), pp 107-133.

Hidrue, M.K., Parsons, G.R., Kempton, W., Gardner, M.P., 2011. Willingness to pay for electric vehicles and their attributes. *Resource Energy Econ.* 33, 686–705.

Horne, M., Jaccard, M., Tiedemann, K., 2005. Improving behavioral realism in hybrid energy-economy models using discrete choice studies of personal transportation decisions. Energy Econ. 27, 59–77.

Huo, H. and Wang, M., 2012. Modeling future vehicle sales and stock in China. Energy Policy, 43, pp.17-29.

Huo, H., Zheng, B., Wang, M., Zhang, Q. and He, K.B., 2015. Vehicular air pollutant emissions in China: evaluation of past control policies and future perspectives. Mitigation and Adaptation Strategies for Global Change, 20(5), pp.719-733.

Jensen, A.F., Cherchi, E., Mabit, S.L., 2013. On the stability of preferences and attitudes before and after experiencing an electric vehicle. *Transportation Research Part D*, 25, pp. 24-32.

Larson, P. D.,Viáfara, J., Parsons, R.V., Elias, A., 2015. Consumer attitudes about electric cars: Pricing analysis and policy implications. *Transportation Research Part A, 69, pp.* 299–314.

Liu, Y. and Cirillo, C., 2015. Measuring transit service impacts on vehicle ownership and use. Public Transport, 7(2), pp.203-222.

Liu, Y., 2013. Integrated Discrete-Continuous Choice Models: Theory and Applications in Estimating Household Vehicle Ownership, Type and Usage. Ph.D. Dissertation, University of Maryland.

Liu, Y., Tremblay, J. M., Cirillo, C., 2014. An Integrated Model for Discrete and Continuous Decisions with Application to Vehicle Ownership, Type and Usage Choices. *Transportation Research Part A: Policy and Practice*, volume 69, pp. 315–328.

Liu, Y. and Cirillo, C., 2015. Model system to evaluate impacts of vehicle purchase tax and fuel tax on household greenhouse gas emissions. Transportation Research Record: Journal of the Transportation Research Board, (2503), pp.51-59.

Liu, Y., Cirillo, C., 2016. Evaluating policies to reduce greenhouse gas emissions from private transportation. *Transportation Research Part D: Transport and Environment* 44, 219-233.

Liu, Y. and Cirillo, C., 2017. A Generalized Dynamic Discrete Choice Model for Green Vehicle Adoption. Transportation Research Procedia, 23, pp.868-886.

Lohr, S., 2009. Sampling: design and analysis. Nelson Education.

Lorincz, S., 2005. Persistence effects in a dynamic discrete choice model. Application to low-end computer servers. Technical report IEHAS Discussion Papers 0510, Institute of Economics, Hungarian Academy of Sciences, Budapest.

Mabit, S.L., Fosgerau, M., 2011. Demand for alternative-fuel vehicles when registration taxes are high. *Transportation Research Part D*, 16(3), pp. 225-231.

Maness, M., Cirillo, C., 2012. Measuring Future Vehicle Preferences: Stated Preference Survey Approach with Dynamic Attributes and Multiyear Time Frame. *Transportation Research Record: Journal of the Transportation Research Board*, 2285, pp. 100-109.

Mannering, F., Mahmassani, H., 1985. Consumer valuation of foreign and domestic vehicle attributes: Econometric analysis and implications for auto demand. *Transportation Research Part A*, 19(3), pp. 243-251.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15, pp. 447-470.

Melnikov O., 2013. Demand for differentiated durable products: The case of the U.S. computer printer market. *Econom. Inquiry* 51(2):1277–1298.

Miller, R., 1984. Job matching and occupational choice. *Journal of Political Economy* 92, 1086-1120.

Mohammadian, A., Miller, E.J., 2003. Dynamic modeling of household automobile transactions. *Transportation Res. Record* 1831:98–105.

Mokhtarian, P.L., Salomon, I., Redmond, L.S., 2001. Understanding the demand for travel: it's not purely 'Derived'. Innovation: The European Journal of Social Science Research 14 (4), 355–380.

Nadarajah, S. and Kotz, S., 2008. Exact distribution of the max/min of two Gaussian random variables. IEEE Transactions on very large scale integration (VLSI) systems, 16(2), pp.210-212.

Nolan, A., 2010. A dynamic analysis of household car ownership. *Transportation research part A: policy and practice*. 2010 Jul 31;44(6):446-55.

Pakes, A., 1986. Patents as options: Some estimates of the value of holding European patent stocks. *Econometrica* 54, 755-784.

Potoglou, D., Kanaroglou, P.S., 2007. Household demand and willingness to pay for clean vehicles. *Transp. Res*. 12D, 264–274.

Rashidi, T., Mohammadian, A., Koppelman, F., 2011. Modeling interdependencies between vehicle transaction, residential relocation and job change. *Transportation* 38(6):909–932.

Rasouli, S., Timmermans, H., 2013. Influence of Social Networks on Latent Choice of Electric Cars: A Mixed Logit Specification Using Experimental Design Data. *Networks and Spatial Economics*, 13(1), pp. 1-32.

Reich, G., 2013. Divide and Conquer: A New Approach to Dynamic Discrete Choice with Serial Correlation. Available at SSRN 2371592. 2013 Dec 24.

Ross, S.M., 1997. Renewal theory and its application Ross SM, ed. Introduction to Probability Models, 6th ed. (Academic Press, San Diego), 351–409.

Rust J., 1987. Optimal replacement of GMC buses: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033.

Schiraldi P., 2011. Automobile replacement: a dynamic structure approach. *RAND Journal of Economics*, Vol. 42, No. 2, Summer 2011, pp.266-291.

Serulle, N.U., 2015. Transportation Resilience Architecture: A Framework for Analysis of Infrastructure, Agency and Users. Ph.D. dissertation, University of Maryland.

Shapiro, A., Dentcheva, D., Ruszczynski, A., 2009. Lectures on Stochastic Programming (SIAM, Philadelphia).

Shcherbakov, O., 2016. Measuring consumer switching costs in the television industry. *The RAND Journal of Economics*. 2016 May 1;47(2):366-93.

Shen, X., Yao, Z., Huo, H., He, K., Zhang, Y., Liu, H. and Ye, Y., 2014. PM 2.5 emissions from light-duty gasoline vehicles in Beijing, China. Science of the Total Environment, 487, pp.521-527.

Sun, B., Ermagun, A., Dan, B., 2017. Built environmental impacts on commuting mode choice and distance: evidence from Shanghai. Transport. Res. Part D: Transp. Environ.

The World Bank, 2009. Urban Transport and CO2 Emissions: Some Evidence from Chinese Cities. Working Paper, June 2009. Georges Darido, Mariana Torres-Montoya and Shomik Mehndiratta.

Train, K., 1986. Qualitative choice analysis: theory, econometrics, and an application to automobile demand. MIT Press, Cambridge, Mass.

Train, K., 2009. Discrete Choice Methods with Simulation. Published by Cambridge University Press, second edition.

US Energy Information Administration, 2014. Oil and natural gas import reliance of major economies projects to change rapidly. Available at: http://www.eia.gov/todayinenergy/detail.cfm?id=14691.

US Energy Information Administration, 2015. Vehicle standards around the world aim to improve fuel economy and reduce emissions. Available at: https://www.eia.gov/todayinenergy/detail.php?id=23572.

Vaage, K., 2000. Heating technology and energy use: a discrete/continuous choice approach to Norwegian household energy demand. Energy Economics, 22(6), pp.649-666.

Valeri, E. and Cherchi, E., 2016. Does habitual behavior affect the choice of alternative fuel vehicles?. International Journal of Sustainable Transportation, 10(9), pp.825-835.

Vredin Johansson, M., Heldt, T., Johansson, P., 2006. The effects of attitudes and personality traits on mode choice. *Transportation Research Part A* 40, 507–525.

Wang, H., Fu, L. and Bi, J., 2011. CO 2 and pollutant emissions from passenger cars in China. Energy Policy, 39(5), pp.3005-3011.

Wolpin, K., 1984. An estimable dynamic stochastic model of fertility and child mortality, Journal of Political Economy, 92, pp. 852–874.

Wu, M., Li, T. and He, W., 2010. Scheme evaluation of bus lane planning considering environmental impact. J. Transp. Eng. Inf, 8, pp.82-87.

Wu, Y., Wang, R., Zhou, Y., Lin, B., Fu, L., He, K. and Hao, J., 2010. On-road vehicle emission control in Beijing: past, present, and future.

Xinhua News Agency, 2014. Car brands represent status, stereotypes in China. Retrieved on 26th May 2016. Available at: http://www.china.org.cn/business/2014-04/21/content_32166181.htm.

Zhang, S., Wu, Y., Wu, X., Li, M., Ge, Y., Liang, B., Xu, Y., Zhou, Y., Liu, H., Fu, L. and Hao, J., 2014. Historic and future trends of vehicle emissions in Beijing, 1998–2020: A policy assessment for the most stringent vehicle emission control program in China. Atmospheric Environment, 89, pp.216-229.