ABSTRACT

Title of dissertation: DYNAMIC DISCRETE CHOICE MODELS FOR CAR OWNERSHIP MODELING

Renting Xu, Doctor of Philosophy, 2011

Dissertation directed by: Assistant Professor Cinzia Cirillo
Department of Civil & Environmental Engineering

With the continuous and rapid changes in modern societies, such as the introduction of advanced technologies, aggressive marketing strategies and innovative policies, it is more and more recognized by researchers in various disciplines from social science to economics that choice situations take place in a dynamic environment and that strong interdependencies exist among decisions made at different points in time. The increasing concerns about climate change, the development of high-tech vehicles, and the extensive applications of demand models in economics and transportation areas motivate this research on vehicle ownership based on disaggregate discrete choices. Over the next five to ten years, dramatic changes in the automotive marketplace are expected to occur and new opportunities might arise. Therefore, a methodology to model dynamic vehicle ownership choices is formulated and implemented in this dissertation for short and medium-term planning.

In the proposed dynamic model framework, the car ownership problem is described as a regenerative optimal stopping problem; when a purchase is made, the current vehicle state (vehicle age, mileage driven, etc.) is regenerated. The model

allows the estimation of the probability of buying a new vehicle or postponing this decision; if the decision to buy is made, the model further investigates the vehicle type choices. Dynamic models explicitly account for consumers' expectations of future vehicle quality or market evolution, arising endogenously from their purchase decisions.

Both static and dynamic formulations are applied first to simulated data in order to test the ability to recover the true underlying parameters of the synthetic population. Results obtained attest that the dynamic model outperforms the static MNL in terms of goodness of fit, parameters bias and predictive power. In particular, it is found that MNL captures the general trends in choice probabilities, but fails to recover peaks in demand and behavioral changes due to rapidly evolving external conditions.

The extension to a real case study required a data collection effort. A preliminary pilot survey was designed and executed in the State of Maryland in fall 2010; the survey was self-administrated and web-based. Choices were made under the hypothesis that an interval time period of six months passed from a decision to the successive decision and choices over a hypothetical time period of six years were recorded.

Finally, the application of dynamic discrete choice models to vehicle ownership decisions in the context of the introduction of new technology is proposed. Results from the real case study confirm our initial expectations, as the model fit is significantly superior to the fit of the static model.

DYNAMIC DISCRETE CHOICE MODELS FOR CAR OWNERSHIP MODELING

by

Renting Xu

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
2011

Advisory Committee:
Professor Cinzia Cirillo, Chair/Advisor
Professor Anna Alberini
Professor Fabian Bastin
Professor Paul Schonfeld
Professor Lei Zhang

Acknowledgments

This research has been done for three years, and finally received approval from the National Science Foundation in summer of 2011. This precious honor greatly encouraged me to continue with my research on dynamic discrete choice modeling and to believe in myself when facing every challenge.

I would like to express my gratitude to all the people who have contributed towards the completion of this thesis.

I am deeply indebted to my advisor Dr. Cinzia Cirillo whose help, stimulating suggestions and encouragement helped me during all the time of research for and writing of this thesis. At every happy and difficult moment in my past four years, she has always been with me and supportive. She is not only a valuable guide in my research career, but also my good friend.

I would also like to thank Dr. Fabian Bastin from the University of Montreal, who has always offered his guidance and suggestions in my modeling formulation and simulation process. I also appreciate all the help he provided me during the time I was in Montreal and working with him.

Additionally I would like to thank Michael Maness, who has always been helping me with C programming problems. I also thank Jean-Michel Tremblay, who helped to calibrate historical data for the dynamic research.

Finally, I would like to thank my family for their support and company. And I really appreciate my mother for traveling to US and helping me with my life in the most difficult time.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

DDCM	Dynamic Discrete Choice Model
BEV	Battery Electric Vehicles
PHEV	Plug-in Hybrid Electric Vehicles
NTRF	National Road Traffc Forecasts
ALTRANS	ALternative Transport Systems
ORL	Ordered Response Logit
MNL	Multinormial Logit
STM	Sydney Strategic Transport Model
VMM	Vehicle Market Model
DETR	Department of the Environment, Transport and the Regions
RP	Revealed Preference
SP	Stated Preference
MNP	Multinomial Probit
DVTM	Dutch Dynamic Vehicle Transaction Model
GEV	Generalized Extreme Value
OGEV	Ordered Generalized Extreme Value
PCL	Paired Combinatorial Logit
CNL	Cross-nested Logit
MDP	Markov Decision Process
CI	Conditional Independence
RUM	Random Utility Maximization
NTS	National Transportation Survey

1. INTRODUCTION

1.1 Background

The classical economic theory of consumer behavior provides a logically consistent foundation for the empirical analysis of many aspects of individual's choice decision. This realm of behavior involves choice among discrete alternatives, taste variation in the population and individual's choice making procedure.

The origin of discrete choice models is in economics. These statistical procedures describe choices made by people among a finite set of alternatives. Daniel McFadden won the Nobel prize in 2000 for his pioneering work in developing the theoretical basis for discrete choice. In marketing research, discrete choice models can be used to study the consumer demand, to predict market share, and to solve some business related problems, such as pricing and product development. In energy and environmental studies, discrete choice models are utilized to make forecasts (e.g., households' and firms' choice of heating system), and to examine people's choice of fishing or skiing site. Some labor economists use discrete choice models to examine occupation choice, retirement choice, and education or training program choice [Aguirregabiria and Mira, forthcoming, 2009]. In transportation area, planners have been using discrete choice models for decades to predict demand for transportation facilities, travelers' transportation mode, route, destination, and time choice, and even to predict the travelers' one-day-activity [Ben-Akiva and Lerman, 1985, Daly, 1982, Edited by Randolph W. Hall, 2003].

In recent years, the objectives of transportation planning have evolved from adding road and transit capacity to managing travel demand, connecting modes/trips, and reducing emissions. Car ownership models play a central role in the planning and decision making of various public agencies and private organizations: a) The US Department of Energy, b) State Departments of Transportation, c) The auto industry and d) Local transit Agencies [Train, 1979]. The Clean Air Act of 1990 strengthens the role of demand-side policies and requires that MPOs with over 200,000 populations have their planning procedures recertified by USDOT every three years. MPOs are now required to greatly improve their capabilities for modeling travel and land development and the effects of the resultant travel and land use patterns on the economy, environment, and social equity. However, some behaviors are generally missing from the MPOs Transportation Model Systems completely (for which new sub-models must be created); this includes car ownership (number of cars and types per household), which strongly affects trip/tour/chain generation and mode choice [Johnston, 2003]. In an ongoing project started in 2003, the federal government (jointly administered by the Federal Highway Administration (FHWA) and the Federal Transit Administration) provides support to MPOs who wishes to conduct a peer review of their travel modeling. Reviewers frequently suggest improving or including a car ownership model in the transportation modeling system.

Rising oil prices and environmental consciousness - in particular on climate change - are major drivers for the global race of developing and promoting high technology cars. So is the concern over energy security, especially at times of turmoil in the Middle East. Technological advances have also brought electric vehicles closer. Energy prices in the twenties century rose sharply and will rise steadily once the global economy fully recovers and creates a competitive marketplace for alternative energy sources. Besides, state and national governments are interested in adjusting public policy to reduce dependence on foreign oil, decrease air pollution, and combat climate change. Therefore, technology, energy, and policy development create an interesting opportunity for changes in the automotive marketplace over the next five to ten years. The traditional static discrete choice models cannot truly make the prediction of consumer preferences for future vehicles under the expected changes in technology and environment awareness.

Dynamic in car ownership choice, both at intertemporal dimension (resistance to change in ownership levels due to uncertainty of financial position) and intratemporal dimensions (acquired taste for a certain lifestyle) has been studied by researchers in the US, but in many cases their analysis is based on panel surveys collected oversees (often the models are based on the Dutch National Mobility Panel) [Kitamura and BUNCH, 1990]. In the majority of these studies the state variable of current period is influenced by state in the past. However, the state of each period is only represented by the number of cars owned by a household but not by exogenous attributes. A real dynamic framework is therefore necessary when modeling consumer demand that explicitly accounts for consumers' expectations of future vehicle quality, evolving market and consumers' outflow from the car market.

The purpose of this research is to present a dynamic discrete choice model of consumers' car ownership and develop an estimation technique for analyzing the impact of technological changes and the marketing evolution on the dynamics of consumers' demand.

1.1.1 In Economics

A significant portion of the literature focusing on the extension of discrete choice models into a dynamic frame can be found in economics and related fields. In dynamic discrete choice structural models, agents are forward looking and maximize expected inter-temporal payoffs; the consumers get to know the rapidly evolving nature of product attributes within a given period of time and different products are supposed to be available on the market. Changing prices and improving technologies have been the most visible phenomena in a large number of important new durable goods markets. As a result, a consumer can either decide to buy the product or to postpone the purchase at each time period. This dynamic choice behavior has been treated in a series of different research studies.

In his pioneer work, John Rust[Rust, 1987] formalized the optimal stopping problem and estimated the optimal stopping time to replace a used bus engine. In this first dynamic version of McFadden's logit model, a single agent was considered, and random components were assumed to be additively separable, conditionally independent and extreme value distributed. Berry, Levinsohn and Parkes [Berry et al., 1995] - BLP had shown the importance of incorporating consumer heterogeneity for obtaining realistic predictions of elasticities and welfare but their models were static and did not account for the inter-temporal incentives of market participants. In 2000, Oleg Melnikov [Melnikov, 2000] expanded the engine replacement model and released the BLP limitations to model the decision of whether to buy a printer machine or to postpone the purchase based on the expected evolution of the product quality and price. The Melnikov formulation was transferred to model the adoption of other durable goods, such as computers, digital products, etc. [Song and Chintagunta, 2003, Gordon, 2006, Nair, 2007] whose quality was rapidly improving overtime. In the Melnikov's framework, the products were heterogeneous while consumers were homogeneous; error terms were in fact assumed to be independently distributed across consumers, products and time periods; furthermore, the purchase was only made once in the consumers' lifetime. In addition, the parameters of the static problem part were estimated separately from the dynamic part; the participation probability of a consumer was directly obtained from observing the number of purchases in the total market. The estimation of dynamic discrete choice models was computationally costly because the solution of the fixed point problem as defined by Rust was required on all points along the estimation algorithm. In conclusion a three-step method was used to solve the estimation problem. Szabolcs Lorincz [Lorincz, 2005] added a persistent effect to the optimal stopping model which completed the standard optimal stopping problem. This persistence means that customers who already had a product may choose to upgrade it, (i.e. upgrade the operating systems). For this application, the model not only included the likely future quality of the product, but also the industry evolution. These dynamic economic models were generally applied to evaluate price and elasticities, intertemporal substitution and the welfare gains from industry innovations. In 2006, Carranza [Carranza, 2006] examined digital cameras market and proposed a logit utility model with one time purchase; the model incorporated fully heterogeneous consumers and extended standard estimation techniques to account for the dynamics in consumers' characteristics. The model was estimated in a reduced-form specification that was relatively easy to compute. Gowrisankaran and Rysman [Gowrisankaran and Rysman, 2007] also analyzed the importance of dynamics when modeling consumer's preferences over digital camcorder industry products using a panel data set on prices, sales and characteristics. Their model combined the BLP techniques for modeling consumer heterogeneity in a discrete choice context and the Rust techniques for modeling optimal stopping decisions. This model was based on an explicit dynamics of consumer behavior and allowed for unobserved product characteristics, repeated purchases, endogenous prices and multiple differentiated products.

1.1.2 In Transportation

In the transportation field, dynamic models have been widely used for dynamic network equilibrium [Lam et al., 2006]. For transportation demand analysis, a number of dynamic models were proposed and calibrated but they were not based on dynamic optimization. Landau et al. [Landau et al., 1981] defined and tested empirically a framework for trip-generation models sensitive to temporal constraints; households decided whether or not to perform a trip for a specific purpose during the day, and which period was taken. Hirsh et al. [Hirsh et al., 1986] estimated a parametric model of dynamic decision-making process for weekly shopping activity behavior. The individual was assumed to proceed from period to period and the observed weekly activity pattern was the outcome of successive decisions. Action plans were then modified on the basis of actual behavior and of the additional information acquired in previous periods. Liu and Mahmassani [Liu and Mahmassani, 1998] calibrated a day-to-day dynamic model of commuters' joint departure time and route switching decisions that took into account commuters' learning from experience. The analysis provided insight into day-to-day effects of real-time traffic information on user decisions.

Most recently, Train [Train, 2002b] gave the concept of dynamic decision making and described a two/more-periods model in his book Qualitative Choice Analysis, which was very well known amongst demand modelers in transportation. Moshe Ben-Akiva and Maya Abou-Zeid [Ben-Akiva and Abou-Zeid, 2007] proposed a dynamic framework to model the evolution of latent variables and observed choices over time. Their approach involved the integration of discrete choice with Hidden Markov chains which contained behavioral dynamics such as individuals' plans, well-being states and actions. Shortly after, the methodology of Hidden Markov chains was used again to model dynamic driving behavior [Choudhury, 2007]. Choudhury in her MIT PhD thesis studied the effects of unobserved plans for four traffic scenarios: freeway lane changing, freeway merging, urban intersection lane choice and urban arterial lane. These dynamic applications of discrete choice model in transportation focused on the evolution of individuals' previous plans and actions but did not consider the changes in external conditions. Possible applications of dynamic discrete choice models in transportation include modeling car ownership for short and medium-term planning applications; customer choices for dynamic pricing schemes in airline or rail industry; route choice and lane change behavior; weekly (or longer term) activity patterns. Therefore, in transportation the development of dynamic discrete choice models has not been as comprehensive as in economics or marketing.

1.2 Objective of the Research

In transportation, dynamic discrete choice models have not been studied extensively and applications are rather limited; this dissertation aims at widening this gap from the methodological perspective and proposes an application of dynamic discrete choice models on car ownership for short and medium-term planning. This research has multiple objectives:

- Integrate dynamic behavioral processes into discrete choice models.
- Propose a general dynamic framework for car ownership.
- Develop an efficient algorithm to estimate dynamic discrete choice models.
- Extend the framework to heterogeneous consumer problems and evolving product quality.
- Generate an efficient and simple method for collecting real behavioral data over time.

- Validate the superiority of dynamic model to the traditional multinomial logit model.

1.3 Outline of the Dissertation

This dissertation is composed of eight chapters. Chapter 2 reviews car ownership forecasting methodology used by transportation researchers; popular static formulations and dynamic models based on panel data are presented. Chapter 3 discusses dynamic discrete choice models (DDCMs) used by econometricians to forecast the demand for durable products. DDCMs are usually specified as an optimal stopping problem, where agents decide the time period of making a stopping decision. Chapter 4 formulates DDCMs for car ownership forecasting and the dynamic nature of the problem is carefully detailed. Chapter 5 proposes results obtained from a simulated experiment; dynamic and static models are compared in terms of coefficients' bias and prediction power. Chapter 6 presents the methodology adopted for survey design and execution; it reports on the revealed preference experiment and on the three stated choice games corresponding to vehicle technology, fuel choice, and taxation policy. In Chapter 7, the statistical analysis of the sample collected is described. The application of the DDCM based on real data is presented in Chapter 8; results are then compared with those deriving from a static model with equivalent specification. Finally, Chapter 9 summarizes the main findings of this dissertation, presents the main contributions and offers avenues for future research.

2. CAR OWNERSHIP FORECASTING METHODOLOGY REVIEW

Different car ownership models are being used for a wide variety of purposes. National governments (notably the Ministries of Finance) make use of car ownership models for forecasting tax revenues and the regulatory impact of changes in the level of taxation. Regional and local governments (particularly traffic and environment departments) use car ownership models to forecast transport demand, energy consumption and emission levels, as well as the likely impact on this of policy measures. Car manufacturers apply models to the consumer valuation of attributes of cars that are not yet on the market. Oil companies want to predict the future demand for their products and might benefit from car ownership models. International organizations, such as the World Bank, use aggregate models for car ownership by country to assist investment decision-making [Fox et al., 2004]. The estimation of future car ownership and car users' preferences are modeled with demand models, using one of the two possible forms: aggregate or disaggregate. The literature on car ownership forecasting models is reviewed in this chapter with a focus on the disaggregate models and their framework, variables specifications, and estimation methods.

2.1 Aggregate Models

In the year 2004, De Jong published a paper on a comprehensive review of car ownership models. In this paper, the models have been classified into nine types [Fox et al., 2004]. I simplify the classification into: (1) aggregate models, (2) static disaggregate car ownership models (number of car choice model, type choice model), (3) joint discrete-continuous models, (4) Pseudo-panel methods, and (5) Dynamic car transaction models with vehicle type conditional on transaction.

Aggregate car ownership models are mainly of three types: (1) time series models [Tanner, 1981, Dargay and Gately, 1999], (2) cohort models [Algers et al., 1989] and (3) car market models [Leuven, 1989]. Aggregate models no longer appear in academic journals but are still used in practice; their major limitation is the impossibility of modeling vehicle type and use; they also usually include limited socio-demographic variables.

- Aggregate time series models

Aggregate time series models usually contain a sigmoid-shape function for the development of car ownership over time as a function of income or gross domestic product (GDP). The function increases slowly in the beginning (at low GDP per capita), then rises steeply, and ends up approaching a saturation level. Examples are the work done in the late 1980s by Tanner [Tanner, 1981] and in the early 1990s by Button [Button et al., 1993]. They mainly used logistic function. In more recent application, Ingram and Liu [K.Ingram and Liu, 1997] used a double logarithmic specification to explain car ownership; the National Road Traffic Forecasts (NTRF) in the UK [Whelan, 2001, Whelan et al., 2000] applied a logistic curve for saturation, and extended this by including the saturation levels (by household type) to the overall disaggregate tree logit calibration. Dargay and Gately [Dargay and Gately, 1999] used the more flexible function to predict the motorization rate (the number of cars per 1000 persons) on the basis of GDP/capita for a large number of countries. These models had the lowest data requirements, and were attractive for application to developing countries. Income was generally considered to be the main driving force behind car ownership growth.

- Aggregate cohort models

Aggregate cohort models segmented the current population into groups with the same birth year (often five-year cohort), and then shifted these cohorts into the future, describing how the cohorts as they became older, acquired, kept and lost cars. Examples are the models of Van den Broecke [den Broecke/Social Research] for the Netherlands, cohort-based car ownership models in France (Madre and Pirotte, 1991) and Sweden. The Van den Broecke car ownership model was a combination of cohort survival model and an econometric model. The econometric component was used for producing the impact of changes in income on car ownership. Aggregate cohort models were most suited for predicting the impact on car ownership of changes in the size and composition of the population. The demographic force behind car ownership growth can be expected to remain important in Western Europe for another couple of decades.

- Aggregate car market models

Examples of aggregate car market models are Mogridge, the Cramer car ownership model [Cramer and Vos, 1985], Manski [Manski, 1983], Berry [Berry et al., 1995], the TREMOVE model [Leuven, 1989], the ALTRANS model [Kveiborg, 2001], and the software package TRESIS [Hensher and Ton, 2002]. Mogridge distinguished between demand for cars and supply for cars in the car market which was different from aggregate time-series models. Cramer's model was based on time-series data and depended on car prices, income, variation of income and development over time in the utility of using a car. The second hand car price was endogenous. Manski's aggregate car demand and supply model had endogenous used car price on the market as well. Berry modeled the market for new cars only, with consumer demand, oligopolistic manufacturers and endogenous prices which was an innovation in the the most car market models.

TREMOVE was designed to analyze cost and emission effects of a wide range of measures in the European Union to reduce emission from road transport. TREMOVE was a simulation model but not a forecasting model. It was specially used to analyze changes in behavior as a result of changes in economic conditions. ALTRANS (ALternative TRANSport systems) was a model developed for analyzing the environmental impact of different policy proposals on car and public transport usage in Denmark. The software package TRESIS was developed in 2002 for integrated strategic planning of transport, land use and the environment. It included disaggregate models for household fleet size, vehicle type choice and car use. The aggregate car demand of the households by vintage in each year was compared with aggregate supply. The used vehicle prices were used to reach equilibrium and the new vehicle prices were exogenous.

2.2 Disaggregate Static Models

Disaggregate car ownership and type choice models have been extensively developed and applied in the last two decades in several countries: The Netherlands [AVV., 2000, HCG, 1989], Norway [HCG and TOI, 1990], Sydney [Hensher et al., 1992, HCG, 2000], and US [Manski, 1983, Bhat and Pulugurta, 1998]. Their success is due to their behavioral foundations and the possibility of including a large number of policy variables, as well as car types and use.

- Car ownership models

This category is discrete choice models and they deal with the number of cars owned by a household. The car ownership submodel within the Dutch national model system (LMS) for transport [HCG, 1989] is an early example. The car ownership choices of the household were conditioned on household license holding, i.e. household without licenses had zero cars, household with one license chose between zero cars or one, and household with two or more licenses chose between one car or two more cars. The models were binary logit based on random utility theory. Monthly income that a household can freely spend was an important explanatory variable from which the monthly expenditures on food, clothing and housing had already been subtracted. Another important variable was fixed car cost. Therefore, if monthly incomes rose, the probability of car ownership would rise accordingly. If the fixed car costs rose, the car ownership probability would decrease. The rest explanatory variables were age, gender, household size, number of workers in the household and region-specific variables. These household car ownership models in the LMS , combined with personal and household license holding, then influenced tour frequencies and mode/destination in the model system.

Bhat and Pulugurta [Bhat and Pulugurta, 1998] in the year 1998 proposed ordered-response choice mechanism to model car ownership choice; they compared it with unordered-response choice mechanism and both of them applied disaggregate models. Ordered-response choice mechanisms were not consistent with global utility maximization. They were based on the hypothesis that a single continuous variable represented the latent car owning propensity of the household. The decision process can be viewed as a series of binary choice decisions. A given household assigned utility values for each car ownership outcome, and then made an independent utility maximization decision for each range. Only one set of M household parameters needed to be estimated in this approach, but variation in sensitivity to income cannot be specified to vary between alternatives. The ordered-response mechanism was Ordered Response Logit (ORL). Unordered-response mechanisms were consistent with the theory of global utility-maximization. The choice was determined by the alternative with highest utility and the process was simultaneous among alternatives. This method had more parameters to estimate and allowed for variation in sensitivity to household income to vary with car ownership alternative. The unordered-response mechanism was Multinomial Logit. ORL and MNL models were estimated. Three socio-economic variables were significant across the data set: number of working adults, number of non-working adults and household income.

After comparison, it was found that the MNL was superior according to the rooted mean-square error measure. The average probability of correct prediction showed the MNL was superior as well. Therefore, Bhat and Pulugurta concluded that the appropriate choice mechanism for modeling car ownership was the unordered-response structure, such as MNL or probit models.

Hague Consulting Group did the Sydney Strategic Transport Model (STM) [HCG, 2000] in 2000 in which company and total car ownership at the household level were estimated. The data set was from two sources, one collected during 1991/1992 and the other one during 1997/1998. Three approaches were tested in the disaggregate models: modeling private and company car ownership behavior independently, modeling private car ownership conditional on company car ownership, and modeling company car ownership conditional on private car ownership. The results showed that the second approach performed better. The model structure was a two-level MNL model system which had company car ownership models as the upper level and total car ownership models as the lower level. Both the company and total car models were dependent on the logarithm of net household income. The total car model accounted for the impact on the net household income of car ownership cost. The number of license holders in the household was a significant factor. Parking cost was significant negative in the lower car ownership zones since parking was more expensive in those areas. The head of the household was identified with the highest income, and it reflected car ownership differences according to the age and gender. The variable accessibility from the home-work mode-destination model was important as well. It accounted for higher car ownership in the certain zones which were accessible to work places.

The UK Department of Transport made a number of possible improvements for NRTF forecast in the year 1999 [Whelan, 2001, Whelan et al., 2000]. The 1997 NRTF included two binary models for each household, a P_{1+} model to predict the probability of owning at least one car each household, and a $P_{2+|1}$ model that was a conditional probability of owning two or more cars given the household owned at least one car. The improved model was introduced in NRTF-2001 with an additional submodel which was the conditional probability of a household owning three or more cars ($P_{3+|2+|1+}$). Considering the impact of company car ownership on total household car ownership, company car dummies were included into the ownership models. This was consistent with the findings of HCG's work in Sydney described before.

Another example is from Rich and Nielsen [Rich and Nielsen, 2001] who modeled a long-term travel demand for households with up to two workers. The model was specified as a nested logit model with two components: a work model (W-model) modeling the choice of work location and car ownership and a residential location model (R-model) modeling the zone and type of residence. Car ownership was treated within this model structure, but not separately estimated. The W-model was at the bottom of the structure, therefore it was assumed that individuals chose the work location depending on their residential location. Car ownership was modeled as a decision conditional on both residential and work location choice, and the alternatives were zero, one or two cars per household. They did not consider company cars in the models.

- Car-type choice models

This category deals with the choice of car type of the household given car ownership. Hensher et al.[Hensher et al., 1992], Manski and Sherman [Manski and Sherman, 1980] and Train [Train, 1986] have made influential studies. Hensher et al. and Train not only included detail vehicle types, but also included the number of vehicles in the household and car use. Disaggregate models for the number of cars per household had usually been developed to provide inputs for multimodal transport model system, while the car-type choice models form a part of standard models to forecast the size and composition of the car fleet.

Among recently developed car ownership models, some new vehicle type models are described. Page et al.[Page et al., 2000] developed a model of new car sales for incorporation within the Vehicle Market Model (VMM) of the UK Department of the Environment, Transport and the Regions (DETR). Both revealed preference (RP) and stated preference (SP) data were used in the model. RP data contained some household socio-economic characteristics and the attributes of the household's vehicle fleet. The SP data collected information from households that were either planning to acquire a new car, or had just bought a new car. The potential vehicle attributes were presented to respondents that included purchase prices, running costs, resale value, engine size, vehicle emissions, safety measurement, fuel type (petrol, diesel or hybrid petrol-LPG) and fuel economy. The SP and RP data were combined to form two nested models. One model predicted the binary choice between a private and company car. The other one predicted a multinomial choice between different vehicle types. Separate models were used for company and private cars. In the private car model, variables were population density, log of annual household income, log of purchase price, number of children, running costs, variations in emissions, safety features, resale value, fuel economy, standing charges, hybrid engine type and diesel engine type. In the company car ownership model, the variables were population density, log of annual household income, log of monthly cost, number of children, fuel cost, engine size, variations in emissions, safety features, hybrid engine type. There was a scale factor used to scale the SP data relative to the RP data. An interesting result in both models was that in areas with high population densities with scarce parking spaces, there was a higher probability of acquiring a smaller vehicle.

Brownstone et al. [Brownstone et al., 2000] compared multinomial logit (MNL) and mixed logit models for data on California households' RP and SP for vehicle type choice. Before estimating joint SP/RP models, separate SP and RP models were estimated. However, some preference were only identified in the SP while some preference only in the RP. In the joint SP/RP models, a scale factor was used. MNL model showed the scale factor was less than 1, indicating the stochastic error term in the SP data had a larger variance than in the RP dataset. Mixed logit model had the scale factor greater than 1 and its preference heterogeneity was captured by fuel-type error components. The results showed that pure SP models predicted unrealistically high sports car market shares compared with the RP/SP model which demonstrated the superiority of combining RP and SP data. The mixed logit models showed results with higher market shares for the alternative fuel vehicles. Because of the

Independence from Irrelevant Alternatives (IIA) properties of MNL, a proportionate share of each new vehicle must come from all other vehicles, whereas the mixed logit models resulted in more plausible results that the market share for electric fuel vehicles came disproportionately from other mini and subcompact vehicles. Therefore, mixed logit models were feasible for joint RP/SP data.

Hensher and Green [Hensher and Greene, 2000] estimated both MNL and mixed logit models with combined RP/SP data for vehicle choice. In the SP survey, vehicles were categorized according to the following attributes: three size categories based on engine size (within a given engine size, respondents were asked to indicate a preferred body type), price of vehicle, registration fee, fuel cost to travel 500 km (variable described as approximate cost of filling a tank so respondents understood levels), fully fuelled range, acceleration and boot size. The SP survey followed a two-stage process. The household member was required to consider three conventionally fuelled vehicles (one from each size class) and choose one in the first stage. In the second stage, three electric vehicles and three alternative fuel vehicles were added to the choice set, and the respondent was asked to choose one vehicle from the nine options. This process was repeated three times. The RP model was defined by a 10-alternative choice set, using a random sampling procedure within each size class to assign vehicles of each vintage to the 10 alternatives given their size class. One nested logit and three mixed logit models were estimated. In the mixed-logit models, random parameters were estimated for the electric and alternative fuel vehicle constants (normally distributed), and for the vehicle price (log-normally distributed to ensure the parameter is alwasnegative). After comparing nested MNL and the mixed logit formulation, MNL was found to over allocate to new fuels market shares and therefore underestimate shares on the conventionally fuelled classes relative to mixed-logit models.

2.3 Joint Discrete-continuous Models

Part of Train's models [Train, 1986] for California and Hensher's et al. [Hensher et al., 1992] for Sydney and the models of De Jong [Jong, 1989a,b, 1991] for the Netherlands belong to this category. These models explain household car ownership and car use in an integrated micro-economic framework.

De Jong [Jong, 1989a] developed two disaggregate models in his studies. Both of his models explained whether or not a household would own a private car, and the number of kilometers driven per year conditional on car ownership. The idea from his models was that decision of household on car ownership and car use were strongly interrelated and should be studied together. Both models were joint discrete-continuous models. The first model was called the "statistical model" and it was under the situation without major policy changes. It assumed that a household had a desired annual kilometrage, which depended on attributes of the household. When this desired kilometrage exceeded a threshold, the household would own a car. The observed kilometrage can deviate from the desired kilometrage through a random disturbance term. Explanatory variables for both models were household income, household size, age, gender and occupation of the head of the household.

De Jong's second model was the "indirect utility model" [Jong, 1989b]. Train

[Train, 1986] and Hensher et al.[Hensher and Greene, 2000] used this model in their studies after that. This model was based on micro-economic theory. The basic idea was that households compared combinations of car ownership and car use with each other and chose the combination that gave the highest utility. Fixed car cost and variable car cost were incdlued besides the variables that were in the statistical model.

Train [Train, 1986] and Hensher et al. [Hensher and Greene, 2000] developed similar "indirect utility" equations for car ownership and annual kilometrage, but embedded these models in a larger framework which also contained car type choice, conditional on car ownership. Hensher et al.'s model system was developed based on panel data for Sydney and contained both static and dynamic vehicle choice and use models.

2.4 (Pseudo)-panel methods

- Panel models

Panel data has been used since 1980s. Early in the 1987, Kitamura [Kitamura, 1987] developed an integrated model simultaneously determining car ownership and the total number of trips in a week. The model contained lagged effects. All the equations were linear. The data set consisted of the first waves from the Dutch National Mobility Panel (LVO). 10 waves were collected between March 1984 and March 1989. Kitamura and Bunch [Kitamura and BUNCH, 1990] used four waves of the same LOV panel data set to develop an ordered-response probit model for the car ownership per household. They included lagged variables to account for state dependence and individual-specific error components to account for unobserved heterogeneity across households. Meurs [Meurs, 1991] also had car ownership models estimated on the panel data of LVO. The models included linear simultaneous equation models of car ownership and use, discrete choice car ownership models, and joint car ownership and mobility models [Meurs, 1993]. Income was used as the variable but car cost variables were not included.

Here are some recent panel models. Nobile [Nobile et al., 1996] estimated a random effect multinomial probit (MNP) model of car ownership level with longitudinal data collected in the Netherlands. Nobile et al. noted that panel data enabled the incorporation of both intertemporal dimensions and intratemporal dimensions. The data source for modeling was drawn from Dutch National Mobility Panel. Waves 3, 5, 7 and 9 of the period were analyzed. The approach used for estimation was Bayesian: a prior distribution of the parameters of the longitudinal MNP model was specified and the posterior was examined using Markov chain Monte Carlo methods. A total of 50,000 draws were used for the Markov chain, with an initial burn-in of 5000 draws excluded to ensure that the Markov chain had stabilized. The results showed wave dummies were all negative, suggesting generic temporal effects. In the cross-sectional terms, standard disaggregate household model term were estimated for one and two or more car alternatives with no cars as the base. These terms included the level of urbanization, number of licenses in the household, number of full and part-time workers, number of adults, number of children, and household income.

Hanly and Dargay [Hanly and Dargay, 2000] used 4-year panel data from the British Household Panel Survey. This panel model had dependent variable, the number of cars owned per household in each year. The dependence on past experience was incorporated by introducing lagged endogenous variables. Three types of models were estimated: a model without a lagged dependent variable, a model with a lagged dependent variable and a model with dummies for the number of cars in the last year.

Golounov et al.[Golounov et al., 2001] in the year 2002 first developed a theoretical model for the purchases and consumption of cars, other durable goods and other day-to-day and long-term purchases. They stated that most existing dynamic car ownership models (panel models, cohort models, duration models) did not have a strong theoretical underpinning. Another theoretical foundation for dynamic ownership and replacement model is from John Rust [Rust, 1987] who combined utility theory from micro-economics with optimal stopping process decision-making rules from dynamic programming. His application concerned the replacement of bus engines in a single agent over time.

- Pseudo-panel models

The pseudo-panel method is a relatively new econometric approach to estimate dynamic transport demand models that circumvents some problems of panel data such as attrition. A pseudo-panel is an artificial panel based on cohort averages of repeated cross-sections. There are some restrictions imposed on pseudo-panel data. One of the important is that the cohorts should be based on time-invariant characteristics of the households, such as the birth year of the head of the household. The cohorts should have homogeneity within them and heterogeneity between them. Another important feature of pseudo-panel data is that averaging over cohorts transforms disaggregate values of variables into cohort means losing information about the individuals.

Dargay and Vythoulkas [Dargay and Vythoulkas, 1999a] used the pseudo-panel dataset of 5-year cohorts constructed from repeated cross-section data contained in the UK Family Expenditure Survey. Their model was a fixed effect model but resulted in an error-in-variables estimator. A generation effect was added to the model proposed by Deaton and a lagged dependent variable was included to estimate the dynamics of the model. There were three other models estimated to compare with the fixed effect model: OLS, random effect specification and random effect with a first-order autoregressive scheme. The dependent variables was the number of cars per household and it indicated the average number of cars for that particular cohort. The explanatory variables were socio-economic characteristics of the household such as number of adults and children, income, metropolitan and rural areas, and a generation effect for the head of the household. Car purchase costs, car running costs and public transport fares were also included. Dargay and Vythoulkas [Dargay and Vythoulkas, 1999b] had another paper which extended the previous paper by defining the pseudo-panel observations not only as 5-year cohorts, but also in terms of area type (e.g. rural, urban).

2.5 *Dynamic car transaction models*

Hocherman et al.[Hocherman et al., 1983], Smith et al.[Smith et al., 1989] and Gilbert[Gilbert, 1992] did some early studies on vehicle transaction models. Hocherman et al. used a nested logit model for vehicle transactions and the conditional vehicle type choice. The transaction options for a zero-car household were purchasing a car or doing nothing; for a one-car household, the options were replacing or doing nothing. For the purchase and replace options, there were type choice models. Smith et al. only studied the transaction of one-car households. Gilbert used duration models to explain car ownership duration. Bunch et al.[Bunch et al., 1996] and the Dutch Dynamic Vehicle Transaction Model (DVTM) are the most recent examples. Duration models in these models determine whether a household will make a purchase. Vehicle type model is used if a transaction is made.

Bunch et al.'s model for California contained transaction models for adding a car, disposing a car and replacing a car for single- and multivehicle households. The overall dynamic simulation system also included the type choice models from Brownstone et al. [Brownstone et al., 2000] and car use equations.

The DVTM model was developed and tested by the Hague Consulting Group. The main objective of the modeling was to extend the static disaggregate modeling approach for the size and composition of the car market into the dynamic models. The DVTM consisted of four submodels. Hazard-based duration models explained the time that elapsed between two household vehicle transactions. They used continuous time and were intrinsically stochastic models. Several functional forms they used in the models were exponential, Weibull, and log-normal. Vehicle type choice models in this study were for households replacing or extending their fleet. Vehicle types were distinguished by brand, model and vintage. For each brand-model-vintage combination, the engine size, weight, average fuel efficiency, fuel type, type of catalytic converter and fixed and variable cost were known. MNL model was used. Model for annual car use was similar to the indirect utility model (discrete-continuous model). Model for style of driving, the last submodel determined a possible deviation from the average fuel efficiency.

In the dynamic vehicle transaction model such as the DVTM or Bunch et al.'s model for California, the number of cars per household was predicted based on current car ownership of the household. The duration model predicted the time (e.g. months) until the next vehicle transaction and the type of transaction (e.g. replacement, disposal, adding a car). Time was discrete in this model. Households that did not transact in year t would have the same vehicle ownership in year $t+1$ as in year t. Households that had transactions involved replacing a car or adding a car, the conditional type choice model would therefore be used to get new type choice probabilities. The duration model then could be used to predict transactions each time based on the car ownership situation of the previous year. Meanwhile, vehicle scrappage transactions could be integrated in the model.

For both duration model and a panel model of vehicle transactions, short term predictions (up to 5 years ahead) might be done without updating the population in the sample used. But for medium and long term forecasts, the population needs to be updated.

The discrete vehicle type choice model was applied conditional on specific vehicle transactions in the DVTM. The choice alternatives were the brand-model-vintage combinations and there were about 1000 distinguished alternatives. In addition, average emission rates and fuel consumption for the brand-model-vintage combination can be used to give outcomes on these variables.

2.6 Summary

This chapter reviewed car ownership models with a classification into five types: aggregate models, static disaggregate models, joint discrete-continuous models, (pseudo)panel models and dynamic models. Table 2.6 compared the car ownership model types discussed above on the basis of 16 criteria proposed by De Jong in 2004 [Fox et al., 2004].

The aggregate models which included time series models, cohort models and car market models could not model vehicle type and use and they lacked a lot of variables. Therefore, the aggregate models were not the right type for the development of a fully fledged car fleet model. They can only predict the total number of cars in the medium and long term and then used the results as a starting point in other models. However, when the data were very scarce, aggregate time series models might be the only method available for forecasting.

The static car ownership models and discrete car-type choice models were suitable for a long-term prediction to forecast the number of cars and the distribution over households and car types. Their advantages compared to the aggregate models were the possibility of including a large number of policy variables, cost and price variables with RP and SP data. Car-type choice models were predicted given car ownership. After comparing nested MNL and mixed logit formulation, mixed logit was found to reasonably allocate the car-type market shares.

Joint discrete-continuous models explained household car ownership and car use in an integrated micro-economic framework. The idea from these models were that decision of household on car ownership and car use were strongly interrelated and should be studied together.

Discrete car type choice models can be added to panel models for the transitions between car ownership states of households. The panel models could then be used to give the evolution of the fleet from the present fleet. For medium- and long-term forecasts, panel models can be carried out when changes in the size and composition of the population need to be predicted. Pseudo-panel models provide an convenient way to get short- and long-term policy-sensitive forecasts of the car ownership based on cohort averages of repeated cross-sections. But the restrictions of losing information about the individuals determine pseudo-panel models cannot take over the role of a choice-based model for the number of cars and car type.

Dynamic transaction models included duration models for determining whether a household would make a purchase. These dynamic models had been combined with detailed policy-sensitive type choice models to predict brand-model-vintage combination. For long-term forecasts, as for panel models, population needed to be updated. Long-term changes in the supply of car types could be simulated through scenarios.

Aspect	Aggregate time series models	Cohort models	Aggregate market models	Static disaggregate ownership models	Static disaggregate type choice models	Joint discrete-continuous models	Panel models	Pseudo-panel models	Dynamic transaction models
Level of aggregation	Aggregate	Aggregate	Aggregate	Disaggregate	Disaggregate	Disaggregate	Disaggregate	Aggregate	Disaggregate
Dynamic or static	Dynamic	Dynamic	Dynamic	Static	Static	Static	Dynamic	Dynamic	Dynamic
Long or short-run forecasts	Short, medium and long	Medium and long	Short, medium and long	Long	Long	Long	Short and long	Short and long	Short and medium
Theory	N/A	N/A	Economic market equilibrium	Can be based on random utility theory	Can be based on random utility theory	micro-economic theory	Can be based on random utility theory or lifetime utility theory	Weak links with random utility theory	Parts can be based on random utility
Data requirements	Light	Light	Light	Moderate	Heavy	Heavy	Very heavy	Moderate	Very heavy
Car types	No car types	No car types	Limited car types	Very limited	Many car types	Very limited	Very limited but could be combined with a type choice model	Very limited	Very limited number in duration model, but very many in car type choice model
Impact of income	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Impact of car cost	Fixed and or variable cost sometimes included	None	Fixed and variable	Fixed cost often included;log-sum includes variable cost	Purchase cost and fuel efficiency often included	Fixed and variable	No policy runs reported, but might be possible	Fixed and variable	Fixed and variable
Impact of license holding	No	Yes	Yes	Possible	No	Possible	No,but possible	No,but possible	No,but possible
Socio-demographic impacts	Limited	Many possible	Limited	Many possible	Many possible	Many possible	Many possible	Limited	Many possible
Scrappage included	No	No	Can be included	No	No	No	Can be included	No	Can be included

Tab. 2.1: Comparison of types of car ownership models

3. DYNAMIC DISCRETE CHOICE MODELS REVIEW

A significant portion of the literature focusing on the extension of discrete choice models into a dynamic frame can be found in economics and related fields. In dynamic discrete choice structural models, agents are forward looking and maximize expected inter-temporal payoffs; the consumers is aware of the rapidly evolving nature of product attributes within a given period of time and different products are supposed to be available on the market. Changing prices and improving technologies have been the most visible phenomena in a large number of important new durable goods markets. This chapter provides a review of dynamic theory and its application in economics, with a special focus on the combination of behavioral dynamics and discrete choice. Successively, possible applications in transportation are discussed. Finally, conclusions and the avenues for future research opportunities in transportation are presented.

3.1 Discrete Choice Models and the Dynamics

Discrete choice models based on Random Utility Maximization (RUM) theory have been of interest to researchers for many years in a variety of disciplines. These methodologies are used to analyze and predict individual choice behavior. Classical formulations assume that utilities are linear, additive and include both individual characteristics and alternative attributes. The multinomial logit (MNL) [Ben-Akiva and Lerman, 1985] model has been the most widely used structure for modeling discrete choices in travel behavior analysis. Nested logit (NL) model [Daly, 1982] relaxes in part MNL model assumptions; it is derived from McFadden's [McFadden, 1978] generalized extreme value (GEV) model. Other relaxations of the MNL model, designed to consider similarity between pairs of alternatives, have been derived from McFadden's GEV model as well. These include the ordered generalized extreme value (OGEV) model [Small, 1987], the paired combinatorial logit (PCL) model [Chu, 1981, 1989] and the cross-nested logit (CNL) model [hua Wen and Koppelman, 2001, Abbe et al., 2007, Papola, 2000]. Non-closed form discrete choice models as Probit [Daganzo et al., 1977] and Mixed logit [McFadden and Train, 2000] have been adopted by researchers to deal with heterogeneity over consumer preferences, correlation across alternatives and state dependency. All these models have been mainly developed in a static context. However, the static framework is limited by the assumption that consumers are not affected by past and future states when choosing their preferred alternative in the present. The gap between discrete choice model and dynamics in individual behavior has spurred various developments that are mainly intended to enrich the basic theory by including in the formulation the changes occurring in the system to be modeled.

A significant portion of the literature focusing on the extension of discrete choice models into a dynamic frame can be found in economics and related fields. In dynamic discrete choice structural models, agents are forward looking and maximize expected inter-temporal payoffs; the consumers is aware of the rapidly evolving nature of product attributes within a given period of time and different products are supposed to be available on the market. Changing prices and improving technologies have been the most visible phenomena in a large number of important new durable goods markets. Although sometimes the future effects are not fully known, or depend on factors that have not yet transpired, the person knows that in the future, he/she will maximize utility among the alternatives that are available at that time. This knowledge enables him/her to choose the alternative in the current period that maximizes his expected utility over the current and future periods [Train, 2002a]. As a result, a consumer can either decide to buy the product or to postpone the purchase at each time period.

This dynamic choice behavior has been treated in a series of different research studies and the modeling procedures were applied in various areas, such as Wolpin's model for women's fertility probability[Wolpin, 1984], Pakes' model about patent options [Pakes, 1986], and Wolpin's model on job search [Wolpin, 1987]. John Rust formalizing the optimal stopping problem and estimating the optimal stopping time to replace a used bus engine have been considered as a breakthrough on dynamic modeling in both transport and economic fields. In this dynamic version of McFadden's logit model, a single agent was considered, random components were assumed to be additively separable, conditionally independent and extreme value distributed. Berry, Levinsohn and Parkes [Berry et al., 1995] - BLP had shown the importance of incorporating consumer heterogeneity for obtaining realistic predictions of elasticities and welfare but their models were static and did not account for the intertemporal incentives of market participants. In 2000, Oleg Melnikov expanded the engine replacement model and released the BLP limitations to model the decision of whether to buy a printer machine or to postpone the purchase based on the expected evolution of the product quality and price. The Melnikov formulation was transferred to model the adoption of other durable goods, such as computers, digital products, etc. [Song and Chintagunta, 2003, Gordon, 2006, Nair, 2007] whose quality was rapidly improving overtime. Szabolcs Lorincz [Lorincz, 2005] added a persistence effect to the optimal stopping model which completed the standard optimal stopping problem. This persistence means that customers who already had a product may choose to upgrade it(i.e. upgrade the operating systems). For this application, the model not only included the likely future quality of the product, but also the industry evolution. These dynamic economic models were generally applied to evaluate price and elasticities, intertemporal substitution and the welfare gains from industry innovations.

In 2006, Carranza examined the digital cameras market and proposed a logit utility model with one time purchase [Carranza, 2006]; the model incorporated fully heterogeneous consumers and extended standard estimation techniques to account for the dynamics in consumers' characteristics. The model was estimated in a reduced-form specification that was relatively easy to compute. Gowrisankaran and Rysman also analyzed the importance of dynamics when modeling consumer's preferences over digital camcorder industry products using a panel data set on prices, sales and characteristics [Gowrisankaran and Rysman, 2007]. Their model combined the BLP techniques for modeling consumer heterogeneity in a discrete choice context and the Rust techniques for modeling optimal stopping decisions. This model was based on an explicit dynamics of consumer behavior and allowed for unobserved product characteristics, repeated purchases, endogenous prices and multiple differentiated products.

In the transportation field, dynamic models have been widely used for dynamic network equilibrium [Lam et al., 2006]. For transportation demand analysis, a number of dynamic models were proposed and calibrated but they were not based on dynamic optimization. In transportation the development of dynamic discrete choice models has not been as comprehensive as in economics or marketing.

3.2 Markov Decision Process and Dynamic Discrete Choice Structure

3.2.1 Theory of Dynamics

According to the formulation proposed by John Rust in 1987, any dynamic problem can be formulated as a Markov decision process (MDP) in which two components should be defined at each discrete period and for each individual: (1) a vector of system state variable s_t and (2) an action or decision variable d_t. The state and action determine current utility $u(s_t, d_t)$ and affect the distribution of the next period's state s_{t+1} via the Markov transition probability $p(s_{t+1}|s_t, d_t)$. In each period t, the individual maximizes the expected utility $V(s) = \max E(\sum_{t=0}^{\tau} \beta^{\tau} u(s_t, d_t)|s_0 = s)$ and decides the optimal decision rule d . In this equation, E denotes expectation with respect to the controlled stochastic process s_t, d_t and $\beta \in (0, 1)$ is the discount factor. By applying the Bellman's principle of optimality the value function can be obtained using a recursive procedure:

$$V(s_t) = \max_{d \in D(s_t)} [u(s_t, d_t) + \beta \int V(s_{t+1}) p(ds_{t+1}|s_t, d_t)] \tag{3.1}$$

and the optimal decision rule is obtained from V by finding a value $d(s) \in D(s)$ that attains the maximum utility in equation (3.1) for each s (Rust, 1994 draft)[Rust, 1994].

$$d_t(s_t) = \arg \max_{d \in D(s_t)} [u(s_t, d_t) + \beta \int V(s_{t+1}) p(ds_{t+1}|s_t, d_t)] \tag{3.2}$$

3.2.2 Dynamic Discrete Choice Models

Dynamic discrete choice models describe the behavior of a forward-looking agent who chooses among some available alternatives repeatedly over time and intends to maximize expected inter-temporal payoffs. The parameters in the dynamic function describe agents' preferences and beliefs about technological and institutional constraints, and the whole utility function contains both the static parameters and the transition probabilities. The ultimate objective is to estimate the structural parameters in preferences, state transition probabilities and the discount factor β.

The application of dynamic discrete choice models in economics are intended for the consumer i to decide whether to buy a product or not at time t, that is the consumer chooses one of J_t products in period t or chooses to postpone buying. From these J_t choices, the consumer chooses the alternative which maximizes the sum of the expected discounted value of utilities at time $t+1$ conditional on the information at time t . Generally, product j is characterized by observed static characteristics x_j, dynamic characteristic y_{jt} (such as price) and unobserved characteristics ξ_j (e.g. policy, technology innovation). Consumer preferences over x_j and y_{jt} are defined respectively by coefficients α_i^x and α_i^y which need to be estimated with ξ_j. It is assumed that x_j and ξ_j stay constant over infinite life of the product. In each period, the consumer obtains a utility from the product that has just been purchased or from the product that has already been owned. The utility function of discrete choice from product j purchased at time t can be generalized as

$$u_{ijt} = \alpha_i^x x_j + \alpha_i^y y_{jt} + \xi_j + \epsilon_{ijt} \tag{3.3}$$

ϵ_{ijt} is an individual-specific random term depending on the individual i , the product j and the time period t . It is usually assumed that ϵ_{ijt} is distributed type I extreme value, independent across consumers, products and time.

The consumer i will decide to buy a product at time period t when the maximum utility is greater than a specific utility which will depend on the expected evolution of products' quality and prices in the future. Let $v_{it} = \max_j u_{ijt}$ denotes the maximum utility consumer i can get from any product purchased at time t . The reservation utility is the value of not purchasing anything at current time period t and postponing until the next period $t+1$ when the individual evaluates the problem again. The reservation utility could be written as:

$$V(\Omega_{it}) = \beta E[\max\{v_{i,t+1}, V(\Omega_{i,t+1})\}\,|\Omega_{it}] \tag{3.4}$$

where Ω_{it} is a vector of sufficient statistics for the distribution of v_{it} and its Markov transition probability. The specific settings of $V(\Omega_{it})$ might differ depending on the specific application considered while the estimation methods used are mostly based on Rust's nested fixed point maximum likelihood algorithm. Both specification and estimation will be discussed in the following Sections.

3.3 Discussion by Model Type

3.3.1 Rust Optimal Stopping Problem

Modeling framework

An early example of dynamic framework for agent decisions is the optimal stopping model proposed by John Rust in 1987 and applied to the problem of bus engine replacement. This work is the basis for later dynamic studies [Melnikov, 2000, Lorincz, 2005, Carranza, 2006, Gowrisankaran and Rysman, 2007]. In this specific case, the optimal stopping rule is defined as "whether or not to replace the current bus engine" in each period and based on observed and unobserved variables. The stochastic dynamic problem formalizes the trade-off between the conflicting objectives of minimizing maintenance costs versus minimizing unexpected engine failures. Rust's framework focuses on two ideas: (1) a "bottom-up" approach for modeling the replacement problem and a (2) "nested fixed point" algorithm for estimating dynamic programming models in the presence of discrete choices.

The bottom-up approach generates replacement investments by aggregating single replacement demands for some specific capital goods such as bus engine

(Rust's case is to aggregate all the models of bus engine). The demand is the sum of a large number of stochastic processes, each characterized by a decision variable d_t , where $d_t = 1$ if a replacement occurs and 0 otherwise, and by a state variable s_t which is the mileage cumulated by the bus engine at time t. At each time period the agent faces the following discrete decisions: (i) perform normal maintenance on the current bus engine and incur operating cost $c = (s_t, \theta_1)$[1] or (ii) cannibalize the old bus engine for scrap value P and install a new bus engine at cost $\overline{P}$ and incur operating cost $c = (0, \theta_1)$. It is also assumed that the mileage travelled each month is exponentially distributed with parameter θ_2. Besides, there are still some variables that can be observed by the agent but not by the econometrician, a solution is to add an error term ϵ_t to the utility function $u(s_t, d_t, \theta) + \epsilon_t(d)$ which realizes single period utility value when alternative d is selected and the state variable is s_t, $\theta = \{\theta_1, \theta_2\}$.

Suppose the vector of state variables obey a Markov process with transition density given by a parameter function $\pi(s_{t+1,\epsilon_{t+1}} | s_t, \epsilon_t, d_t, \theta)$. The behavioral hypothesis is that agent chooses a decision rule to maximize his expected discounted utility over an infinite horizon where the discount factor $\beta \in [0, 1)$. The solution to this optimal stopping problem is given by the recursive Bellman's equation:

$$V_\theta(s_t, \epsilon_t) = \max_{d_t \in D(s_t)} [u(s_t, d_t, \theta) + \epsilon_t(d_t) + \beta E V_\theta(s_t, d_t, \epsilon_t)] \tag{3.5}$$

[1] Costs are in general not directly observable, so they are inferred from observations. In Rust case study a total cost function is estimated with parameter θ_1

where the utility function u is given by:

$$u(s_t, d_t, \theta_1) = \begin{cases} -c(s_t, \theta_1) + \epsilon(0) & \text{if } d_t = 0 \\ -[\overline{P} - P + c(0, \theta_1)] + \epsilon(1) & \text{if } d_t = 1 \end{cases} \tag{3.6}$$

In function (3.5), $V_\theta(s_t, \epsilon_t)$ is the maximum expected discounted utility obtained by the agent when the state variable is (s_t, ϵ_t). The expected function EV_θ is defined by

$$EV_\theta(s_t, \epsilon_t, d_t) = \int V_\theta(s_{t+1}, \epsilon_{t+1})\pi(ds_{s+1}, d\epsilon_{t+1}|s_t, \epsilon_t, d_t, \theta) \tag{3.7}$$

The transition probability defines the regeneration property through evolution of the mileage variable s_t:

$$p(s_{t+1}|s_t, d_t, \theta_2) = \begin{cases} \theta_2 exp\left\{\theta_2(s_{t+1} - s_t)\right\} & \text{if } d_t = 0, s_{t+1} \geq s_t \\ \theta_2 exp\left\{\theta_2(s_{t+1})\right\} & \text{if } d_t = 1, s_{t+1} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

With all the functions defined above, it is concluded that Section by saying that (s_t, d_t) is a realization of a controlled stochastic process whose solution is an optimal decision rule d_t that attains the maximum in Bellman's equation (3.5). The objective is to use the observed data to infer the unknown parameter vector θ.

Estimation

Maximum likelihood is the method used to infer the unknown parameters and to derive the probability density function $L(s_1, ..., s_T, d_1, ...d_T|\theta)$ from the data and to compute the estimate $\hat{\theta}$ which maximizes the likelihood function. Rust set Conditional Independence (CI) Assumption yielding a simple formula for the likelihood function so the procedure to compute (3.5) is substantially simplified.

$$\text{CI } \pi(s_{t+1}, \epsilon_{t+1}|s_t, \epsilon_t, d_t, \theta) = p(s_{t+1}|s_t, d_t, \theta)q(\epsilon_t|s_t, \theta) \tag{3.9}$$

CI limits the pattern of dependence in (s_t, ϵ_t) in two ways. First, s_{t+1} is a sufficient statistic for ϵ_{t+1} so that any statistical dependence between ϵ_t and ϵ_{t+1} is transmitted entirely through the vector s_{t+1} . Second, the probability density for s_{t+1} depends only on s_t and not on ϵ_t . If it is assumed that q yields some specific functional form such as multivariate extreme value distribution, the likelihood function can be written as:

$$L(s_1, ...s_T, d_1, ...d_T|\theta) = \prod_{t=1}^{T} P(d_t|s_t, \theta)P(s_t|s_{t-1}, d_{t-1}, \theta) \tag{3.10}$$

Where the conditional choice probability $P(d|x, \theta)$, is given by the standard multinomial logit formula:

$$\frac{\exp u(s, d, \theta) + \beta EV_\theta(s, d)}{\sum_{d' \in D(x)} \exp u(s, j, \theta) + \beta EV_\theta(s, j)} \tag{3.11}$$

where EV_θ is the fixed point to the contraction mapping $T_\theta(EV_\theta)$ computed by:

$$EV_\theta(x,d) = T_\theta(EV_\theta)(s,d) \equiv \int log \left[\sum_{d^{'} \in D(s^{'})} \exp \left\{ u(s^{'}, d^{'}, \theta) + \beta EV_\theta(s^{'}, d^{'}) \right\} \right] p(ds^{'}, d, \theta) \tag{3.12}$$

T_θ is a contraction mapping and EV_θ is the unique solution to (3.12). To conclude, the nested fixed point optimization finds a θ that maximizes the likelihood function (3.10). Further details about the optimization algorithm can be found in Rust, 1988.

3.3.2 Melnikov Demand Model for Differentiated Durable Products

The bus engine replacement problem only describes one single agent's choosing behavior that limits the application of dynamic discrete choice models. Another example of dynamic demand framework is the Melnikov's model for computer printers. The computer hardware market is similar to many other high-technology product markets; the quality rapidly improves over time and product durability impacts the evolution of prices and sales [Melnikov, 2000]. In Melnikov's model, only one purchase is made; this is the same assumption made by Rust in his optimal stopping problem. Furthermore, all consumer heterogeneity is captured by a term that is independently distributed across consumers, products and time. The significant difference between the two approaches is that Melnikov mainly deals with differentiated durable products rather than homogenous products (i.e. the bus engine in Rust's example). The framework is divided into three parts: consumer optimal stopping problem, industry evolution and sales dynamics and aggregation.

Consumer Optimal Stopping Problem

The consumer optimal stopping problem gives a general formulation of this choice decision. In each period t , consumer i has two options, $S_{it} = \{0, 1\}$. $s_{it} = 0$ means i does not own any product at t; $s_{it} = 1$ otherwise; in the latter case consumers are out of market. In each period t consumers who have no product either choose to buy one of the products j or to postpone the purchase until the optimal time. If the consumer buys a product, the terminal payoff which is the utility when the consumer decides to buy is:

$$u_{ijt} = f(x_j, y_{jt}; \theta_i) + \epsilon_{ijt} \tag{3.13}$$

where x_j is a vector of static product attributes for product j, y_{jt} is a vector of dynamic characteristics such as price for product j at time t, θ_i is a vector of parameters for homogenous consumer preferences over x and y, so it can be simplified as θ under the author's assumption; random terms ϵ_{ijt} are individual-specific random utility components of J-dimensional random vector ϵ which are assumed to be independent and identically-distributed amongst individuals and periods. ϵ is also required to follow generalized extreme value (GEV) distribution. Based on the description above, u_{ijt} are therefore i.i.d amongst individuals as well. We can neglect the different individuals in (3.13) because of the assumption of homogeneity and decompose it as $u_{jt} = \delta_{jt} + \epsilon_{jt}$, where δ_{jt} is the mean utility $E[u_{ijt}]$.

Generally, the consumer makes the decision following two steps: first he chooses j_t^* that maximizes the utility from set J and then he decides whether to buy or to postpone the purchase until the next period. j_t^* is the product which contributes the maximum utility and set J includes all the products j available to the consumer. This optimal stopping problem can be generated as the following formula:

$$D(u_{it}, ..., u_{Jt}) = max_{\tau} \left\{ \sum_{k=t}^{\tau-1} \beta^{k-t} c + \beta^{\tau-t} E[max_{j \in J} u_{J\tau}] \right\} \tag{3.14}$$

where β is a common discount factor; c is the utility payoff and E denotes a conditional expectation. Let $v_t = max_{j \in J} u_{j\tau}$ and v_t has type I extreme value (Gumbel) distribution according to the described assumption about ϵ_{ijt}. The distribution of v_t is Gumbel distributed with a scale factor 1 (because of the assumption defined in this paper), so

$$F_v(z; r_t) = exp(-exp(-(z - r_t))) \tag{3.15}$$

where r_t in formula (3.15) is the mode of the distribution of v_t given by $r_t = lnG(exp(\delta_{1t}, ..., \delta_{Jt}, t)) = lnR$ (proof see Appendix A of Melnikov's paper). The consumer's decision can be finally transformed from (3.14) into:

$$D(v_t, c_t) = max \left\{ v_t, c + \beta E[D(v_{t+1})] \right\} \tag{3.16}$$

Industry evolution

Melnikov's model contains a very important factor r_t which characterizes the distribution of the maximum utility; it represents the evolution of the industry and it is formulated as the mode of the Gumbel distribution of v_t. It is also assumed that the evolution of the mean utility can be characterized by a homogenous Markov process with transition density $\Phi(r_{t+1}|r_t, \theta_r)$. Besides, r_t here follows a diffusion process defined by:

$$r_{t+1} = \mu(r_t) + \sigma(r_t)\nu_{t+1} \tag{3.17}$$

where ν_t are assumed to be i.i.d. standard normal $N(0,1)$. $\mu(r)$ and $\sigma(r)$ are continuous and almost everywhere differentiable and $\mu(r) > r$. The diffusion process can be expressed by means of different formulations; those formulations are reported in Melnikov's paper but not implemented into the framework presented. Here r_t has a homoschedastic random walk with drift, $r_{t+1} = r_t + \gamma + \sigma\nu$(where $\gamma \geq 0$). In this case, the Bellman equation (3.16) becomes:

$$D(v_t, r_t) = max\,\{v_t, c + \beta E[D(v_{t+1}(r_{t+1}))|r_t]\} \tag{3.18}$$

where v_t has Gumbel distribution with mode r_t. Meanwhile, the stopping set is $\Gamma(r) = \{v|v \geq c + \beta E[D(.)|r]\}$ and it is convenient to define $W(r) = c + \beta E[D(.)|r]$ as the reservation utility. $W(r)$ can be integrated as:

$$W(r) = c + \beta \int_{\infty}^{\infty} \int_{\infty}^{\infty} max(v, W(z))dF(v|z)d\Phi(z|r) \tag{3.19}$$

where from (3.15), $F(v|z) = exp(-exp(-(v-z)))$, $d\Phi(z|r) = \phi(\frac{z-\mu(r)}{\sigma(r)})dz$ and $\phi(.)$ is the standard normal density.

- **Demand structure**

The dynamics of the demand structure is determined by the probability of postponing the purchase, which the author denotes as:

$$\pi_{0t}(r_t) = P\left\{S_{i,t+1} = 0 | S_{it} = 0, r_t\right\} = F_v(W(r_t), r_t) = exp(-exp(-(W(r_t) - r_t))) \tag{3.20}$$

The probability of buying the product is defined as the individual hazard rate of the product adoption, $h(r_t) = 1 - \pi_{0t}(r_t)$. Furthermore, product-specific purchase probability is:

$$\pi_{jt}(r_t, .) = P\left\{u_{jt} \geq u_{kt}, \forall k \neq j; u_{jt} \geq W(r_t)\right\} \tag{3.21}$$

$$= P\left\{u_{jt} \geq W(r_t) | u_{jt} \geq u_{kt}, \forall k \neq j\right\} P\left\{u_{jt} \geq u_{kt}, \forall k \neq j\right\} \tag{3.22}$$

$$= P\left\{u_{jt} \geq W(r_t)\right\} p\left\{u_{jt} \geq u_{kt}, \forall k \neq j\right\} \tag{3.23}$$

$$= h(r_t)\frac{exp(\delta_{jt})G_j(e^{\delta_{j1}}, ..., e^{\delta_{jt}})}{G(e^{\delta_{j1}}, ..., e^{\delta_{jt}})} = h(r_t)\frac{exp(\delta_{jt})G_j(.)}{R_t} \tag{3.24}$$

$G_j(.)$ is the partial derivative of $G(.)$ with respect to jth argument.

One important issue in this Section is the calculation of the hazard rate with equation (3.20). By setting $Y(r_t) = W(r_t) - r_t$ and by combining (3.17), (3.18),(3.19), $Y(r_t)$ can be integrated as:

$$Y(r_t) = c + \beta\mu(r) - r + \beta \int_{\infty}^{\infty} E[max(\epsilon, Y(r_{t+1}))]\phi \left\{ \frac{z - \mu(r)}{\sigma(r)} \right\} dz \tag{3.25}$$

Recall that equation (3.17) and r_t's random walk with drift, $Y(r_t)$ is obtained from (3.25). Thus, the hazard rate h can be computed from (3.20).

- **Aggregation**

The transition of consumer state can be presented by a Markov matrix $H : \{0,1\} \rightarrow \{0,1\}$:

$$H_1(r_t) = \begin{bmatrix} \pi_{0t}(r_t) & h(r_t) \\ 0 & 1 \end{bmatrix} \tag{3.26}$$

The model can also accommodate product's "break down", which is given a probability q; the consumer i under the state $s_{it} = 1$ has probability q to return to the market. Therefore the transition matrix can be expressed by:

$$H_2(r_t) = \begin{bmatrix} \pi_{0t}(r_t) & h(r_t) \\ q & 1-q \end{bmatrix} \tag{3.27}$$

Participation rate is composed by two components: (1) the market share that does not own a product and (2) the market share that has break-down product (i.e. $\varphi_t = P[s_{it} = 0]$). The participation rate evolves over time according to the Kolmogorov-Chapman equation (3.28).

$$\varphi_{t+1} = \varphi_t \pi_{0t}(r_t) + q(1-q) \tag{3.28}$$

The hazard rate and product-specific purchase probability of (3.24) are adjusted into:

$$h_t^{'} = \varphi_t h_t \tag{3.29}$$

$$\pi_{jt}^{'} = \varphi_t h(r_t) \frac{exp(\delta_{jt}) G_j(.)}{R_t} \tag{3.30}$$

Rather than using Rust nested fixed point maximum likelihood algorithm, Melnikov uses an easier three-stage method to estimate the models that includes: (1) identifying static parameters by OLS, (2) using maximum likelihood to get parameters γ and σ from transition density $\Phi(r_{t+1}|r_t, \theta_r)$, and (3) estimating the remaining parameters (c, β, q, φ_0) by fitting predicted sales to the data with the moment condition. This method is based on the assumption that sales of product j can be aggregated, total market size is known and that the consumers are homogeneous.

3.3.3 Computer Server Choice Model with Persistence Effect

In the previous examples, dynamic discrete choice models are applied by Rust to describe the optimal stopping time for bus engine replacement decision and by Melnikov to model the choice from a set of differentiated durable goods with quality stochastically improving over time. Lorincz's paper incorporates a persistence effect into the Melnikov optimal stopping problem [Lorincz, 2005]. If the consumer already has one product, he can upgrade it without getting rid of the old one. Hence, besides deciding about the optimal time to buy a product, the consumer who already has a product can choose between simply using the original product and specifically upgrading its format. Overall, this model is built on three principles: product differentiation, optimal stopping problem and persistence effect. In this example, the model is applied to low-end server computers where formats are represented by operating systems (OSs). Since reliability and security are essential characteristics for servers, upgrades of OSs often need to be carried out. Meanwhile, servers are very important parts in a computer network which is ever changing, evolving and being upgraded; so the right server choice needs to be based on a more sophisticated forward-looking behavioral model. In particular, a dynamic nested logit model is estimated here, where nests are represented by different operating systems.

General dynamic nested logit model

Lorincz represents the evolution of the state vector by a Markov-transition probability and models the problem by using the Bellman equation:

$$V(s_t) = max_{j \in J(s)} \left[u_j(s_t) + \beta \int V(s_{t+1}) p(ds_{t+1}|s_{t,j}) \right] \tag{3.31}$$

The choice set $J(s)$ is partitioned into $G+1$ mutually exclusive subsets: $J(s) = \bigcup_{g=0}^{G} g(s)$. The subset $g = 0$ means that customers are not buying any product. The other G subsets correspond to different OSs which are nested. The state is composed of three elements: x, y and ϵ. x is a set of product specific state variables such as characteristics and price; y is the customer specific state variable observed by econometrician and $y \in 0, 1, ..., G$. $y = 0$ indicates that the customer does not own anything at the beginning of the current period. $y = g$ indicates that the product owned currently belongs to nest g. This latter specification differentiates this approach from the Melnikov's formulation where only states $y = 0$ (not owning a product) and break-down probability are considered.

Different utilities need to be specified depending on the conditions y and g:

In case 1, $y = 0$ and $g = 0$ the customer owns nothing and does not buy, $u_j = c + \epsilon_0$. The constant c is a payoff.

In case 2, $y = 0$ and $g \in \{1, ...G\}$, $j \in g$ the customer owns nothing but buys one from nest g. Payoff is then the sum of a product specific value $u_j = x_j\gamma_g + \epsilon_g + (1 - \sigma_g)\epsilon_j$ where γ_g is a vector of parameters and $\sigma_g \in (0, 1)$ governs correlation in nest g . The terms ϵ_g and ϵ_j represent the heterogeneity of nests and products within nests respectively. And they are distributed identically and independently across nests and periods with extreme value distributions.

In case 3, $y \in \{1, ..., G\}$ and $g = 0$ the customer does not buy anything when he already owns one product. So he gets a format specific "continuation value"' c_y. $u_j = c_y + \epsilon_0^u$.

In case 4, $y \in \{1, ..., G\}$ and $g \in \{1, ..., G\}$, $j \in g = y$ the customer already has a product and decides to upgrade it. So the customer chooses an alternative j from the upgrade nest y of the original product. $u_j = x_j^u\gamma_g^u + \epsilon_g^u + (1 - \sigma_g^u)\epsilon_j^u$.

Some assumptions are given. ϵ_0 and ϵ_0^u are iid distributed across all alternatives and periods with extreme value. $\epsilon_g + (1 - \sigma_g)\epsilon_j$ and ϵ_g are iid distributed across nests and periods with extreme value, that is the same as $\epsilon_g^u + (1 - \sigma_g^u)\epsilon_j^u$ and ϵ_g^u.

Then, transition probabilities are specified as following:

$$p(x_{t+1}, y_{t+1}, \epsilon_{t+1}|x_t, y_t, \epsilon_t, j) = h(\epsilon_{t+1}|x_{t+1}, y_{t+1})f(x_{t+1}|x_t)l(y_{t+1}|y_t, j) \tag{3.32}$$

Simplified dynamic nest logit model

The customer is supposed to choose between nests; this assumption reduces the state and choice dimensional space. Since the specific product index j is identified by its nest g, the author replaces j by g in the transition probability function of state y . So in equation (3.32) $l(y_{t+1}|y_t, j)$ can be changed into $l(y_{t+1}|y_t, g)$ while equation (3.31) becomes $V(s_t) = max_{j \in J(s)} \left[u_j(s_t) + \beta \int V(s_{t+1}) p(ds_{t+1}|s_t, g)\right]$. Therefore, it is assumed that the formats of all products belonging to the same nest g are the same and that customer specific persistence effect is carried out through time by the format but not by the product itself.

p_j is defined as the probability of choosing product j belonging to nest g. As in classical nested logit model p_j can be obtained by multiplying the conditional probability of choosing j from g and the probability of choosing g , that is $p_j = p_{j|g} p_g$. Let $w_g(s) = \int V(s_{t+1}) p(ds_{t+1}|s_t, g)$. So p_j of case 2 is represented by the following nested logit structure:

$$p_j = \frac{exp\left[(x_j \gamma_g / (1 - \sigma_g))\right]}{R_g} \frac{exp\left[(1 - \sigma_g) ln R_g + \beta w_g(s)\right]}{\sum_{g'=1}^{G} exp\left[(1 - \sigma'_g) ln R_{g'} + \beta w'_g(s)\right]} \tag{3.33}$$

where $R_g \equiv \sum_{j \in g} exp\left[x_j \gamma_g / (1 - \sigma_g)\right]$. In this formula, the first term and the second term are both standard logit models. The mean utility of the first term is $(x_j \gamma_g / (1 - \sigma_g))$. The value $\gamma_g \equiv ln R_g$ is the expected maximum utility of the conditional choice problem. The mean utility of the second term is the weighted sum of the value γ_g of this nest g and the discounted value of the next period problem. Similar formula can be generated for case 4 where the corresponding inclusive value is γ_g^u .

Through reducing the state vector of the problem, the state is composed of (γ, y, ϵ) with transition probability $p(\gamma_{t+1}, y_{t+1}, \epsilon_{t+1}|\gamma_t, y_t, \epsilon_t, j) = h(\epsilon_{t+1}|\gamma_{t+1}, y_{t+1})$ $f(\gamma_{t+1}|\gamma_t)l(y_{t+1}|y_{t+1}, g)$. Here γ_t is the vector of γ_g's and γ_g^u's. The Bellman equation is updated as

$$V(s_t) = max_{g\in(0,1,\ldots,G)} \left[u_g(z_t) + \beta \int V(z_{t+1})p(dz_{t+1}|z_t, g)\right] \qquad (3.34)$$

Estimation

The model is estimated following three steps. First, specify static conditional logit models of within nest choices are estimated; second, the transition probabilities for the models' inclusive values are calculated; then a dynamic logit model of choice between nests including the results from the last two steps is calibrated. More technical details can be found in Lorincz's (2005).

3.3.4 Dynamic Durable Goods Demand with Consumer Heterogeneity

In previous examples, consumers are assumed to be homogeneous and randomly i.i.d. Under this assumption the parameters of the static problem can be estimated separately from the dynamic one. Homogeneity simplifies the problem formulation although the computation cost associated to the fixed point algorithm is still high. Furthermore, when extending the original technique to fully heterogeneous consumer problems, the integration of the individual demand function over the distribution of consumers' characteristics is needed. In this context, Juan Esteban Carranza [Carranza, 2006] models digital camera demand by using models similar to those described in previous Sections but incorporates fully heterogeneous consumers into a reduced form of the participation probability. The author is then able to estimate the joint distribution of consumers' preferences and the parameters associated to the participation function which is based on the observed number of purchases.

A dynamic model of demand

Suppose individual i buys product j, the lifetime utility of this purchase is:

$$u_{ijt} = \varsigma_{ij} + \alpha_j^x x_j - \alpha_i^p p_{jt} + \epsilon_{ijt} \tag{3.35}$$

Similarly with the framework presented in Section 2, ς_{ij} is an unobserved product attribute common to all consumers who purchase product j at time t; p_{jt} is the price of product j at time t and x_j is the vector of observed static characteristics of product j. Preference parameters $(\varsigma_{ij}, \alpha_i^x, \alpha_i^p)$ vary across consumers. The author lets $\varsigma_{ij} = \varsigma_j + \sigma_\varsigma e_{i\varsigma}, \alpha_i^x = \alpha^x + \sigma_x e_{ix}$, and $\alpha_i^p = \alpha^p + \sigma_p e_{ip}$, where e_i is drawn from a know iid distribution F_e . So (3.35) can be rewritten as:

$$u_{ijt} = (\varsigma_j + \alpha^x x_j - \alpha^p p_{jt}) + (\sigma_\varsigma e_{i\varsigma} + \sigma_x e_{ix} x_j - \sigma_p e_{ip} p_{jt}) + \epsilon_{ijt} \tag{3.36}$$

$$= \delta_{jt}(x_j, p_{jt}; \theta_0, \varsigma_j) + \mu_{ijt}(x_j, p_{jt}; \theta_1, e_i) + \epsilon_{ijt} \tag{3.37}$$

In this formula, $\theta_0 = (\alpha^x, \alpha^p)$, $\theta_1 = (\sigma_\varsigma, \sigma_x, \sigma_p)$ and $e_i = (e_{i\varsigma}, e_{ix}, e_{ip})$. The utility function has mean δ_{jt} which is common to all consumers, variance μ_{ijt} which captures the variability of tastes across consumers and an idiosyncratic product-consumer random component ϵ_{ijt} .

The reservation utility of the consumer is the value of not purchasing anything at time t and waiting until the next period to decide. The problem can be formulated as:

$$W(S_{it}) = 0 + \beta E\left[MAX\left\{V_{i,t+1}, W(S_{i,t+1})\right\}|S_{it}\right] \tag{3.38}$$

Let $v_{it} = max_j\left\{u_{ijt}(.)\right\}$ be the maximum utility consumer i can get from any product purchased at t and it is assumed that its distribution $F_{v_{it}}$ is known (recall that in Melnikov case $F_{v_{it}}$ is GEV distributed). The probability that the consumer buys any product at time can be expressed as a hazard rate (see 3.2.2) and obtained from the known distribution of v_{it}:

$$Pr(purchase) \equiv h_{it}(S_{it}) = P\left[v_{it} > W(S_{it})\right] = 1 - F_{v_{it}}(W(S_{it})) \tag{3.39}$$

It is assumed that ϵ_{ijt} have an independent extreme value distribution. According to Melnikov's deduction (Melnikov, 2000-see appendix), v_{it} has extreme value distribution with mode r_{it} :

$$r_{it}(.) = log\left[\sum_{k\in J_t} exp(\delta_{kt}(.) + \mu_{ikt}(.))\right] \tag{3.40}$$

Since v_{it} is assumed to be Markovian, state S_{it} in formula (3.39) and (3.40) can be replaced by r_{it} . Then the specific product purchase probability is

$$h_{ijt}(., e_i) = h_{it}(r_{it}(., e_i)) \frac{exp\delta_{jt}(.) + \mu_{ijt}(., e_i)}{exp(r_{it}(., e_i))} \tag{3.41}$$

Estimation

To obtain the predicted market share for product j, (3.41) has to be integrated across consumers which can be based on the distribution of F_g.

$$s_{jt}(\theta_0, \theta_1) = \int \left[h_t(r_t(\theta_0, \theta_1, e)) \frac{exp(\delta_{jt}(\theta_0) + \mu_{jt}(\theta_0, \theta_1, e))}{exp(r_t(\theta_0, \theta_1, e))} \right] dF_e \tag{3.42}$$

$\theta = (\theta_0, \theta_1)$ can be obtained by equating the predicted and observed demand but the premise is the observed demand for product j at t and market size are known, that is:

$$M_t s_{jt}(\theta_0, \theta_1) = Q_{jt} \tag{3.43}$$

where Q_{jt} is the observed demand for product j at t and M_t the market size. The integration of (3.42) can be simplified by using simulation techniques, to obtain N draws of $\{e_n\}_{n=1,\ldots,N}$. See formula (3.44).

$$s_{jt}(\theta_0, \theta_1) \approx \frac{1}{N} \sum_{n-1}^{N} \left[\Psi_{nt} h_{nt}(r_{nt}(\theta_0, \theta_1, e_n)) \frac{exp(\delta_{jt}(\theta_0) + \mu_{njt}(\theta_0, \theta_1, e_n))}{exp(r_{nt}(\theta_0, \theta_1, e_n))} \right] \tag{3.44}$$

In (3.44), $\Psi_{n,1} = 1$ and $\Psi_{n,t>1} = \Psi_{n,t-1}(1 - h_{n,t-1})$ is the probability that consumer n is still in the market in period t. When computing mean utility δ_j from (3.43), a fixed point for each simulated consumer is required. This procedure is computationally costly and it is not clear whether the computable points have a contraction across all the relevant parameter space. To circumvent the nested computation, the author imposes a parametric structure on h_{it} and estimate its parameter θ_{2i} as a part of the whole model. If the transition probability of r_{it} follows a Makov process, the participation probability can be approximated as $h_{it} = \tilde{h}_i(r_{it}(.); \tilde{\theta_{2i}})$ and the function $\tilde{h(.)}$ varies across consumers. (3.44) will be updated as:

$$s_{jt}(\theta_0, \theta_1, \theta_2) \approx \frac{1}{N} \sum_{n-1}^{N} \left[\Psi_{nt} \tilde{h_{nt}}(r_{nt}(.); \tilde{\theta_{2n}}) \frac{exp(\delta_{jt}(\theta_0) + \mu_{njt}(\theta_0, \theta_1, e_n))}{exp(r_{nt}(\theta_0, \theta_1, e_n))} \right] \tag{3.45}$$

The detail estimation procedure of θ_0, θ_1 and $\tilde{\theta_{2n}}$ can be referred from Carranza's paper in 2006.

3.3.5 Dynamic Durable Goods Demand with Repeat Purchases

Carranza's model incorporates consumer heterogeneity into differentiated product demands but does not account for repeat purchases. Gowrisankaran and Rysman [Gowrisankaran and Rysman, 2007] generate a dynamic model of consumer preference for the digital camcorder. It allows for unobserved product characteristics, repeat purchases, endogenous prices and differentiated products.

It is assumed that a consumer who purchases product j at t would receive a net flow utility $u_{ijt} = \delta^f_{jit} - \alpha^p_i ln(p_{jt}) + \epsilon_{ijt}$, where $\delta^f_{jit} = \alpha^x_i x_{jt} + \varsigma_{jt}$. δ^f_{jit} is the gross flow utility from product j purchased at time t . x_{jt} is observed characteristics and ς_{jt} is unobserved; p_{jt} is price; ϵ_{ijt} is an idiosyncratic unobservable parameters. Let Ω_t denote current product attributes and it evolves according to the Markov process $P(\Omega_{t+1}|\Omega_t)$. The author defines a consumer who does not purchase a new product at time t has net flow utility as well: $u_{i0t} = \delta^f_{i0t} + \epsilon_{i0t}$. Then the value function could be $V(\epsilon_{it}, \delta^f_{i0t}, \Omega_t)$ and the expectation of the value function is $EV_i(\delta^f_{i0t}, \Omega_t) = \int_{e_{i,t}} V(\epsilon_{it}, \delta^f_{i0t}, \Omega_t)dP_\epsilon$. ϵ_{it} is iid and it satisfies the conditional independence assumption in Rust's 1987 paper. Bellman equation is represented as:

$$V_i(\epsilon_{it}, \delta^f_{i0t}, \Omega_t) = max\left\{u_{i0t} + \beta E[EV_i(\delta^f_{i0t}, \Omega_{t+1})|\Omega_t], max_{j=1,...J_t}\left\{u_{ijt} + \beta E[EV_i(\delta^f_{ijt}, \Omega_{t+1})|\Omega_t]\right\}\right\} \tag{3.46}$$

The problem so far is the large dimensionality of Ω_t that leads to the heavy difficulties to compute (3.46). Therefore, the author substitutes Ω_t with a scalar variable, the logit inclusive value of purchasing in time t :

$$\delta_{it}(\Omega_t) = ln(\sum_{j=1,...J_t} exp(\delta_{ijt}(\Omega_t))) \tag{3.47}$$

Besides, there is a main simplifying assumption, the logit inclusive value depends only on the current logit inclusive value that is termed Inclusive Value Sufficiency. This assumption indicates that if two states have the same inclusive value δ_{it} for consumer i at current time t , they have the same distribution of inclusive value for this consumer for the future time. The simplification from this assumption is represented in this formula:

$$EV_i(\delta^f_{i0t}, \Omega_{it}, E[\delta_{it+1}, \delta_{it+2}, ...|\Omega]) = EV_i(\delta^f_{i0t}, \delta_{it}) \tag{3.48}$$

To specify the density $P(\delta_{i,t+1}|\delta_{it})$, a simple function is assumed with linear autoregressive specification with drift $\delta_{i,t+1} = \gamma_{1i} + \gamma_{2i}\delta_{it} + u_{it}$,where u_{it} is normally distributed with mean 0 and γ_{1i} , γ_{2i} are parameters.

The estimation algorithm includes three levels of optimization. The inner loop evaluates the predicted market shares as a function of $\bar{\delta}^f_{jt}$ and parameters by solving the consumer dynamic programming problem for the simulated consumers and then integrating across consumer types. The middle loop performs a fixed point equation and iterates until the new and old $\bar{\delta}^f_{jt}$ converge. The outer loop is a search over the parameters. Details can be found in paper (Gowrisankaran and Rysman, 2007).

The model allows for consumers' repeat purchases but does not introduce any new parameters over the static model that is because there are some strong assumptions for the product. The assumptions include: durable goods do not wear out; there is no resale market for them; and there are no households with more than one good at the same type. Therefore, the second purchased good will only have new features which are observed and very different from previous good's type.

3.4 Summary of Dynamic Demand Models in Economics

Finally the five dynamic models are compared and presented in Table 2.6 which includes the case description, the main formulation and the estimation results.

Rust's optimal stopping problem provides the basic model framework and the estimation method for the dynamic models developed later in the literature. It is a single agent problem describing the decision of time to make one purchase over a set of products with homogeneous attributes (bus engines with different models). The estimation method is the nested fixed point algorithm that computes the maximum likelihood estimates and reduces the computational burden of solving the contraction fixed point EV_θ. Melnikov's dynamic demand model of computer printers contains the concept that product quality rapidly improves over time and the product durability impacts the evolution of prices and sales. Same as Rust's example, only one purchase is made and all consumer heterogeneity is captured by a term that is independently distributed across consumers, products and time. The difference is that it deals with differentiated durable products rather than homogenous products. The estimation method is a three-stage procedure that replaces the more complicated nested fixed point maximum likelihood algorithm. Then Lorincz's model extends Melnikov's optimal stopping problem with a persistent effect. The consumer can choose to upgrade the product instead of getting rid of it. Given that different product alternatives and two conditions are considered: without a product (when alternatives are not to buy and to buy a new product) and with the current product (when alternatives are not to upgrade and to upgrade the owned product), thus the decision problem in this case is specified as a dynamic nested logit model. The estimation follows a sequential procedure with three steps. Juan Esteban Carranza incorporates fully heterogeneous consumers into a reduced form of the participation probability for a digital camera demand problem. He estimates the joint distribution of consumers' preference and parameters of the participation function which is based on the observed number of purchases. The distribution of preference is defined as a continuous parametric distribution. The complicated integration across consumers in the estimation part needs simulation. Gowrisankaran and Rysman's dynamic model for digital camcorder demand allows for repeat purchases which is different from previous studies. The estimation algorithm includes three levels of optimization and the repeat purchases estimation could be simplified only with strong assumptions. Table 3.1 shows the summary of these dynamic models.

Name	Bus engine replacement	Computer printer demand	Low-end computer server demand with persistence effects	Digital camera demand	Digital camcorder demand
Author (year)	John Rust (1987)	Oleg Melnikov (2000)	Szabolcs Lorincz (2005)	Juan Esteban Carranza (2006)	Gautam Gowrisankaran and Marc Rysman (2009)
Data	ten years of monthly data on bus mileage and engine replacements for a subsample of 104 buses	monthly data on sales and average prices of computer printers and multifunction devices. 462 models from 27 manufacturers. 1998-1999	quantities, prices and technical characteristics for all server models in three regions. 1996-2001	a panel of sales, prices and characteristics of digital cameras, 1998-2001	monthly level for 378 models and 11 brands, number of units sold, price, others Mar 2000-May 2006
Characteristics	Single agent, one purchase, homogeneous attributes of the products	Homogeneous consumers with one purchase, differentiated durable products. Potential market size is required.	Homogeneous consumers with one purchase, differentiated servers and upgraded formats	Fully heterogeneous consumers and differentiated durable products. Potential market size is required.	Repeat purchases, heterogeneous consumers and differentiated products
Main formula	Described recursively by Bellman's principle of optimality	Formulate the timing of consumers' purchase as an optimal stopping problem and the solution defines the hazard rate of production adoptions	The utility function in the Bellman equation has four cases. The nested logit assumptions describe the unobserved heterogeneity term.	The endogenous participation probability has a reduced form. The identification of the participation function is based on the observation over time.	Described recursively by Bellman's principle of optimality with logit inclusive value of purchasing in a given time.
Estimation method	Nested fixed-point maximum likelihood algorithm that computes theta and associated value function	A nested three-step method that allows for sequential parameters with aggregate data from relatively short time series	Estimated by a sequential procedure, specifying static conditional logit models of within nest choices, estimating transition probabilities and the dynamic logit model of choice between nests	Integrating across consumers by simulation methods to obtain the market demand for each product. Estimate the parameter vector by equating the predicted and observed demand.	Three levels of non-linear optimization: a search over parameters outside, a fixed point calculation of population mean flow utilities outside and calculation of predicted market shares inside

Tab. 3.1: Comparison of the five dynamic models

4. DYNAMIC CAR OWNERSHIP FORMULATION

Discrete choice models based on Random Utility Maximization (RUM) theory have been of interest to researchers for many years in a variety of disciplines. These methodologies are used to analyze and predict individual choice behavior [Ben-Akiva and Lerman, 1985, Daly, 1982, McFadden, 1978, Small, 1987, Chu, 1981, hua Wen and Koppelman, 2001, Papola, 2000]. However, discrete choice methods are commonly based on a static framework which is limited by the assumption that consumers are not affected by past and future states when choosing their preferred alternative in the present. The gap between discrete choice model and dynamics in individual behavior has spurred various developments that are mainly intended to enrich the basic theory by including the changes occurring in the system over time.

This chapter presents a comprehensive modeling framework for car ownership modleing; it includes the consumer utility specification, the definition of the dynamic programming problem, the industry evolution equation and the optimization algorithm.

4.1 Car Ownership Formulation

4.1.1 General Consumer Stopping Problem

We consider a consumers set $\mathcal{I} = \{1, \ldots, M\}$, where each consumer $i \in \mathcal{I}$ can be in one of two possible states at each time period $t \in \{0, 1, \ldots, T\}$. More precisely, we have the state space

$$S = \{0, 1\},\ \forall i \in \{1, \ldots, M\},\ t \in \{0, 1, \ldots, T\}.$$

Each state $s_{it} \in S$ can therefore take two values:

$$s = \begin{cases} 0 & i \text{ in the market,} \\ \\ 1 & i \text{ out of market.} \end{cases}$$

'In the market' typically means the consumer, also referenced as the individual has the possibility to buy a product while 'out of the market' means the individual never considers to make a purchase at all. State is evolving from period to period depending on the consumer's decision as well as some external factors. In other words, in an optimal stopping problem, a consumer in state 0 tries to choose the best transition period in order to attain state 1. The decision process continues even when he/she reaches state 1 because the framework is used for repeated purchases. In the car ownership case, 'in the market' means the individual considers to buy a car no matter whether he/she currently owns a car. If the individual does not own a car, it is quite possible he/she considers to buy one; if he/she does own a car but with some problematic condition (or plan to sell the previous car), he/she can also consider to replace it. ‘Out of market’ means the individual does not consider to buy a car at all.

The car ownership problem is described by a regenerative optimal stopping problem, i.e. when the individual reaches state 1, this state is replaced by the state 0, and some variables of the problem such as current vehicle age and mileage are reinitialized. The regeneration can sometimes happen at each period in state 1 with some probability (strictly less than 1).

In each time period t, consumer i in state $s_{it} = 0$ has two options

1. to buy one product $j \in \mathcal{J}_t$ and obtain a terminal period payoff u_{ijt}, where $\mathcal{J}_t = \{1, \ldots, J_t\}$ is the set of products available at time t;

2. to postpone and obtain a one-period payoff c_{it}, which is a function of individual i's attributes and the characteristics of current product owned by i, i.e. $c(x_{it}, q_{it}; \theta_i, \alpha_i)$. x_{it} is a vector of attributes for individual i at time t, e.g., sex, education, income, age, etc., and q_{it} is the vector of characteristics of current product owned by this individual. θ_i and α_i are parameters vectors for x_{it} and q_{it} respectively.

It is here assumed that the choice set $\mathcal{J}_t$ is consistent in each time period t, so the subscript t from $\mathcal{J}_t$ and J_t can be dropped, and keep $\mathcal{J}$ and J respectively. The payoff $\boldsymbol{u}_{ijt}$ is expressed as a random utility function

$$\boldsymbol{u}_{ijt} = \boldsymbol{u}(x_{it}, d_j, y_{jt}, \theta_i, \gamma_i, \lambda_i, \epsilon_{ijt}), \tag{4.1}$$

where

- $x_{it}, \theta_i \in \Re^Q$ are defined as above;

- $d_j \in \Re^K$ is a vector of static attributes for potential choice j and γ_i is a vector of parameters related to d_j;

- $y_{jt} \in \Re^H$ is a vector of dynamic attributes for product j at time t; y_{jt} can be energy (typically fuel) cost per mile[1], buying cost, environment incentives, etc. λ_i is a vector of parameters related to y_{jt}.

- $\boldsymbol{\epsilon}_{ijt}$ is an individual-specific random term depending of i, the product j and the time period t. It is assumed the random terms $\boldsymbol{\epsilon}_{ijt}$ are components of J-dimensional random vectors $\boldsymbol{\zeta}_{it}$ ($i = 1, \ldots, M$, $t = 0, \ldots, T$), which are independent and identically-distributed amongst individuals and periods, and have zero means (so for convenience, the subscripts i and t are sometimes dropped). It is also required that $\boldsymbol{\zeta}_{it}$ follows the generalized extreme value (GEV) distribution, characterized by the cumulative joint distribution function $F_{\boldsymbol{\zeta}}(\epsilon_1, \ldots, \epsilon_J)$ of the form $e^{-G(e^{-\epsilon_1}, \ldots, e^{-\epsilon_J})}$. The function $G(a_1, \ldots, a_J) = G(a)$ has the following properties:

  (i) $G(a) \geq 0, \forall a \in J,\ a \geq 0$;

  (ii) $G(a)$ is homogenous of degree $\kappa > 0$; here let $\kappa = 1$;

  (iii) $\lim_{a_j \to \infty} G(a_1, \ldots, a_J) = \infty, \forall j = 1, \ldots, J$;

[1] This allows us to summarize car consumption and current fuel price into one attribute.

(iv) for any distinct sequence $(j_1, \dots, j_k)$, $\partial^k G / \partial a_{j1} \dots \partial a_{jk}$ is greater than 0 if k is odd and less than 0 if k is even.

Besides, $G_j(\cdot)$ is the first partial derivative of $G(\cdot)$ with respect to j^{th} argument. It is further assumed that equation (4.1) can be rewriten with error acting in an additive way:

$$\boldsymbol{u}_{ijt} = V_{ijt} + \boldsymbol{\epsilon}_{ijt},$$

where V_{ijt} is the mean utility, i.e. $V_{ijt} = E[\boldsymbol{u}_{ijt}]$. It also assumed so far that these parameters are the same over individuals, i.e. $\theta_i = \theta$, $\alpha_i = \alpha$, $\gamma_i = \gamma$, and $\lambda_i = \lambda$, $i = 1, \dots, M$ (in other words, there is no heterogeneity between individuals).

Relying on McFadden seminal paper [McFadden, 1978], $\boldsymbol{\zeta}$ follows a multivariate extreme value distribution. An example of a quite general G function is

$$G(a) = \sum_{n=1}^{N} \left(\sum_{j \in B_n} a_j^{\frac{1}{1-\delta_n}} \right)^{1-\delta_n} \tag{4.2}$$

where $B_n \subseteq \{1, \dots, J\}$, $\cup_{n=1}^{N} B_n = \{1, \dots, J\}$, and $0 \leq \delta_n < 1$. Each B_n ($n = 1, \dots, N$) can therefore be seen as a nest, with possible overlappings between the nests. δ_n can be interpreted as an index of the similarity of the unobserved attibutes in B_n. The choice probabilities for the function (4.2) satisfy

$$\begin{aligned} P_i &= \sum_{n=1}^{N} P[i \,|\, B_n] P[B_n] \\ &= \frac{\sum_{i \in B_n} e^{\frac{V_i}{1-\delta_n}} \left(\sum_{j \in B_n} e^{\frac{V_j}{1-\delta_n}} \right)^{-\delta_n}}{\sum_{n=1}^{N} \left(\sum_{k \in B_n} e^{\frac{V_k}{1-\delta_n}} \right)^{1-\delta_n}}, \end{aligned} \tag{4.3}$$

where

$$P[i \mid B_n] = \begin{cases} \dfrac{e^{\frac{V_i}{1-\delta_n}}}{\sum_{j \in B_n} e^{\frac{V_j}{1-\delta_n}}} & \text{if } b \in B_n; \\ 0 & \text{if } b \notin B_n. \end{cases}$$

In the special case $G(a_1, \dots, a_J) = \sum_{j=1}^{J} a_j$, we have $F_{\boldsymbol{\zeta}}(\epsilon_1, \dots, \epsilon_J) = F(\epsilon_1), \dots, F(\epsilon_J)$, so ϵ_j are all independent, and $\delta_n = 0$ $(n = 1, \dots, N)$. When all alternatives are available, probabilities (4.3) simplify to usual multinomial logit probabilities.

The two-step decision process is that, at each period, first, the consumer decides to buy or to postpone the purchase until the optimal time period τ, that is the time when the consumer decides to buy instead of postponing; then, the consumer chooses the product j_t^* that maximizes utility (4.1) from $\mathcal{J}$. The consumer deciding to buy or postpone is the optimal stopping problem at time t:

$$D(\boldsymbol{u}_{i1t}, \dots, \boldsymbol{u}_{iJt}, c_{it}, t) = \max_{\tau} \left\{ \sum_{k=t}^{\tau-1} \beta^{k-t} c_{it} + \beta^{\tau-t} E\left[\max_{j \in \mathcal{J}} \boldsymbol{u}_{ij\tau} \right] \right\} \tag{4.4}$$

where

- β is a discount factor in [0,1);
- c_{it} is the payoff function of individual i's attributes and the characteristics of current product owned by i when choosing to postpone the purchase, as defined before.

Let $\boldsymbol{v}_{it} = \max_{j \in \mathcal{J}} \boldsymbol{u}_{ijt}$. According to the previously described assumption about $\boldsymbol{\epsilon}_{ijt}$, $\boldsymbol{v}_{it}$ is Gumbel distributed with a scale factor equals to 1 since it is assumed in property (ii) that $G(a)$ is homogenous of degree one. In other terms, there are

$$
\begin{aligned}
F_{\boldsymbol{v}}(z; r_{it}) &= e^{-e^{-(z-r_{it})}}, \\
f_{\boldsymbol{v}}(z; r_{it}) &= e^{r_{it}} e^{(-e^{-(z-r_{it})}-z)} \\
&= -e^{-(z-r_{it})} F_{\boldsymbol{v}}(z; r_{it}),
\end{aligned} \tag{4.5}
$$

where r_{it} is the mode of the distribution of $\boldsymbol{v}_t$, that is

$$r_{it} = \ln G(e^{V_{i1t}}, \ldots, e^{V_{iJt}}). \tag{4.6}$$

Later, r_{it} will replace $r_{it}(y_{it})$ in order to stress the functional relationship between the distribution mode and the dynamic attributes. Based on dynamic programming theory, the consumer's decision can be transformed from (4.4) into:

$$D(\boldsymbol{v}_{it}, c_{it}) = \max\left\{\boldsymbol{v}_{it}, c_{it} + \beta E[D(\boldsymbol{v}_{i,t+1})]\right\} \tag{4.7}$$

4.1.2 Utility Formulation

The Bellman equation (4.7) becomes:

$$D(\boldsymbol{v}_{it}, c_{it}) = \max\left\{\boldsymbol{v}_{it}, c_{it} + \beta E[D(\boldsymbol{v}_{i,t+1}(\boldsymbol{y}_{t+1}, c_{i,t+1})) \,|\, y_t]\right\} \tag{4.8}$$

where $y_t = (y_{1t}, ...y_{Jt})$. This is a standard regenerative optimal stopping problem. The stopping set is given by:

$$\Gamma(y_t) = \{v_{it} \,|\, v_{it} \geq c_{it} + \beta E[D(\cdot) \,|\, y_t]\} \tag{4.9}$$

$W(y_t)$, as the reservation utility level is defined by function:

$$W(y_t) = c_{it} + \beta E[D(\boldsymbol{v}_{i,t+1}(\boldsymbol{y}_{t+1}, c_{i,t+1})) \,|\, y_t] \tag{4.10}$$

and consider the optimal policy:

$$\begin{cases} v_{it} & \text{if } v_{it} \geq W(y_t) \\ W(y_t) & \text{otherwise.} \end{cases}$$

Using (4.10), (4.7) can be simplified as:

$$D(v_{it}) = \max \{v_{it}, W(y_t)\}.$$

At this step, the way to calculate expectation utility E is complicated and will be discussed later.

As presented in (4.9), the consumer i will buy some product at time t only when $v_{it} > W(y_t)$. The probability of postponing the purchase until the next period can therefore be written as:

$$\pi_{i0t}(y_t) \stackrel{def}{=} P\left[v_{it} \leq W(y_t)\right] = P\left[\text{postpone} \,|\, s_{it} = 0, y_t\right]$$
$$= F_v(W(y_t), y_t) = e^{-e^{-(W(y_t) - r_{it})}} \tag{4.11}$$

The probability of the product adoption is $h(y_t) = P\left[\text{buy} \,|\, s_{it} = 0, y_t\right] = 1 - \pi_{0t}(y_t)$, and the product-specific purchase probability is:

$$\pi_{ijt}(y_t) \stackrel{def}{=} P[U_{ijt} \geq U_{ikt}, \forall k \neq j, u_{ijt} \geq W(y_{it})]$$
$$= P[U_{ijt} \geq W(y_{it}) \,|\, U_{ijt} \geq U_{ikt}, \forall k \neq j] P[U_{ijt} \geq U_{ikt}, k \neq j]$$
$$= P[v_{it} \geq W(y_t)] P[U_{ijt} \geq U_{ikt}, k \neq j]$$
$$= h(y_{jt}) \frac{e^{V_{ijt}} G_j(e^{V_{ij1}}, \ldots, e^{V_{ijt}})}{G(e^{V_{ij1}}, \ldots, e^{V_{ijt}})}. \tag{4.12}$$

As introduced in Section 4.3.1, $G_j(\cdot)$ is the partial derivative of $G(\cdot)$ with respect to j^{th} argument.

4.1.3 Industry Evolution

As expressed in Section 4.3.1, y_{jt} represents the evolution of the product j's attributes and the market environment. $\boldsymbol{y}_{jt}$ here is assumed to follow a normal diffusion process:

$$\boldsymbol{y}_{j,t+1} = \mu(y_{jt}) + L(y_{jt}) \boldsymbol{\nu}_{j,t+1}, \tag{4.13}$$

where

- $\boldsymbol{\nu}_{jt}$ ($j = 1, \ldots, J$, $t = 1, \ldots, T$) are i.i.d. multivariate standard normal random vectors, $\mathcal{N}(0, I)$, where I denotes the identity matrix;

- $\mu(\boldsymbol{y}_{jt}) : \Re^H \to \Re^H$ and $L(\boldsymbol{y}_{jt}) : \Re^{H\times H} \to \Re^{H\times H}$ are continuous and have Jacobian matrices for almost every y_{jt}; $L(y_{jt})L(y_{jt})^T = \Sigma(y_{jt})$, the variance-covariance matrix of the random vector $\boldsymbol{y}_{j,t+1}$; $\Sigma(\boldsymbol{y}_{jt})$ is semi-definite positive, for almost every y_{jt};

- $\lim_{n\to\infty} \beta^n \mu^n(y_{jt}) < +\infty$, where $0 \leq \beta < 1$, $\mu^0(y_{jt}) = \mu(y_{jt})$ and $\mu^n(y_{jt}) = \mu(\mu^{n-1}(y_{jt}))$.

The random vector $\boldsymbol{y}_{j,t+1}$ therefore follows a multivariate normal distribution where L is the Cholesky factor of the variance-covariance matrix Σ and is lower triangular. The vector μ is the expected value of $\boldsymbol{y}_{j,t+1}$. $\boldsymbol{y}_{j,t+1}$ can for instance be specified as a random walk with drift η_j, i.e. $\boldsymbol{y}_{j,t+1} = \psi_j y_{jt} + \eta_j + L\boldsymbol{\nu}_{j,t+1}$. For simplicity, it is assumed that ψ_j and η_j are the same over all the alternatives. Therefore, $\mu(y_{jt}) = \psi y_{jt} + \eta$ and $L(y_{jt}) = L$.

4.1.4 Objective Function and Parameters to Estimate

We can summarize the parameters to estimate in the car ownership problem:

- θ, a vector of stationary consumer preference parameters related to individual attributes x_{it}.

- γ, a vector of parameters related to attributes for potential choice d_j.

- λ, a vector of parameters related to dynamic attributes of product j, y_{jt}. $\lambda = (\psi, \eta, L)$.

- β, the discount factor. It is usually set to 1.

- α, a vector of parameters for characteristics of current owned car q_{it}.

The parameters estimation is finally performed by maximizing the likelihood function:

$$\mathcal{L}(\theta, \gamma, \lambda, \beta, \alpha) = \prod_{i=1}^{M} \prod_{t=1}^{T} P_{it}[\text{decision}]. \tag{4.14}$$

The decision probability is presented as:

$$\begin{aligned} P_{it}[\text{decision}] &= P_{it}[\text{decision}, s_{it} = 0] + P_{it}[\text{decision}, s_{it} = 1] \\ &= P_{it}[\text{decision} \mid s_{it} = 0] P[s_{it} = 0] + P_{it}[\text{decision} \mid s_{it} = 1] P[s_{it} = 1] \end{aligned}$$

In the car ownership problem, since the individual decision is observed in the survey,

- if the individual reports consider to buy, $s_{it} = 0$, therefore $P[s_{it} = 0] = 1$ and $P[s_{it} = 1] = 0$, and

$$P_{it}[\text{decision}] = P_{it}[\text{decision} \mid s_{it} = 0];$$

- if the individual reports not consider to buy, $s_{it} = 1$, therefore $P[s_{it} = 0] = 0$

and $P[s_{it} = 1] = 1$, and

$$P_{it}[\text{decision}] = P_{it}[\text{decision} \mid s_{it} = 1].$$

Under this second condition, the respondent is out-of-market, so $P_{it}[\text{not to buy}] = 1$ and $P_{it}[\text{to buy}] = 0$. When the individual's decision is 'not to buy', the probability does not affect the result of the likelihood function (whatever are the parameters). If and only if one interviewee reports he will buy but under the condition that $s_{it} = 1$, the likelihood function (4.14) becomes $\mathcal{L}(\theta, \gamma, \lambda, \beta, \alpha) = 0$. Thus, the whole system collapses (indicating a problem in the underlying dataset). As a result, in car ownership example, the complete likelihood function is:

$$\mathcal{L}(\theta, \gamma, \lambda, \beta, \alpha) = \prod_{(i,t)\in V} P_{it}[\text{decision} \mid s_{it} = 0], \tag{4.15}$$

where $V = \{(i, t) \mid i \in 1, \ldots, M, t \in 1, \ldots, T \text{ and } s_{it} = 0\}$. The "decisions" include postponing and product-specific purchase. So $P_{it}[\text{decision} \mid s_{it} = 0] = \{\pi_{i0t}, \pi_{ijt}\}$.

4.1.5 Dynamic Estimation Process

Maximum log-likelihood estimation method is used to optimize the function (4.15). First π_{i0t} must be obtained and then π_{ijt} calculated. In the function $\pi_{i0t}(y_t) = e^{-e^{-(W(y_t)-r_{it})}}$ (4.11), r_{it} can be obtained by (4.6) and $W(y_t)$ is the reservation utility in time period t. $W(y_t)$ can be calculated by $c_{it}+\beta E[D(\boldsymbol{v}_{i,t+1}(\boldsymbol{y}_{t+1}, c_{i,t+1})) \mid y_t]$ (4.10), which is composed by the current product utility and the expected utility in next time period $t+1$, $E[D(\boldsymbol{v}_{i,t+1})]$. The key point during the whole process is to figure out how to calculate the expected utility.

At each time period, the respondent might have a perspective about the future scenarios in the short-term horizon which is characterized by alternatives' attributes changing over time (i.e. fuel price, rapidly technology development, etc.); therefore the expectation utility should account for the possible market conditions in the respondent's perspective scenarios. It is assumed that the respondent has the perspective on a limited number of time periods, such as T. Starting from the generic time period t, the respondent faces two possible alternatives, buy one type of car and not buy; at t+1, each of the two scenarios from time t generates another two buy and not buy scenarios, for a total of four scenarios. Therefore, the decision process is formulated with a scenario tree (see Figure 4.1). This scenario tree constitutes the base for the calculation of the expected utility. An example is provided here on the procedure adopted to calculate $\pi_{i0,0}$ and $E[D(\boldsymbol{v}_{i1})]$ which will be indicated for simplification purpose by $E[D_1]$ because all the expectations in the example are for individual i.

- Assumption. At each time period, the respondent has expectation over a limited number of future time periods, which is limited to two in order to reduce the number of leafs in the tree scenarios. At time period 0, the respondent can anticipate the possible alternative characteristics (i.e. fuel price, MPG) for time periods 1, and 2. $E[D_3] = 0$ because he/she knows nothing for time period 3 when he/she decides at time period 0.

- Calculate $E[D_1]$. Since the reservation utility $W(y_0) = c_0 + E[D_1]$ can be calculated according to (4.10), $E[D_1]$ must be calculated first in order to get $\pi_{i0,0}$. At time 0, the respondent has two alternatives for successive time 1, buy the car with the highest utility or not buy (see Figure 4.1). The right side of the utility function $E[D_1] = E\{\max[v_1, c_1 + \beta E[D_2]]\}$ represents the utility of the "not buy" alternative; therefore when calculating $E[D_2]$ we only take the terms corresponding to the right leave of the tree in the Figure 4.1. The calculation of $E[D_2]$ ($E[D_2] = E\{\max[v_2, c_2 + \beta E[D_3]]\}$) demands the same function to be calculated for period 3 ($E[D_3]$), which is assumed to be zero according to the above assumption.

  The process of calculating $E[D_1]$ is recursive with known utility at the end of the perspective horizon (assumed to be two periods long in this formulation). Having calculated $E[D_1]$, reservation utility at time 0 $W(y_0)$ can be obtained.

- These steps can be repeated to calculate $\pi_{i0,1}$ with the assumption that the respondent can anticipate alternative characteristics for time periods 2, 3 and $E[D_4]$ set equal to zero.

For a receding horizon, a terminal value for the expected utilities has to be fixed, therefore the expectation of the last time period under the person's perspective, $E[D(\boldsymbol{v}_{iT})]$ must have a constant value . Since it is difficult to predict a particular value, we assume it to be zero. In the long term, the individual has not enough information to predict the future; he/she cannot anticipate the utility of buying or postponing. Under this assumption, after a limited number of time periods information on future market trends is just ignored.

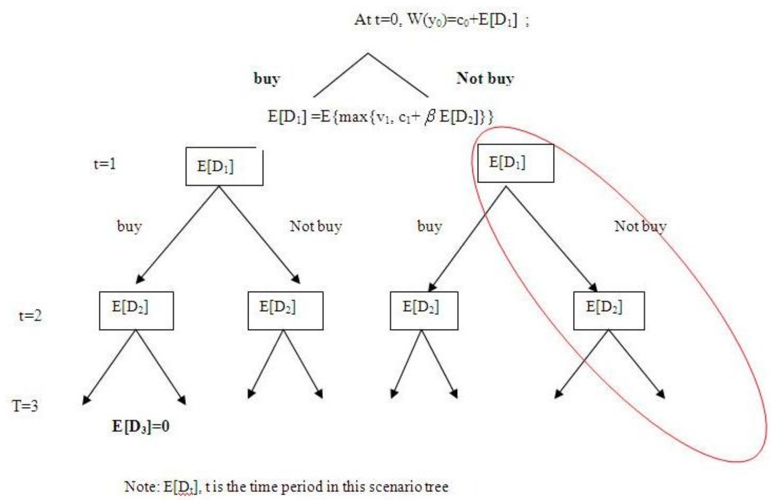

Fig. 4.1: Scenario tree

4.2 Conclusions

Dynamic models based on a regenerative optimal stopping problem of car ownership, and with non-linear formulation were presented in this Chapter. The car ownership model targeted consumers with heterogeneous characteristics and allowed the repeated purchases along the time periods. For each individual, we intended to explore the probability of buying or postponing the purchase of a vehicle; furthermore, if the individual chose to buy, the probability of buying a certain type of vehicle would be explored as well. Calculating the probability of postponing was the most significant step and it was calculated through the Reservation Utility. The reservation utility was constituted of the current vehicle information utility and the expected utility for the successive time period. The way to calculate the expected utility for each time period was a recursive process with the known terminal utility value at the final time period in the individual's perspective time horizon. The individual's perspective time horizon at each current time point was assumed to be two time periods; at the third time period which was also the terminal time of perspective time horizon, the individual was supposed to know nothing about the market trend information and the expected utility was set to zero. With the expected utility calculated, reservation utility was obtained and so was the probability of postponing. Maximum log-likelihood function was used as the optimization method to estimate parameters.

5. EXPERIMENTS USING SIMULATED DATA

Synthetic households' choices over different time periods have been simulated to validate the proposed dynamic discrete choice formulation. Simulated data are necessary to ensure that the model to be estimated is fully coherent with the hypothesis formulated by the analyst. In this case, data are created with known characteristics to test the ability of the dynamic formulation to recover the value of parameters used to generate the data (true values) and to reproduce individual choices over time (observed choices). The synthetic sample is composed of 200 individuals. Each of them is supposed to provide responses in 12 time periods starting from the current year; each time period is assumed to be six month long. A total of 2,400 observations are then generated. There are four alternatives in the choice set: (1)gasoline vehicle, (2) hybrid vehicle, (3) electric vehicle, and (4) keep the current vehicle. One important assumption made in the simulated process is that at each time period previous decisions affect the alternatives in the current choice set. If the previous period choice is "to buy", the current vehicle situation will be regenerated; the current vehicle will be the newly bought car and its age re-set to zero. If the previous period choice is "not buy", the current vehicle age is adjusted to reflect the fact that six months have passed since last decision.

5.1 Simulated Data Format and Generation

In total, five datasets are generated: household characteristics, current vehicle attributes, potential vehicle static attributes, potential vehicle dynamic attributes, and choice. In each file, the first column is the household identification number and constitutes the key variable used to merge the information from the five files. In appendix A, more details about the formatting of the five datasets are provided. The criteria that have determined the characteristics of the variables generated are described in the following subsections.

5.1.1 Household Characteristics

This set includes two variables: number of family members and household income. The number of family members is assumed to be uniformly distributed in the range 1-6. Household income is also assumed to be uniform and varies on four levels of variation: (1) low, (2) medium, (3) medium high and (4) high.

5.1.2 Current Vehicle Attributes

Car age is the only current vehicle attribute used in this experiment. Age for the first time period varies in the range: 0-10 years. Once age is generated randomly for the first time period, it is increased by 0.5 for each successive time period, unless a new car is bought which will imply that the car age in the following time period is 0.5.

5.1.3 Static Potential Vehicle Attributes

Vehicle size and vehicle price are the two attributes that characterize new vehicles. Vehicle size is classified as (1) small, (2) medium and (3) large. Small size vehicles have price range between \$15,000 and \$25,000. Medium vehicles have price range between \$25,000 and \$50,000. Large vehicles have price range between \$50,000 and \$70,000. The price for each size of vehicle is generated from an uniform defined within the corresponding range. Vehicle price in the simulated dataset is multiplied by 10^{-3}.

5.1.4 Dynamic Attributes

We limit our analysis to one dynamic attribute, which is future gasoline price. The function that defines gasoline price over time is generated using historical monthly prices in the past thirty years. Based on the assumption discussed in section 4.3.3, the dynamic variable is assumed to follow a normal diffusion process and specified as a random walk with drift. The calibration function is :

$$y_{t+1} = 0.9757 * y_t + 4.49 + draw \tag{5.1}$$

where: drift $\eta_j = 4.49$ and auto-regressive factor $\psi_j = 0.9757$. Monte Carlo simulation with 1000 draws is applied to generate gasoline fuel price for each time period. The average gasoline price for the initial time period is set to be \$3.5; future prices are generated from Equation 4.13 and transformed into equation 5.1; draws are obtained from a normal distribution $N(0, 16)$. In the utility function the price is expressed in cent.

5.1.5 Choice

Respondents choose between four alternatives (gasoline car, hybrid car, electric car and not buy). Decision in each time period is made after comparing totally purchase probabilities of different alternatives. Based on the methodology described in Chapter 4, the purchase probabilities among four alternatives at time period 0 are calculated, and compared. The individual makes the choice (with the highest probability) for time period 0. If the choice is buying a car, the car age in the following period will be 0.5; if the choice is not buying, the car age will be increased by 0.5 in the successive time period. After generating car age for the following time period, choices at this time period will be made by comparing all the probabilities again.

5.2 Utility Specification

Generally, respondents choose between two alternatives: buy and not buy a vehicle. If respondents choose to buy, they also need to decide which vehicle type they are going to buy amongst gasoline vehicle, hybrid vehicle or electric vehicle. Therefore, there are actually four alternatives in the choice set: gasoline car, hybrid car, electric car and not buy. Details of the utility specifications are given in equation 5.2. In the model, only one static potential vehicle attribute is used which is vehicle size.

$$
\begin{aligned}
&U_{i1t} = ASC1 + gap_t * \beta_{gap} + \epsilon_{i1t} + \epsilon_i \\
&U_{i2t} = ASC2 + hh\_no * \beta_{hh\_no} + income * \beta_{inc} + veh\_size2_t * \beta_{veh\_size2} \\
&+gap_t * \beta_{gap} + \epsilon_{i2t} + \epsilon_i \\
&U_{i3t} = hh\_no * \beta_{hh\_no} + income * \beta_{inc} + veh\_size3_t * \beta_{veh\_size3} + \epsilon_{i3t} + \epsilon_i \\
&U_{i4t} = current\_age_t * \beta_{current\_age} + \epsilon_{i4t} + \epsilon_i
\end{aligned}
\tag{5.2}
$$

where ASC is alternative specific constant; gap_t is gas price; hh_no is household number; veh_size2_t is vehicle size for hybrid car; veh_size3_t is vehicle size for electric car; $current\_age_t$ is current vehicle's age. ϵ_{ijt} is the random error term for each alternative at a given time period. ϵ_i is the individual error term that is assumed to be constant across all observations derived by the same respondent.

5.3 Model Estimation

Using the simulated data and the specification defined above two models were estimated: dynamic model and static MNL model.

The dynamic model algorithm derived from the formulation in Chapter 4 is coded in C language and makes use of a number of optimization tools derived from AMLET (Another Mixed Logit Estimation Tool), a mixed logit estimation software created by Fabian Bastin (http://amlet.slashbin.net/). Appendix E presents the C code developed for the probability calculation and the estimation process.

Static model is estimated using the software Alogit (http://www.alogit.com/).

In the static model, respondents are not considering future market evolution and possible vehicle technology improvements when making decisions in each time period. The model is simply formulated as the traditional MNL model with four alternatives; utilities include both static and dynamic variables, for consistency with the dynamic model formulation.

There are two experiments conducted with the simulated data sets. The first one is to estimate coefficients by both static and dynamic models based on the full sample size; the second one is to estimate coefficients by both models based on the first 160 samples out of 200, and then use the estimated parameters to forecast market shares of the rest 40 samples. Experiment two aims to show that the dynamic model can be used in different populations.

Table 5.1 and 5.2 show the results from the static and the dynamic models estimated on the whole and part set of simulated data. As expected, the fit of the model improves when considering the dynamic nature of the problem. In experiment one, the rho-squared increases from 0.5, the value obtained with the logit model and nine degrees of freedom to 0.7 for the dynamic model and eight degrees of freedom. In experiment two, rho-squared increases from 0.49 for the static model to 0.52 for dynamic model.

All estimated parameters, although not very close to the true values, are highly significant. The alternative specific constant values of static model have higher bias than those estimated with the dynamic formulation.

In order to validate which model better recover the true values, Root Mean Square Deviation (RMSD) is adopted as measure of the differences between the true values and the predicted values. The bigger the RMSD, the poorer the model's ability to reproduce the true phenomenon. The RMSD is defined as

$$RMSD(\hat{\theta}) = \sqrt{E((\hat{\theta} - \theta)^2)} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}{n}} \tag{5.3}$$

which n is the number of parameters. The last row of Table 5.1 and 5.2 reports the RMSD.

Alternative	Gas	Hybrid	Electric	Current	True	MNL		Dynamic	
					Values	Estim	t-Stat	Estim	t-Stat
ASC_gas	X				2.2	-0.97	0.9	2.79	6.6
ASC_gas-ASC_ele						2.29			
ASC_hyb		X			-0.4	-5.74	5.1	0.03	0.1
ASC_hyb-ASC_ele						-2.48			
ASC_ele			X			-3.26	11		
hh_no		X	X		-0.65	-0.14	3.8	-1.06	32.3
income		X	X		0.35	0.21	3.4	0.42	2.7
veh_size2		X			1.9	2.15	12.4	2.63	26.2
veh_size3			X		-0.1	-0.17	1.9	-0.13	2.9
gas_price	X	X			-1.4	-1.04	3.5	-2.12	19.2
current_age				X	-1.8	-1.37	17.3	-1.63	60
N observed						2,400		2,400	
LL(0)						-3327.10		-2472.27	
LL(final)						-1656.05		-685.05	
Likelihood ratio index						0.5		0.7	
RMSD						0.79		0.39	

Tab. 5.1: Model Estimation of Experiment One

5.4 Model Application

Coefficient estimates are used in application to calculate the prediction power of the models. The choice probability for each alternative observed and predicted together with a measure of errors are reported in Table 5.3 and 5.4. The absolute

Alternative	Gas	Hybrid	Electric	Current	True	MNL		Dynamic	
					Values	Estim	t-Stat	Estim	t-Stat
ASC_gas	X				2.2	-1.30	1.2	2.69	23
ASC_gas-ASC_ele						1.60			
ASC_hyb		X			-0.3	-6.11	4.9	-2.67	12
ASC_hyb-ASC_ele						-3.21			
ASC_ele			X			-2.90	9		
hh_no		X	X		-0.6	-0.16	3.5	-0.53	21
income		X	X		0.3	0.15	2.2	0.75	18
veh_size2		X			2	2.17	11	2.62	33
veh_size3			X		-0.1	-0.22	2.2	-0.11	2.9
gas_price	X	X			-1.4	-0.88	2.7	-1.23	34
current_age				X	-1.8	-1.30	15	-1.21	59
N observed						1,920		1,920	
LL(0)						-2661.68		-2369.82	
LL(final)						-1348.43		-1148.27	
Likelihood ratio index						0.49		0.52	
RMSD						1.10		0.92	

Tab. 5.2: Model Estimation of Experiment Two

error D is defined as

$$D = |M_{pred} - M_{obs}| \tag{5.4}$$

where D is error norm; M_{pred} is vector of model shares predicted; M_{obs} is vector of model shares observed. The results indicate that the dynamic model has stronger prediction power than the static model in the full sample experiment. In experiment one, D value from dynamic model is smaller than the value from static model; in experiment two, D values are the same for both models.

For experiment one, Figure 5.1, 5.2, 5.3, and 5.4 present the observed and predicted market trends of gasoline vehicle, hybrid vehicle, electric vehicle and keeping the current vehicle along the twelve time periods in the six years considered. The vertical axes indicates choice probability. In general, the market shares trends are

Alternative	Observed	Predicted (Static)	Predicted (Dynamic)
Gas car 1	0.005	0.0185	0.005
Gas car 2	0.025	0.039	0.015
Gas car 3	0.045	0.0595	0.035
Gas car 4	0.06	0.0695	0.03
Gas car 5	0.05	0.0585	0.035
Gas car 6	0.05	0.056	0.035
Gas car 7	0.085	0.06	0.06
Gas car 8	0.035	0.058	0.035
Gas car 9	0.085	0.0535	0.06
Gas car 10	0.06	0.044	0.03
Gas car 11	0.035	0.034	0.025
Gas car 12	0.055	0.0375	0.035
Hybrid car	0.015	0.0315	0.02
Hybrid car 2	0.08	0.066	0.075
Hybrid car 3	0.075	0.0825	0.07
Hybrid car 4	0.095	0.1037	0.105
Hybrid car 5	0.095	0.088	0.085
Hybrid car 6	0.1	0.095	0.1
Hybrid car 7	0.11	0.095	0.105
Hybrid car 8	0.1	0.0945	0.095
Hybrid car 9	0.105	0.0815	0.095
Hybrid car 10	0.045	0.0615	0.04
Hybrid car 11	0.055	0.059	0.035
Hybrid car 12	0.035	0.0525	0.02
Electric car 1	0.035	0.054	0.03
Electric car 2	0.065	0.09	0.06
Electric car 3	0.11	0.124	0.085
Electric car 4	0.19	0.15	0.15
Electric car 5	0.12	0.15	0.105
Electric car 6	0.165	0.144	0.14
Electric car 7	0.14	0.1357	0.13
Electric car 8	0.1	0.122	0.085
Electric car 9	0.14	0.135	0.125
Electric car 10	0.18	0.1395	0.165
Electric car 11	0.115	0.1395	0.125
Electric car 12	0.19	0.167	0.19
Current car 1	0.945	0.8955	0.945
Current car 2	0.83	0.8045	0.85
Current car 3	0.77	0.7345	0.81
Current car 4	0.655	0.6775	0.715
Current car 5	0.735	0.705	0.775
Current car 6	0.685	0.705	0.725
Current car 7	0.665	0.7075	0.705
Current car 8	0.765	0.725	0.785
Current car 9	0.67	0.73	0.72
Current car 10	0.715	0.755	0.765
Current car 11	0.795	0.7675	0.815
Current car 12	0.72	0.743	0.755
D		1.005	0.88

Tab. 5.3: Model Validation: Market Shares of Experiment One

Alternative	Observed	Predicted (Static)	Predicted (Dynamic)
Gas car 1	0	0.0225	0.0136
Gas car 2	0.025	0.045	0.022
Gas car 3	0.025	0.0575	0.059
Gas car 4	0	0.0675	0.034
Gas car 5	0.05	0.0675	0.043
Gas car 6	0.025	0.065	0.085
Gas car 7	0.075	0.075	0.044
Gas car 8	0	0.063	0.023
Gas car 9	0.075	0.0525	0.03
Gas car 10	0.05	0.038	0.03
Gas car 11	0	0.038	0.031
Gas car 12	0	0.045	0.070
Hybrid car	0	0.03	0.038
Hybrid car 2	0.125	0.085	0.104
Hybrid car 3	0.075	0.108	0.079
Hybrid car 4	0.05	0.12	0.151
Hybrid car 5	0.15	0.098	0.169
Hybrid car 6	0.075	0.085	0.091
Hybrid car 7	0.15	0.0925	0.102
Hybrid car 8	0.175	0.118	0.099
Hybrid car 9	0.225	0.088	0.089
Hybrid car 10	0	0.053	0.054
Hybrid car 11	0.025	0.055	0.029
Hybrid car 12	0.05	0.05	0.061
Electric car 1	0.025	0.06	0.04
Electric car 2	0.075	0.1	0.078
Electric car 3	0.075	0.118	0.065
Electric car 4	0.275	0.14	0.107
Electric car 5	0.075	0.153	0.106
Electric car 6	0.15	0.156	0.143
Electric car 7	0.1	0.16	0.136
Electric car 8	0.15	0.133	0.091
Electric car 9	0.125	0.13	0.097
Electric car 10	0.1	0.108	0.099
Electric car 11	0.175	0.135	0.155
Electric car 12	0.225	0.173	0.161
Current car 1	0.975	0.888	0.907
Current car 2	0.775	0.773	0.795
Current car 3	0.825	0.718	0.798
Current car 4	0.675	0.67	0.709
Current car 5	0.725	0.685	0.682
Current car 6	0.75	0.693	0.681
Current car 7	0.675	0.673	0.719
Current car 8	0.675	0.685	0.786
Current car 9	0.575	0.728	0.785
Current car 10	0.85	0.805	0.819
Current car 11	0.8	0.773	0.784
Current car 12	0.725	0.733	0.708
D		2.0	2.0

Tab. 5.4: Model Validation: Market Shares of Experiment Two

quite realistic; the probability of keeping the current car is relatively high, starting at 95% in the first time period and declining to about 75% after four-five time periods. New gasoline vehicles, hybrid and electric vehicles occupy smaller market shares; all types become more popular in later time periods with the exception of hybrid vehicles that have a significant drop after the tenth period. The four figures indicate that the choice probability obtained from the dynamic model are closer to the observed market shares when compared to those obtained from the static model. The static model is able to capture the general trend through the years but fails to correctly recover peaks and suddenly changing behaviors.

Fig. 5.1: Market Trend for Gasoline Car-Experiment 1

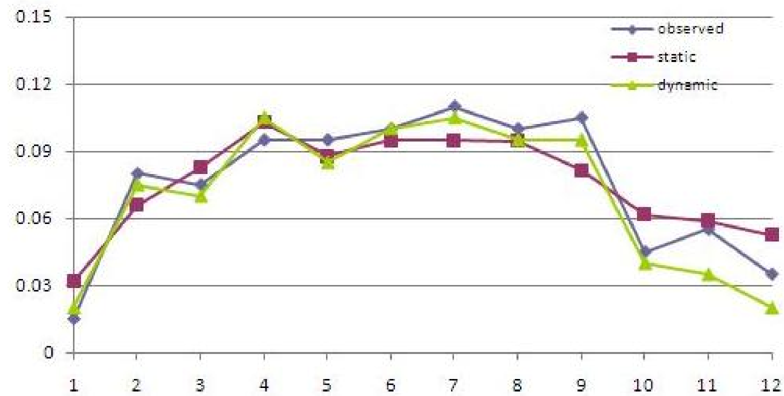

Fig. 5.2: Market Trend for Hybrid Car-Experiment 1

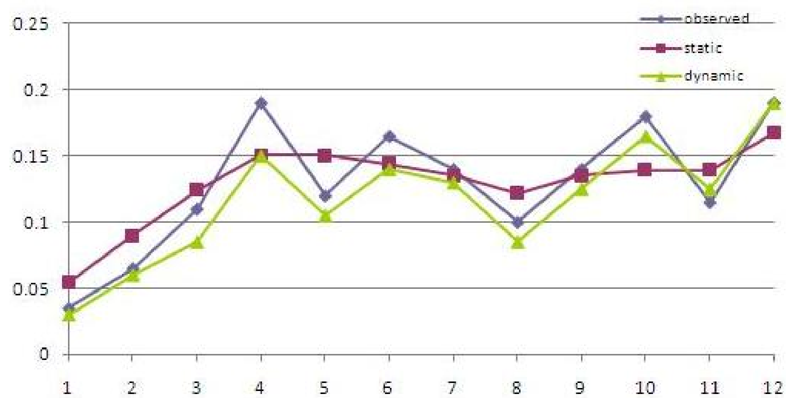

Fig. 5.3: Market Trend for Electric Car-Experiment 1

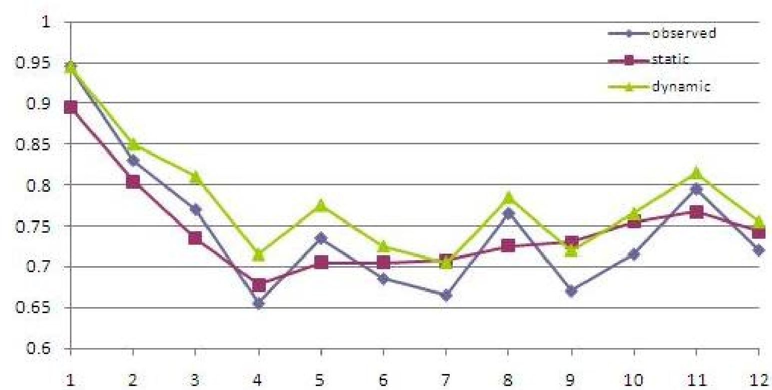

Fig. 5.4: Market Trend for Current Car-Experiment 1

For experiment two, Figure 5.5, 5.6, 5.7, and 5.8 present the observed and predicted market shares of the rest 40 samples. The four figures show similar market trends of the alternative vehicles as the trends from experiment one. But the predicted values have bigger bias to the observed ones, although the dynamic model is still able to capture the changes of choices through time periods.

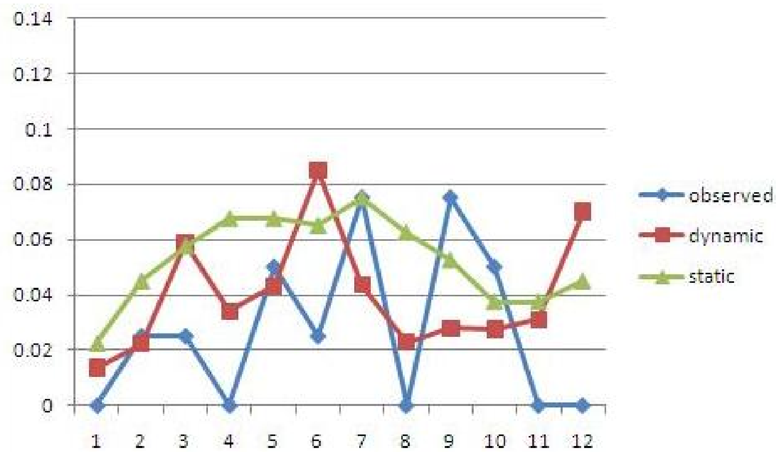

Fig. 5.5: Market Trend for Gasoline Car-Experiment 2

Fig. 5.6: Market Trend for Hybrid Car-Experiment 2

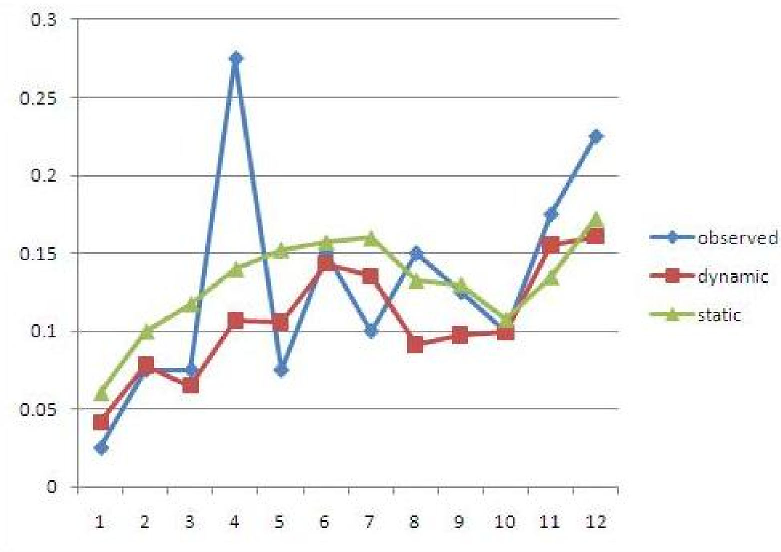

Fig. 5.7: Market Trend for Electric Car-Experiment 2

Fig. 5.8: Market Trend for Current Car-Experiment 2

5.5 *Conclusion*

In this Section results obtained from both static and dynamic model formulations were presented. Models were applied to a simulated dataset containing 200 households each observed along a twelve-time period. Two experiments were done in this chapter. One was estimating coefficients and applying them to models based on the full sample size; another experiment was estimating coefficients from the first 160 samples and applying the predicted values to models based on the rest 40 samples. The analysis indicates that the dynamic model outperforms the static model in terms of goodness of fit and that it produces estimators which are closer to the true values used to produce the synthetic population (smaller bias). More investigations are, however, needed in order to understand why significant bias exists for both types of models. The models have been also applied in order to test the ability to reproduce the observed choices. Again, the dynamic model is superior to the traditional logit model, as the absolute error D between observed and predicted choices is smaller for the dynamic case in experiment one. It can be observed that the static logit model recovers the general trend in the market, but fails to detect peaks in choice distributions as results of rapidly changing market conditions. Bias is bigger in magnitude when estimating on the part set of samples and validating on the rest of the samples (due to the smaller sample sizes); but the dynamic model is still capable to mimic the real trends of market shares in experiment 2.

6. SURVEY DESIGN AND METHODOLOGY

In order to apply the models discussed in Chapter 4 to real data, a survey has been designed and executed. This Chapter presents the design of a car ownership survey that includes Revealed Preference (RP) questions and three games of Stated Preference (SP) questions. This survey is web based and was designed and implemented by Michael Maness [Maness, 2010].

6.1 Survey Design

The survey aimed at exploring consumers' preferences over future gasoline vehicle, hybrid vehicle, and electric vehicle. Forecasting the demand for new technology in transportation requires information about users' preferences for services that do not exist in the current system. SP data are commonly used to provide behavioral choice information in hypothetical contexts. The survey was divided into three parts: (1) household characteristics; (2) current vehicle and (3) Stated Preference games. The survey consisted of approximately 50 questions; a printed version of the survey is in appendix B.

6.1.1 Household Characteristics

In this section, the individual is asked to describe his her socioeconomic situation via the following questions:

- Gender
- Age
- Education Level
- Head of Household
- Work Status
- Driver's License
- Commute Distance
- Work Parking

The questions also include household information:

- Household Income
- Number of Kids
- Number of Adolescents
- Number of Adults
- Number of Workers
- Household Location

- Home Type
- Five-Year Vehicle Purchase Plans
- Number of Vehicles
- Make/Model of Vehicles
- Home Parking

6.1.2 Current Vehicle

Current Vehicle section gathers data on households' primary vehicle characteristics for possible use in the SP games and modeling. This section has the following questions:

- Vehicle Type
- Model Year
- Purchase Year
- Miles Traveled per Year
- Fuel Type
- Price
- Fuel Economy
- Tank Capacity
- Seating Capacity

6.1.3 Stated Preference

The stated preference (SP) portion of the survey presents respondent with one of the three SP experiments: (1) Vehicle Technology; (2) Fuel Type; and (3) Tolling and Taxing. Each respondent randomly receives one SP game. Game (1) has a 50% chance of being picked while games (2) and (3) each have a 25% chance.

Each stated choice game generates multiple SP scenarios over a six year time period, from 2010 to 2015. The variables in the scenarios change from year to year. For example, vehicle price generally increases over time, hybrid tax credit decreases with time, and the range for gasoline vehicles remains constant. Two scenarios per year are presented for a total of 12 observations.

Respondents are given the following instructions when making decisions:

- Make realistic decisions. Act as if the respondent were actually buying a vehicle in a real life purchasing situation. If you would not 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 your current living situation has moderate increases in income from year to year.

- Each scenario is independent from one another. For example, if you purchase a vehicle in 2011, then in the next scenario forget about the new vehicle and just assume you have your current real life vehicle. This instruction is not adapted to the dynamic model data collection rule. Because in the dynamic model, if the person chooses to buy a vehicle in 2011, he/she is supposed to keep the new vehicle and increase vehicle age by 0.5 in the next scenario based on the description in Chapter 5.

Game 1 - Vehicle Technology

The vehicle technology game focuses on presenting respondents with different characteristics for the vehicles and pricing to discover preferences for vehicle technology. This game design consists of four alternatives and five variables. Each variable has 16 - 24 levels of variation per alternative (about four levels per vehicle size). Respondents have a choice set size of eight.

The four alternatives include the current owned vehicle and a new gasoline vehicle , hybrid vehicle, and electric vehicle. Gasoline vehicles are the traditional option. Hybrid vehicles are growing in the market which is led by the Toyota Prius, in the US. Although electric vehicles are new to the marketplace, major automobile manufacturers have significant interest in exploring this paradigm, such as the Nissan Leaf.

The variables of interest in the vehicle technology game include: vehicle price, fuel economy, refueling range, emissions, and vehicle size.

Vehicle price is a major factor in the household vehicle purchase decision. Prices depend on the size of the vehicle and increase from year to year. For gasoline and hybrid vehicles, the base price was determined from the average vehicle price for each vehicle type and size. For electric vehicles, the base price was determined by the average projected price of future electric vehicles and/or European prices (if a similar vehicle is sold in Europe). This base price is increased by 2% per year.

From this base price, the other three levels are determined by increasing the base price by 4.5%, 9%, and 18%.

Vehicle fuel economy which is presented in miles per gallon (MPG), has an impact on the operating cost of a vehicle. For electric vehicles, the fuel economy is not presented (when research began, a standard for displaying fuel economy of electric vehicles was not established by the Environmental Protection Agency). Fuel economy begins with a base value that is the average of current vehicle MPG per vehicle type and class. This base was set to be constant from 2010 to 2015 because the recent history record shows that the average fuel economy changes slightly. The fuel economy for the other three classes varies linearly from factors of 1.07, 1.13, and 1.18 in 2010 to factors of 1.10, 1.25, and 1.50 respectively in 2015. This formulation accounts for the uncertainty in fuel pricing over the next five years.

A vehicle's range which is miles between refueling periods plays an significant role in the decision of choosing electric vehicles considering commute distances and their long recharge times. Gasoline and hybrid vehicle have refueling ranges of approximately 300-500 miles. Electric vehicles have refueling ranges primarily dependent on vehicle size. In this stated choice game, the refueling range for gasoline and hybrid vehicles do not vary in the periods. For electric vehicles, the range level chosen are dependent on projected ranges for current and future vehicles. The levels are set in 2010 and are generally increased geometrically by a factor between 1.05 and 1.1 depending on the detail in the data collected on range estimates by size.

A vehicle emissions variable was introduced to test if emission levels significantly influenced household vehicle purchasing decisions. The designer chose to present emission level as a percent difference versus the average 2010 vehicle(about 24 mpg). Levels for electric vehicle emissions were zero.

Vehicle sizes chosen are based only on designs that could be found in literature. The size system used is an abbreviated version of the EPA size class: small/compact car, midsize car, large car, minivan, sports utility vehicle, and pickup truck.

The choice set for the vehicle technology experiment includes all combination of buying a new vehicle (gasoline, hybrid, or electric) or not buying with selling or retaining the current vehicle. A summary of the vehicle technology game is available in Table.

Variables	Vehicle Price Fuel Economy Refueling Range Emissions Vehicle Size
Alternatives Shown	Current Vehicle New Gasoline Vehicle New Hybrid Vehicle New Electric Vehicle
Choice Set	I Will KEEP My Current Vehicle I Will BUY the Gasoline Vehicle And SELL My Current Vehicle I Will BUY the Hybrid Vehicle And SELL My Current Vehicle I Will BUY the Electric Vehicle And SELL My Current Vehicle I Will BUY the Gasoline Vehicle And KEEP My Current Vehicle I Will BUY the Hybrid Vehicle And KEEP My Current Vehicle I Will BUY the Electric Vehicle And KEEP My Current Vehicle I Will SELL My Current Vehicle and NOT REPLACE It

Tab. 6.1: Vehicle Technology Game Summary

Game 2 - Fuel Type

The fuel type game presents respondents with different fuel options for future vehicle purchases. This game design consists of four alternatives and four variables.

Each variable has three or six levels of variation per alternative. Respondents have a choice set size of seven. The four alternatives (fuel types) are: gasoline, alternative, diesel, and electricity. These fuel types are currently available in Maryland's marketplace - gasoline, alternative (ethanol, E85), diesel via some gas stations, and electricity via the home.

The variables of interest in the fuel type game include: fuel price, fuel tax, average fuel economy, and refueling availability.

Fuel price (pre-tax) is a operating cost of a vehicle and measured in US dollars per unit of energy. Gasoline price is in dollars per gallon of gasoline. Alternative fuel price is in dollars per gallon of alternative fuel. Diesel price is in dollars per gallon of diesel fuel. Electricity price is in dollars for 33.7 kwh of electricity, which is the electrical equivalent of the energy in one gallon of gasoline. The fuel price for the liquid fuels (gasoline, alternative, and diesel) is based on historical data in the Mid-Atlantic region, mostly from the US Department of Energy. Gasoline and diesel prices range from \$2.00 to \$4.00 in 2010. Alternative fuel prices, which are based on E85, are 10% less than gasoline prices based on historical data (due to subsidies and lower energy density). Electricity prices for 2010 scenarios are based on residential prices in Maryland during June 2009 and a four cent per kwh change in price for the levels. The prices were assumed to vary geometrically at an annual rate of 1.10 for liquid fuels and 1.03 for electricity.

Fuel tax is measured in dollars per fuel unit. The fuel tax varied by three levels, the current tax and two higher tax rates. Tax levels did not change annually because there has been no history in Maryland of tying fuel tax rates to inflation.

Fuel efficiency means how much fuel is used per mile traveled and provide respondents with an idea of how efficient their vehicle choice could be in fuel economy. For liquid fuels, the fuel efficiency base level is based on current data regarding average fuel economy in the US by fuel type with one level being higher and one level being lower. Electric efficiency was difficult to find as there are various methods of presenting the efficiency of electric vehicles. Therefore the base vehicle used was the preliminary fuel economy sticker from a Mini E, which has a fuel efficiency of 100 miles per gallon equivalent (mpge). To be conservative, the other two levels for electric fuel efficiency were lower. Fuel efficiency was assumed to increase annually by 2 mpg for gasoline and diesel, 1 mpg for alternative fuel, and 5 mpge for electricity.

Fueling station availability was represented by the distance from home to the nearest station for liquid fuels and the time to charge the vehicle at home for electricity. This variable does not change over time for gasoline and diesel fuels. Availability increases (distance from home decreases) for alternative fuel over time and the charging time decreases over time for electricity.

The choice set for this game includes keeping and selling the respondent's current vehicle or buying a new vehicle that runs on one of the fuel choices. A summary of the fuel type game is available in Table 6.2[Maness, 2010].

Game 3 - Taxation policy

The taxation policy game presents respondents with different toll and tax policies impacting on their effect on future vehicle purchases. For the 2010 and 2011 scenarios, the game design consists of four alternatives and two variables with three

Variables	Fuel Price, Before Tax Fuel Tax Fuel Efficiency Fueling Station Availability
Alternatives Shown	Gasoline Fuel Alternative Fuel Diesel Fuel Electric
Choice Set	I Will KEEP My Current Vehicle I Will BUY a Gasoline Vehicle (or normal hybrid) that runs on Gasoline I Will BUY an Alternative Fuel Vehicle that runs on Alternative Fuel I Will BUY a Diesel Vehicle that runs on Diesel Fuel I Will BUY an Electric Vehicle that runs on Electric Fuel I Will BUY a Plug-In Hybrid Electric Vehicle that runs on Gasoline and Electric Fuel I Will SELL My Current Vehicle and NOT REPLACE It

Tab. 6.2: Fuel Type Game Summary

levels of variation per alternative. The choice set size is eight. For the 2012 through 2015 scenarios, the game design consists of four alternatives, three variables with three levels of variation per alternative, and nine choices.

The four alternatives include the current vehicle and a new gasoline vehicle, hybrid vehicle, and electric vehicle. The variables of interest in the tolling and taxing game include: income tax credits, toll cost, and vehicle-miles traveled (VMT) tax rate (2012-2015).

The income tax credit is shown for hybrid and electric vehicles based on current federal guidelines for credits. It is used to encourage adoption of new technology through reducing one's tax burden.

The toll policy variable attempts to encourage adoption of new technology by reducing toll costs for users of that technology. This variable is presented to respondent as the percent reduction in normal toll prices for users of hybrid and electric vehicles.

The VMT tax is to encourage adoption of new technology by reducing the operating cost of using the vehicle. The VMT tax rate is presented as a cost (in US dollars) per 1000 miles traveled that is collected by the respondent's insurance provider.

The choice set for the taxation policy experiment includes all combination of buying a new vehicle (gasoline, hybrid, or electric) or not buying with selling or retaining the current vehicle. 8 possible choices for 2010 and 2011 scenarios are available. For the 2012 through 2015 scenario, an additional choice is added to keep one's current vehicle and drive less.

Variables	Income Tax Credit Toll Price VMT Tax (2012-2015)
Alternatives Shown	Current Vehicle New Gasoline Vehicle New Hybrid Vehicle New Electric Vehicle
Choice Set	I Will KEEP My Current Vehicle I Will KEEP My Current Vehicle and DRIVE LESS (2012-2015) I Will BUY the Gasoline Vehicle And SELL My Current Vehicle I Will BUY the Hybrid Vehicle And SELL My Current Vehicle I Will BUY the Electric Vehicle And SELL My Current Vehicle I Will BUY the Gasoline Vehicle And KEEP My Current Vehicle I Will BUY the Hybrid Vehicle And KEEP My Current Vehicle I Will BUY the Electric Vehicle And KEEP My Current Vehicle I Will SELL My Current Vehicle and NOT REPLACE It

Tab. 6.3: Tolling and Taxing Game Summary

6.2 Survey Methodology

The survey was conducted by volunteers from September to October in 2010. The target population are households living in suburban and urban Maryland State and five counties are selected as the sample districts. Volunteers used two modes to conduct the survey: face to face to help people finish the survey and distribute flyers printed with the link of website to people and make a very short introduction.

6.2.1 Sample design

A two stage cluster design was adopted for this survey. Household were clustered by county then by zip code. This approach was used for cost and human capital reasons. Therefore there may be some sampling error from this technique as some biases may develop from a pseudo-random clustering of zip codes. Volunteers were also allowed to decide on the recruitment method in those zip codes which was either limited to door-to-door flyer handouts or flyer handouts at a local gathering place (e.g. mall, supermarket). The distribution of respondent location is available in appendix D.

6.3 Platform for the Web-based Survey Design

Free software commonly used for online surveys handles simple questions (e.g. multiple choice, open-ended) and basic question ordering functions; however, to the best of our knowledge, web-based survey applications suited for stated choice games are not available in open-source. Therefore, a web-based survey framework (called Julie) that could perform stated-choice experiments was specifically created to collect real data. JULIE depends on the following:

- Ruby, a programming language
- Rails, a web-application framework
- SQLite, a database
- JULIA, a domain-specific language for creating surveys

JULIE was built using the dynamic, reflective, object-oriented Ruby programming language. Ruby is the backbone of JULIE and provides it with the capabilities it needs to perform calculations and conditional logic in order to create and customize surveys to different purposes and different respondents. In JULIE, the model and controller are based in Ruby while the view is based in HTML. The basic premise is that JULIE displays a sequence of questions to the user, records his responses, and ensures that the responses are in the expected format. This process is essential as the surveys are designed to be self-administered online, an environment where the researcher cannot help the respondent.

The view is written in HTML with embedded Ruby. The view visually provides the user with a place to see and respond to questions. The view primarily consists of two parts: the survey and scenario views. These views have corresponding controllers with the same name. The survey view displays questions corresponding to all question types except choice experiments. The scenario view displays the scenarios from the stated choice experiments.

6.4 Conclusion

This Chapter describes in detail the design of the car ownership survey which includes RP questionnaire and three games of SP questions. The SP experiments were designed in order to collect future vehicles preferences in the case that (1) new technology vehicles will be available in the market; (2) fuel prices will drastically increase and that new and more environmental friendly fuel types could be purchased; and (3) tolls will be introduced and new tax schemes implemented. The survey is web-based and created by software JULIE which is self-programmed. The target population were residents of the State of Maryland.

7. DESCRIPTIVE STATISTICS

The web-survey was executed from September 10th through October 31st, 2010. It is observed that once the respondents accepted to login in our website the majority of them completed the questionnaire. This resulted into a completion rate of 94%; the final sample include 141 valid responses, while 13 incomplete questionnaires were discarded. Given the relative small size of the sample available, it is not possible to generalize the findings derived from both the descriptive statistics and from the model estimation. Real data are collected in this context to verify that dynamic behavior can be captured by way of Stated Preference methods and that the deriving responses can be used to estimate a dynamic discrete choice model. The remaining of this chapter essentially describes the main characteristics of the sample and gives details about the stated choices collected from Game (1), which has been used for estimation. The contents given here heavily draw from Michael Maness's Master thesis completed in December, 2010.

7.1 Socioeconomics Results

Socioeconomic information includes respondent's gender, age, education, household position, occupation, and commute time; and household characteristics, income, number of workers in the household, location, and building type. The results are shown in 7.1

- Gender. 52% of respondents were male.

- Age. The average respondents' age was 43 and the median age 41. The youngest respondent was 18 and the oldest respondent was 83.

- Education. 77 respondents had graduate or professional degrees, 30 respondents had bachelor degrees, and 7 respondents had associate degrees. Of the remaining respondents, 1 did not have a high school diploma, 12 respondents had high school diplomas, and 14 respondents had some college coursework.

- Head of Household. 61% of respondents were the head of their household.

- Income. The income distribution was generally above the Maryland median. 22% of households had incomes above $150,000. 21% of household had incomes between $100,000 and $149,999. 18% of households had incomes between $75,000 and $99,999. 12% of households had incomes between $50,000 and $74,999. 15% of households had incomes between $25,000 and $49,999. 8% of households had incomes less than $25,000 with the remaining households (4%) refusing to answer the question.

- Household Age Distribution. The average household size was 2.74 people with 2.07 adults per household, 0.45 children under 12 years old and 0.22 adolescents. The median household size was 2.00.

- Workers. The average number of workers per household was 1.63 with a median number of workers 2.00. One household chose not to respond.

- Location. The majority of respondents (138) lived in the state of Maryland with one respondent each from Virginia, Washington, D.C. and "Other." The most common zip codes of respondents were 20770 (Greenbelt), 20850 (Rockville), 21213 (Baltimore), and 20877 (Gaithersburg). By county, 48% of all households were located in Montgomery County, 25% in Prince George's County, 9% in Anne Arundel County, 6% in Baltimore City, 5% in Howard County, and the remainder from Kent and Frederick Counties and outside of Maryland. This corresponds to an eligible respondent rate of 93%. Appendix D shows the distribution of locations on a map. Two respondents did not respond to the zip code question and two respondents provided invalid responses.

- Home Type. A majority of respondents lived in single-family dwellings: 62 respondents lived in detached houses and 45 respondents lived in townhouses or rowhouses. In the remaining samples, 21 described their home as an apartment, 9 lived in condominiums, and 1 lived in student housing. One household did not respond to the question.

- Work Status. The work statuses of respondents were generally full-time (104 respondents). Eight respondents described themselves as part-time workers, six were homemakers, five were students, twelve were retired, and three were "Other." Only three described themselves as "looking for work" and it is not clear whether the people in the "Other" category were unemployed or not.

- Commute Time. The round-trip commute time of working respondents (full-time, part-time, and students) was 30 minutes on average. The median commute time was 24 minutes with a maximum commute of 130 minutes. 39% of commuters had commute times of 15 minutes or less, while 25% had commute times between 16 and 30 minutes. 25% of commuters had commute time between 31 and 60 minutes. 8% of commuters had round-trip commutes of over an hour. One household chose not to respond.

- Driver's License. 96% of respondents had driver's licenses.

- Home Parking. 20% of respondents have personal garages, 28% have driveways, 20% park on-street, and 23% park in outdoor lots.

- Work Parking. 87% of workers said that parking was available at their workplace. The median parking cost was $0 with 75 out of 100 workers stating that they had free work parking. Of the workplaces with parking costs, the average parking cost was $100 per month. The highest parking cost stated was $300.

7.2 Current Vehicle Characteristics

In the RP survey, respondents were asked about their household vehicles and the characteristics of their current primary vehicles.

- Vehicles Per Household. 35% of household had one vehicle, 34% had two vehicles, and 21% had three vehicles.

- Primary Vehicle Make and Model. Of the 134 households with at least one vehicle, only 11 respondents did not provide a model name and 3 respondents provided a model name without corresponding make. This is a 90% appropriate response rate.

- Primary Vehicle Size. 38% of respondents used a compact/small car as their primary vehicle. 24% drove a mid-size car and 10% drove a large car. 16% of households used a pickup truck as a primary vehicle. Of the remaining household, 6% drove a van as primary transport and 1% of households used a sports utility vehicle (SUV) as a primary vehicle.

- Primary Vehicle Age. The average age of primary vehicles was 6.37 years with a median age of 6.00 years. 36% of primary vehicles were less than five years old, 44% were six to ten years old, and 20% were over ten years old. Two households skipped this question.

- Primary Vehicle Mileage. The average annual mileage was about 15,000 miles. The median mileage was 10,000 miles. Twenty-five respondents (18%) did not know the average annual mileage of their primary vehicle.

- Primary Vehicle Hybrid. 7% of households used a hybrid electric vehicle as their primary vehicle.

- Primary Vehicle Purchase Condition. The purchase condition of 63% of the primary vehicles was new, with the remaining 37% of vehicles purchased used or pre-owned.

- Primary Vehicle Purchase Price. The average vehicle purchase price was \$19,245 with a median price of \$18,000. The minimum purchase price was \$1,500 and the maximum purchase price was \$46,000. The average purchase price of new vehicles was \$23,763. The average purchase price of used vehicles was \$11,367. This question had a 4% nonresponse rate.

- Primary Vehicle Fuel Economy. The average fuel economy was about 27 miles per gallon. The median fuel economy was 25 mpg. 24% of respondents did not know their vehicle's fuel economy.

- Primary Vehicle Fuel Capacity. The average fuel economy of vehicles was 14.59 gallons with a median of 12 gallons. 22% of respondents did not know their vehicle's fuel capacity.

- Purchase Plans. 62% of respondents planned to buy a vehicle within five years. Of those respondents, 36 respondents plan to buy a new vehicle, 40 respondents plan to buy a used vehicle, and 10 respondents had no preference.

7.3 Stated Preference experiment: vehicle technology game

This Section describes the main characteristics of the sample collected from the SP survey on vehicle technology. Eighty-nine respondents were given this experiment, each of them was presented to 12 scenarios; a total of 1068 valid responses were collected from this game.

Table 7.3 summarizes how often respondents bought new vehicles in the proposed scenarios. Data in the second and third columns describe the choice frequency of respondents who intended to purchase a vehicle over the next five years. The fifth and sixth columns describe choice frequencies for respondents who did not intend to purchase a vehicle over the next five year.

As expected, those who expressed the intention to buy a vehicle had a higher choice frequency for new vehicles than people who did not intend to buy. However, those who initially were not inclined to buy, showed a propensity towards the purchase of new vehicles, especially when technology offered interesting options or fuel price was at very high rates.

Table 7.4 summarizes how often respondents bought new technology vehicles in the scenarios. Data in the second and third columns describe the choice frequency of respondents who intended to purchase a vehicle over the next five years. The fifth and sixth columns describe choice frequencies for respondents who did not intend to purchase a vehicle over the next five year.

By comparing results between Table 7.3 and Table 7.4, we see that respondents who did not intend to purchase a vehicle tended to buy non-conventional vehicles over gasoline vehicles. This may imply that household individuals who did not intend to buy a new vehicle because their current vehicle are their most preferable gasoline vehicle. It may also imply that as gasoline prices increase or a VMT tax is implemented, all households will tend to explore non-conventional options, such as hybrid and electric vehicles.

Table 7.5 summarizes the choices made by all the respondents. The percentage of respondents choosing to keep their current vehicle gradually decreases over time from about 75% to about 50%. New gasoline, hybrid, and electric vehicles occupy similar shares of the market. This shows that maturing vehicle technology may have an impact on adoption rates as the expected. Finally, it should be noted that almost none of our respondents decide to sell the car and to not replace it.

7.4 Conclusion

The SP sample collected in Maryland in fall 2010 is characterized by respondents that are middle aged, highly educated, have middle or high income, and live in townhouses or detached houses. The majority of them owns one or more cars, most of these family cars are compact and mid-sized, and in average five year old. The current cars have average fuel economy. Most of the respondents plan to purchase a car in the next five years. In the near future, people will gradually give up their current vehicle. Maturing vehicle technology has an impact on people's willingness to buy a new car and in particular on the adoption of new hybrid and electric vehicles.

	Category	Percentage
Gender	Male	52%
Age	26-35	23%
	36-45	24%
	46-55	18%
	56-55	17%
Education	Less than High School	1%
	High School Diploma or Equivalent	9%
	Some College	10%
	Associate	5%
	Bachelor Degree	21%
	Graduate or Professional Degree	55%
Head of Household	Yes	62%
Income	above $150,000	22%
	$100,000 - $149,999	21%
	$75,000 - $99,999	18%
	$50,000 - $74,999	12%
	$25,000 and $49,999	15%
	less than $25,000	8%
Home Type	Dorm/Student Housing	1%
	Apartment	15%
	Condo	6%
	Townhouse	30%
	Rowhouse	1%
	Detached Home	44%
Work Status	Full Time	74%
	Part Time	6%
	Retired	9%
	Student	4%
	Homemaker	4%
Commute Time	15 minutes or less	39%
	16-30 minutes	25%
	31-60 minutes	25%
	over 60 minutes	8%
Home Parking	Personal Garage	20%
	Personal Driveway	28%
	On-street	20%
	Outdoor Parking Lot	23%
	Parking Garage	2%
Work Parking	Yes	87%

Tab. 7.1: Socioeconomics Results

	Category	Percentage
Cars per Household	None	5%
	One	35%
	Two	34%
	Three	21%
	Four	4%
	More than Four	1%
Primary Vehicle Size	Compact car	38%
	Mid-size car	24%
	Large car	10%
	Van	6%
	SUV	1%
	Pickup truck	16%
	No car	5%
Primary vehicle age	Less than five years old	36%
	Six to ten years old	44%
	Over ten years old	20%
Primary vehicle anuual mileage	Above 20000	17%
	15000-19999	11%
	10000-14000	28%
	5000-9999	18%
	0-4999	4%
	Unknown	18%
Primary Vehicle Hybrid	Yes	7%
Primary Vehicle Purchase Condition	New	63%
	pre-owned	37%
Primary Vehicle Purchase Price ($)	40000-50000	4%
	35000-39999	3%
	30000-34999	9%
	25000-29999	11%
	20000-24999	12%
	15000-19999	22%
	10000-14999	9%
	5000-9999	8%
	0-4999	16%
Primary Vehicle Fuel Economy	20-24	21%
	15-19	7%
	25-29	19%
	30-39	16%
	40-49	6%
Primary Vehicle Fuel Capacity	11-15	37%
	16-20	23%
Purchase Plans	Plan to Buy a New Vehicle within 5 Years	62%

Tab. 7.2: Current Vehicle Characteristics

SP Game	Intend to Buy	Bought a New Vehicle	Rate	Not Intend to Buy	Bought a New Vehicle	Rate
Game1:Vehicle Tech	696	312	44.8%	372	108	29.0%

Tab. 7.3: Scenarios In Which Respondents Bought a Vehicle

SP Game	Intend to Buy	Bought a New Vehicle	Rate	Not Intend to Buy	Bought a New Vehicle	Rate
Game1:Vehicle Tech	696	189	27.2%	372	70	18.8%

Tab. 7.4: Scenarios in Which Respondents Bought a New Non-Conventional Gasoline Vehicle

Vehicle Type Choice	Index	2010	2011	2012	2013	2014	2015
Current vehicle	0	74%	66%	63%	63%	47%	50%
New Gasoline Vehicle	1	10%	13%	14%	10%	13%	13%
New Hybrid Vehicle	2	8%	10%	11%	15%	17%	16%
New Electric Vehicle	3	8%	11%	11%	10%	21%	20%
Sell Current Vehicle	4	0%	0%	0%	1%	1%	2%

Tab. 7.5: SP Game 1 Vehicle Type Choice as Percentage

8. EXPERIMENTS USING DATA COLLECTED

Data collected in Maryland from a multiple-game stated preference experiment are used to estimate a dynamic model of new vehicle adoption over a time horizon of six years. For this modeling exercise we use SP game 1 on vehicle technology, for which a higher number of observations was available. Although 141 respondents completed the survey, only 53 of them could be included in the final sample. Each respondent expressed preferences in 12 time periods spanning over six years. However, given the definition of the scenario tree in the dynamic model, utilities and probabilities can be calculated for the first 10 time periods only. For consistency, the static model is estimated on the same dataset and using the same specification as for the dynamic formulation. In this Chapter, results from both dynamic and static model calibration are presented and the performance of both models in terms of goodness of fit and models' predictions is assessed.

8.1 Static Model Results

Game 1 of the SP survey has a structure that is similar to the one used to generate the simulated data. Four alternatives constitute the choice set: current car (not buy), new gasoline vehicle, hybrid vehicle and electric vehicle. Variables included in the final model specification are: gasoline car price, hybrid car price, electric car price, electric car range, mpg, and current car age. A multinomial logit model was first estimated; Equation 8.1 shows the specification of the model.

$$
\begin{aligned}
&U_{i1t} = gas\_price * \beta_{price\_st} + gas\_mpg\_known * \beta_{mpg\_known} \\
&+gas\_mpg\_unknown * \beta_{mpg\_unknown} + \epsilon_{ij} \\
&U_{i2t} = ASC2 + hyb\_price * \beta_{price\_st} + gas\_mpg\_known * \beta_{mpg\_known} \\
&+hyb\_mpg\_unknown * \beta_{mpg\_unknown} + \epsilon_{ij} \\
&U_{i3t} = ASC3 + ele\_price * \beta_{price\_dy} + ele\_range * \beta_{range} + \epsilon_{ij} \\
&U_{i4t} = ASC4 + veh\_age * \beta_{veh\_age} + \epsilon_{ij}
\end{aligned}
\tag{8.1}
$$

Two coefficients for vehicle price were estimated; β_{price_dy} is specific to the alternative electric car; β_{price_st} is common to both gasoline and hybrid vehicles. Generally, vehicle prices, as cost factors, should only have one corresponding coefficient to estimate, so that the estimated parameter can be used for the elasticity calculation. But here, the electric car price will be chosen as the only dynamic variable in the dynamic model formulation, and the elasticities will not be discussed in the thesis, so two coefficients for vehicle prices (one is for static variables and the other one is for dynamic variable)are required in this special case.

For fuel economy, respondents were split into groups based on their knowledge of the current vehicle fuel economy, measured in MPG. For respondents who knew their vehicle MPG, the difference between the current vehicle MPG and the MPG of the new vehicle was used for estimation. For respondents who did not know their vehicle MPG, the actual new vehicle MPG was used for estimation. All the coefficients estimated have the correct sign and are statistically significant, except for the price of the non-electric vehicles and for the alternative specific constant of electric car. The results are presented in Table 8.1.

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

Tab. 8.1: Static Logit Model Estimation

8.2 Dynamic Model Results

The dynamic model has the same specification of the MNL presented. The price of the electric car is treated as a dynamic variable, and is assumed to vary according to a random drift. Electric car prices generated for the SP scenarios were used to calibrate the auto-regression factor, drift, and variance of random draws under the hypothesis of residuals distributed as normal. After calibration, the dynamic variable function assumes the following form:

$$\begin{aligned}&\text{for electric car price}\\&y_{j,t+1} = -0.103 * y_{jt} + 2.617 + N(0, 1.78)\end{aligned} \tag{8.2}$$

Respondent's perspective dynamic car prices in the scenarios tree are then generated according to equation 8.2.

Unfortunately, the auto-regressive factor is very small; contrary to the one used for the simulated case and relative of the evolution of fuel price, which was estimated on a time-series of real observations. That is due to the fact that scenarios in the survey were designed to be independent. Further research is needed to generate time dependent scenarios in the context of experimental design. The dynamic model estimation results are presented in Table 8.2.

The estimated coefficients are all significant except the static vehicle price. As expected, the fit of the model improves when considering the dynamic nature of the problem; the rho-squared increases from 0.22, the value obtained with the logit model to 0.42 for the dynamic model.

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

Tab. 8.2: Dynamic Model Estimation

8.3 Model Application

Coefficient estimates are used in application to calculate the prediction power of the models. The market share of each alternative observed and predicted together with a measure of errors are reported in Table 8.3.

The error norm D from the dynamic model is smaller than the value obtained by applying the static model. Figure 8.1, 8.2, 8.3, and 8.4 present the observed and predicted market trends of gasoline vehicle, hybrid vehicle, electric vehicle and keeping the current vehicle along the ten time periods in the five years considered. The probability of keeping the current car is relatively high, starting around 70% in the first time period; acceptance of new vehicles starts already in early stages of the time horizon, although volatility is observed at some points. New gasoline vehicles, hybrid and electric vehicles occupy smaller market shares (around 10% each) at the end of the five year periods; all new typologies become more popular after the fifth

Alternative	Observed	Predicted (Static)	Predicted (Dynamic)
Gas car 1	0.151	0.117	0.081
Gas car 2	0.170	0.134	0.120
Gas car 3	0.189	0.128	0.098
Gas car 4	0.208	0.147	0.137
Gas car 5	0.151	0.136	0.118
Gas car 6	0.208	0.142	0.136
Gas car 7	0.113	0.147	0.141
Gas car 8	0.132	0.151	0.102
Gas car 9	0.226	0.168	0.139
Gas car 10	0.189	0.158	0.113
Hybrid car 1	0.094	0.119	0.072
Hybrid car 2	0.094	0.111	0.071
Hybrid car 3	0.151	0.125	0.110
Hybrid car 4	0.094	0.106	0.058
Hybrid car 5	0.113	0.111	0.083
Hybrid car 6	0.170	0.130	0.088
Hybrid car 7	0.245	0.142	0.143
Hybrid car 8	0.151	0.138	0.114
Hybrid car 9	0.170	0.121	0.116
Hybrid car 10	0.226	0.147	0.105
Electric car 1	0.057	0.245	0.062
Electric car 2	0.019	0.245	0.038
Electric car 3	0.094	0.262	0.098
Electric car 4	0.132	0.262	0.076
Electric car 5	0.132	0.264	0.056
Electric car 6	0.132	0.245	0.108
Electric car 7	0.113	0.272	0.092
Electric car 8	0.113	0.270	0.075
Electric car 9	0.189	0.306	0.107
Electric car 10	0.151	0.279	0.111
Current car 1	0.698	0.519	0.786
Current car 2	0.717	0.509	0.770
Current car 3	0.566	0.483	0.694
Current car 4	0.566	0.485	0.730
Current car 5	0.604	0.489	0.742
Current car 6	0.491	0.483	0.668
Current car 7	0.528	0.442	0.624
Current car 8	0.604	0.442	0.709
Current car 9	0.415	0.408	0.638
Current car 10	0.434	0.415	0.671
D		3.24	2.93

Tab. 8.3: Model Validation: Market Shares

time period. Static models have a tendency to average choice probabilities over time and are incapable of recovering peaks in the demand function. More specifically, multinomial logit underestimates the market share of the "not buy" alternative, and dramatically overestimates the share occupied by electric vehicles in the next five years; it predicts quite well the market for new gasoline vehicles and for hybrid vehicles. Dynamic model formulation overestimates the number of respondents keeping their current vehicles, but it is capable to reproduce the descending trend for this alternative. DDCM does an excellent job in recovering market trend for the electric vehicles, that starts at 5% and terminates at around 10% in 2015. For the remaining alternatives the DDCM captures the general behavior, but shows gaps in model prediction higher than those delivered by the static model.

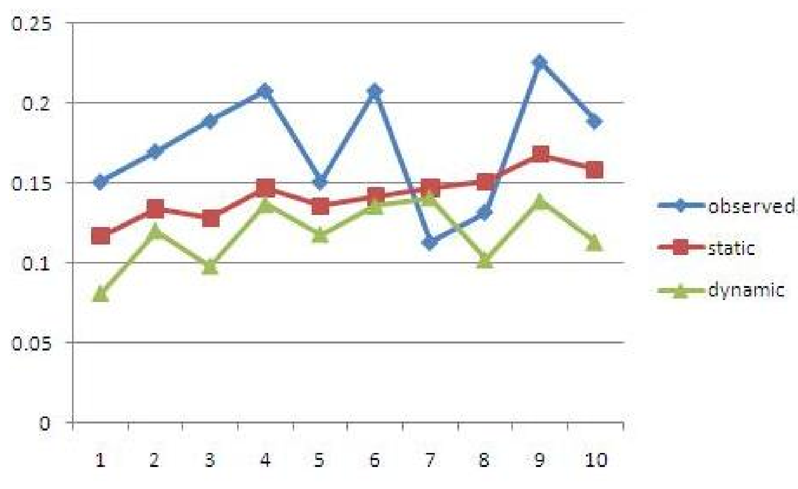

Fig. 8.1: Market Trend for Gasoline Car

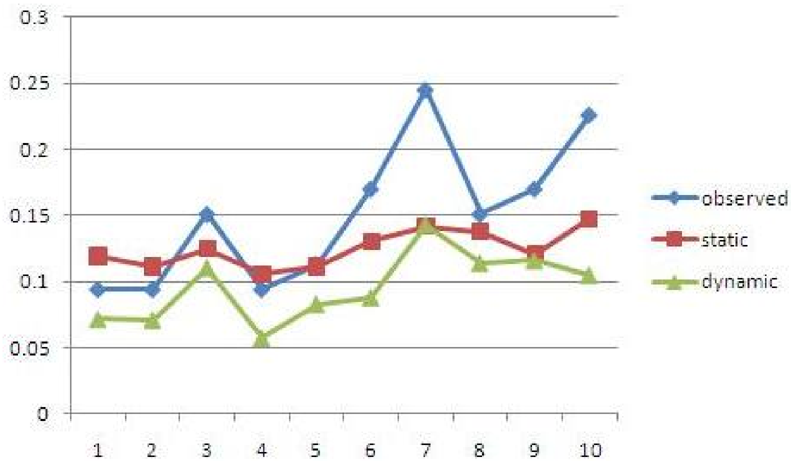

Fig. 8.2: Market Trend for Hybrid Car

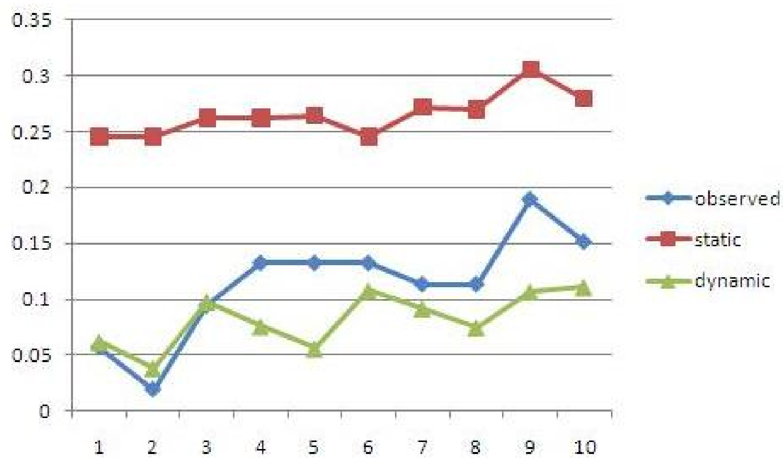

Fig. 8.3: Market Trend for Electric Car

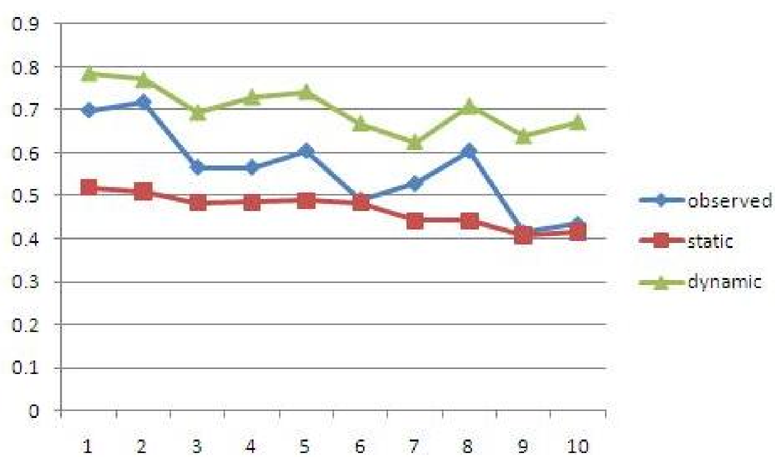

Fig. 8.4: Market Trend for Current Car

8.4 Conclusion

This chapter presented the results obtained from the application of both static and dynamic models to data derived from a SP survey on vehicle preferences, with focus on new technology vehicles. Overall, DDCM is superior to MNL in model fit and in prediction. However, better results can be obtained if more observations would be available and if the data are collected in accordance to the dynamic principles that are expected to characterize the dynamics in the automobile industry products and in consumer behavior. In addition, this first attempt to estimate dynamic models just accounts for two stages in time when calculating individual expectations. Future research should allow for more flexibility when working with the number of scenarios defined in the tree. Finally, estimation and validation on real data collected over past consequent periods of time is desirable; in this case however, it would be difficult to estimate market penetration of new technology vehicles.

9. CONCLUSIONS

Increasingly, energy efficient and environmentally friendly highway transportation technologies are under development and could be available on the market in the near future; they could ensure the freedom of mobility and energy security, while lowering costs and reducing the impact on the environment. Estimation techniques for analyzing the impact of technological improvements and rapid changes in energy costs are necessary to understand the mobility of tomorrow and adapt the products of our car industry. This dissertation has developed a dynamic econometric model that accounts for the evolving characteristics of the products offered by the automobile industry and consumers' expectations of future vehicle quality. The timing of consumers' purchases is formalized as an optimal stopping problem where the agent (consumer) must decide on the optimal time of purchase. The modeling framework is further enriched by explicitly considering the consumer's choice from a set of different types of vehicles whose quality changes stochastically over time. The proposed approach extends the theory of discrete choice models on a temporal basis and improves existing dynamic discrete choice models based on a pure dynamic programming perspective. The modeling framework has been applied to both simulated and real data. In both cases, results show that dynamic models are superior to static models based on MNL; in particular, they are able to recover peaks in the demand evolution over time, while static models fail to detect dramatic changes due to the rapid mutation of external conditions.

The remaining of this chapter summarizes the findings from this research work, identifies the main contributions and proposes future research directions.

9.1 Contributions

1. Static discrete choice models assume that individual utilities are linear functions of the alternatives and individuals characteristics; attributes are not time-dependent and can be defined for the actual situation (RP data) or future scenarios (SP data). Dynamic models estimated in this thesis are non-linear in utilities; utilities include information on both current alternatives and individual expectations about future alternatives.

2. The model proposed is more complex than traditional dynamic discrete choice models used in economics. Usually stopping problems are characterized by just two state options, assume homogeneous population and choice sets are composed of just one product. More importantly, the consumer is considered out of the market when his/her status changes. In the problem modeled here decision makers have more than one starting condition; each household can actually not own any vehicle, own one or multiple vehicles. Moreover, the population is heterogeneous, and every time a household decides to change his/her status, there are multiple alternatives available, each characterized by different vehicle technologies. A regenerative process allows for the consideration

of multiple purchases.

3. The optimization problem for parameters' estimation is solved using Maximum-Likelihood estimation method. Dynamic discrete choice models based on dynamic programming are usually estimated by using the nested fixed point optimization algorithm, which is valid under the restrictive assumption that the time horizon is infinite.

4. The dynamic discrete decision process is solved by generating a scenario tree from the underlying stochastic process. This approach is generally adopted in stochastic optimization programs that consist of a stochastic model and an optimization model. In my formulation the stochastic problem is related to the nature of the attributes that change stochastically over scenarios and to the uncertainty in the individuals' future expectations; the optimization problem concerns individuals' utility maximization according to the random utility maximization paradigm. The optimization problem is expressed and solved in a recursive manner. Such model formulation and presentation, as well as the associated visual structure has suggested a computational method of solution. This provides a graphically intuitive model construction and evaluation capabilities for transportation modelers who may be less familiar with stochastic modeling and algorithms.

5. A pilot survey is designed and executed in order to estimate dynamic models in a real context and to test their performance with respect to traditional static models. Hypothetical scenarios are separated by six month intervals

and span a six year period. The designs correspond to changing vehicle technology, fuel type, and taxation policy. Between scenarios, the vehicle and fuel attributes dynamically change to mimic marketplace conditions. Empirical analysis shows that respondents are able to create trade-offs between different vehicle technology as well as the price of various fueling options. In addition about 65% respondents intend to buy a vehicle in the next six years; this result definitely shows that a potential market exists for new and more efficient gasoline cars and for electric and hybrid vehicles.

9.2 Future Work

This dissertation does the pilot research on dynamic discrete choice models in transportation and has proposed an application to car ownership modeling and forecasting. It is expected that this work will generate innovations in demand modeling and that it will be extended to other problems which are dynamic in nature. The following points indicate possible avenues for future research:

The model formulation allows for just one dynamic attribute in the utility specification; the random walk is actually estimated on a univariate time series. The analysis should be extended to multivariate random walks.

The number of scenarios considered for the calculation of the expected utility is limited to two. This is a rather restrictive assumption; it would be desirable to extend the time horizon over which the respondents consider future information about new alternatives.

Data collection techniques should be improved in order to capture interdependency amongst successive observations in time. Methods to incorporate random walks into orthogonal experimental design (for SP data) should be developed.

The dynamic model should be estimated on a Revealed Preference panel dataset. In France, INRETS organized and maintained a panel dataset that follows the evolution of car ownership and car use from 1984 to 2000.

From the optimization perspective, it would be interesting to compare the results obtained from maximum likelihood estimation with those obtained from the nested fixed point procedure and to demonstrate if the underlying hypotheses are valid for our case, which is developed for a finite horizon problem.

This dynamic framework can be adapted and transferred to other case studies: dynamic pricing for revenue management, route choice behavior under dynamic tolling, activity scheduling for activity based analysis just to cite a few.

APPENDIX

A. SIMULATED INPUT DATA FILE FORMAT

resp_no	hh_no	income
1	3	3
2	6	3
3	4	2
4	4	3
5	6	2
6	2	2
7	5	2
8	5	2
9	1	2
10	4	4
11	5	4
12	2	3
13	3	4
14	1	2
15	3	3
16	5	1
17	4	2
18	4	2
19	4	1
20	4	2

Fig. A.1: Household Characteristics

resp_no	current_age1	current_age2	current_age3	current_age4	current_age5	current_age6	current_age7	current_age8	cur
1	0.6	1.1	1.6	2.1	0.5	1	1.5	0.5	
2	0.8	1.3	1.8	0.5	1	1.5	0.5	0.5	
3	0.8	1.3	1.8	2.3	0.5	1	1.5	2	
4	0.5	1	0.5	1	1.5	0.5	0.5	1	
5	0.5	1	1.5	0.5	1	1.5	2	0.5	
6	0.4	0.9	1.4	1.9	2.4	0.5	1	1.5	
7	0.9	0.5	1	1.5	0.5	1	0.5	1	
8	1	1.5	0.5	1	1.5	2	0.5	1	
9	0.4	0.9	1.4	1.9	0.5	1	0.5	1	
10	0.2	0.7	1.2	1.7	2.2	2.7	0.5	1	
11	0.9	1.4	0.5	1	0.5	1	0.5	0.5	
12	0.1	0.6	1.1	1.6	0.5	1	1.5	0.5	
13	0.2	0.7	1.2	1.7	2.2	0.5	1	1.5	
14	1	1.5	0.5	1	1.5	2	0.5	1	
15	0.3	0.8	1.3	1.8	2.3	0.5	1	1.5	
16	0.5	1	1.5	0.5	1	1.5	0.5	1	
17	0.3	0.8	1.3	0.5	1	1.5	2	0.5	
18	0.6	1.1	1.6	2.1	0.5	1	1.5	2	
19	0.6	1.1	1.6	0.5	1	0.5	1	1.5	
20	0.2	0.7	1.2	1.7	0.5	1	1.5	2	
21	0.9	1.4	1.9	2.4	2.9	0.5	1	1.5	
22	0.1	0.6	1.1	1.6	2.1	2.6	3.1	0.5	
23	0.5	1	1.5	2	0.5	1	1.5	2	
24	0.7	1.2	1.7	2.2	2.7	0.5	1	1.5	
25	0.1	0.6	1.1	1.6	0.5	1	0.5	1	
26	0	0.5	1	1.5	2	0.5	1	1.5	
27	1	0.5	1	1.5	2	0.5	1	1.5	
28	0.8	1.3	1.8	0.5	1	1.5	2	2.5	
29	0	0.5	1	1.5	0.5	1	0.5	1	
30	0.7	1.2	1.7	2.2	0.5	1	1.5	2	
31	0.5	1	1.5	2	0.5	1	1.5	2	
32	0	0.5	1	1.5	2	0.5	1	1.5	
33	0.5	0.5	1	1.5	0.5	1	0.5	1	
34	0.2	0.7	1.2	1.7	2.2	2.7	0.5	1	
35	0.5	1	1.5	2	0.5	1	1.5	2	
36	1	1.5	0.5	1	1.5	0.5	1	0.5	
37	0.7	1.2	1.7	2.2	2.7	0.5	1	1.5	

Fig. A.2: Current Vehicle Attributes

resp_no	veh size11	veh price11	veh size21	veh price21	veh size31	veh price31	veh size12	veh price12	veh size22	veh price22
1	1	24	2	66	2	53	3	65	2	53
2	1	21	2	56	1	22	2	52	3	60
3	3	68	2	63	3	56	2	57	2	66
4	3	69	2	65	1	25	1	24	2	54
5	2	53	1	24	2	50	2	66	2	68
6	2	63	3	61	1	24	2	62	1	24
7	2	64	3	64	3	64	2	68	2	64
8	2	61	3	51	2	69	3	67	3	63
9	2	70	3	53	1	20	2	61	2	55
10	2	55	2	58	2	58	1	21	1	23
11	1	17	1	21	2	51	2	66	2	60
12	3	57	1	17	1	21	2	70	3	50
13	3	59	2	63	1	23	3	50	1	19
14	1	15	2	67	2	63	3	66	1	19
15	2	62	1	21	2	56	2	51	2	54
16	2	52	2	57	1	24	2	62	1	22
17	3	67	3	61	2	58	2	57	2	64
18	2	57	1	15	3	57	2	57	1	16
19	2	53	2	54	2	62	2	56	2	69
20	2	53	1	21	2	55	1	22	3	55
21	1	17	1	18	1	25	3	52	2	59
22	3	69	2	57	2	55	3	57	1	19

Fig. A.3: Potential Vehicle Attributes

resp_no	gap1	gap2	gap3	gap4	gap5	gap6	gap7
1	3.5	3.277349	3.060108	2.848145	2.641334	2.439548	2.242666
2	3.5	3.420437	3.342807	3.267063	3.19316	3.121053	3.050698
3	3.5	3.536285	3.571688	3.60623	3.639934	3.672818	3.704903
4	3.5	3.679306	3.854255	4.024952	4.191502	4.354004	4.512558
5	3.5	3.483449	3.467301	3.451545	3.436171	3.421172	3.406536
6	3.5	3.323211	3.150718	2.982417	2.818205	2.657984	2.501656
7	3.5	3.48229	3.46501	3.44815	3.431699	3.415649	3.399988
8	3.5	3.668365	3.83264	3.992922	4.14931	4.301897	4.450776
9	3.5	3.724576	3.943695	4.157489	4.366088	4.569618	4.768203
10	3.5	3.733025	3.960387	4.182225	4.398672	4.609859	4.815914
11	3.5	3.599805	3.697185	3.792199	3.884903	3.975355	4.063609
12	3.5	3.54886	3.596533	3.643047	3.688431	3.732713	3.775918
13	3.5	3.525796	3.550966	3.575523	3.599484	3.622863	3.645674
14	3.5	3.471225	3.443149	3.415755	3.389026	3.362948	3.337503
15	3.5	3.65025	3.796849	3.939886	4.079447	4.215617	4.348478
16	3.5	3.477431	3.45541	3.433924	3.41296	3.392506	3.372549
17	3.5	3.478838	3.458189	3.438043	3.418386	3.399207	3.380494
18	3.5	3.877197	4.245229	4.604317	4.95468	5.296528	5.63007
19	3.5	3.620193	3.737465	3.851888	3.96353	4.072459	4.178742
20	3.5	3.586322	3.670546	3.752723	3.832904	3.911136	3.987467
21	3.5	3.347048	3.197813	3.052204	2.910134	2.771515	2.636266
22	3.5	3.306865	3.118423	2.93456	2.755166	2.58013	2.409348
23	3.5	3.41782	3.337638	3.259404	3.18307	3.108592	3.035924

Fig. A.4: Dynamic Attributes

resp_no	CHOSEN11	CHOSEN21	CHOSEN31	CHOSEN41	CHOSEN12	CHOSEN22	CHOSEN32	CHOSEN42
1	0	0	0	1	0	1	0	0
2	0	1	0	0	0	0	0	1
3	0	0	0	1	0	0	0	1
4	0	0	0	1	0	0	0	1
5	0	0	0	1	0	0	0	1
6	0	0	0	1	0	0	0	1
7	0	0	0	1	0	0	0	1
8	0	0	0	1	0	0	0	1
9	0	0	1	0	0	0	0	1
10	0	0	0	1	0	0	0	1
11	0	0	0	1	0	0	0	1
12	0	0	1	0	1	0	0	0
13	0	0	0	1	1	0	0	0
14	0	0	0	1	0	0	1	0
15	0	1	0	0	0	0	0	1
16	0	0	1	0	0	0	0	1
17	0	1	0	0	0	0	0	1
18	0	0	0	1	0	0	1	0
19	0	0	0	1	0	0	0	1
20	0	0	0	1	0	0	1	0
21	0	0	0	1	0	1	0	0
22	0	0	1	0	0	0	0	1
23	0	0	1	0	0	1	0	0

Fig. A.5: Choice

B. LIST OF POSSIBLE QUESTIONS FOR THE SURVEY

Q Please indicate your gender.

- Male
- Female

Q What is your age?
Q What is your level of education?

- Less than high school
- High school graduate
- Some college
- Associate degree
- Bachelor's degree
- Graduate or professional degree

Q During most of last week, were you...

- Working full time (35 hours per week or more)
- Working part time (less than 35 hours per week)
- Temporarily absent from a job or business
- Looking for work
- A homemaker
- Going to school
- Retired
- Other

Q How far (in miles) is your commute to work or school?
Q Do you have a driver license?
Q Are you the head of the household?
Q What is your household income?

- Less than $24,999
- $25,000 to $49,999
- $50,000 to $74,999
- $75,000 to $99,999
- $100,000 to $149,999
- $150,000 or more

Q How many children in age 12 or under live in your household?

- 0
- 1
- 2
- 3
- 4
- 5
- More than 5

Q How many people in age 13 through 17 (13, 14, 15, 16, or 17) live in your household?

- 0
- 1
- 2
- 3
- 4
- 5
- More than 5

Q How many adults (including yourself) in age 18 or over live in your household?

- 0
- 1
- 2

- 3
- 4
- 5
- More than 5

Q How many people in your household work?

- 0
- 1
- 2
- 3
- 4
- More than 4

Q What state do you currently live in?

- Maryland
- Virginia
- District of Columbia
- Other

Q What is the zip code of your living place?
Q How many car does your household have?

- No Cars in the Household
- One Car in the Household
- Two Cars in the Household
- Three Cars in the Household
- Four Cars in the Household
- More than 4 Cars in the Household

Q Which of the following best describes your home?

- College Dorm or similar student-based housing
- Apartment
- Condominium

- Townhouse
- Rowhouse
- Single-Family Home, Detached House, or Separated House
- Other

Q Where do you typically park your vehicle when at home?

- Personal Garage
- Personal Driveway
- On-street
- Outdoor Parking Lot
- Parking Garage or Covered Parking Lot
- Other

Q Is parking available at your job or school?
Q How much does it cost you to park at work per month? If free, type in 0.
Q What is the make and model of your primary vehicle?
Q Which of the following types best describes your primary vehicle?

- Compact / Small Car
- Mid-size Car
- Large Car
- Luxury Car
- Sports Car
- Minivan / Van
- Pickup Truck
- Sports Utility Vehicle (SUV)

Q How old (in years) is your primary vehicle?
Q When did you purchase your primary vehicle?
Q On average, approximately how many miles does this vehicle travel per year?
Q What type of fuel does your primary vehicle use?

- Gasoline
- Hybrid

- Diesel
- Electric
- Alternative Fuel (for example: ethanol, natural gas, biodiesel, propane, hydrogen)
- Other

Q Did you buy this vehicle new? (The car has not been owned by anyone else)
Q Approximately, how much did this vehicle cost?
Q What is your primary vehicle's fuel efficiency (MPG)?
Q How many gallons of fuel can your vehicle hold?
Q What is the seating capacity of your vehicle?

- 2
- 3
- 4
- 5
- 6
- 7
- 8
- More than 8

Q How far (in miles) do you travel off of your normal route to find refuel your vehicle? If your refueling station is usually on your route, then input 0.

Q Approximately, how much do you pay in tolls on an average day?
Q Approximately, how much did you your car cost?
Q How much were your tax deductions (if any) on your car?

C. SAMPLE SCENARIO DESIGNS

Price and Emissions Survey

Start Over (Testing)

A survey about customer preference for alternative fuel vehicles in relation to price and emissions

Question 29.

The following vehicles characteristics are available for vehicles in 2010:

	Current Vehicle	Gasoline Vehicle	Hybrid Vehicle	Electric Vehicle
Vehicle Price	**$4500**	**$30300**	**$46100**	**$21800**
Fuel Economy (Miles per Gallon)	**40** mpg	**21** mpg	**21** mpg	**No fuel needed** Runs on electric power
Range Between Refueling	**400** miles	**450** miles	**400** miles	**100** miles
Vehicle Emissions	**20% less than** average vehicle	**30% greater than** average 2010 vehicle	**25% greater than** average 2010 vehicle	**No Direct Emissions**
Vehicle Size	**Compact / Small Car**	**SUV**	**Pickup Truck**	**Small Car (5-Seats)**

Which option would you prefer for your vehicle ownership in **2010**?

- **I Will KEEP My Current Vehicle**
- **I Will BUY the Gasoline Vehicle And KEEP My Current Vehicle**
- **I Will BUY the Hybrid Vehicle And KEEP My Current Vehicle**
- **I Will BUY the Electric Vehicle And KEEP My Current Vehicle**
- **I Will BUY the Gasoline Vehicle And SELL My Current Vehicle**
- **I Will BUY the Hybrid Vehicle And SELL My Current Vehicle**
- **I Will BUY the Electric Vehicle And SELL My Current Vehicle**
- **I Will SELL My Current Vehicle and NOT REPLACE It**

Fuel Tax Survey

Start Over (Testing)

A survey about customer preference for alternative fuel vehicles in relation to taxation and availability of fuel

Question 22.

The following fuel characteristics are available in 2010:

	Gasoline Fuel	Alternative Fuel	Diesel Fuel	Electricity
Fuel Price, Pre Tax	$3.00	$3.15	$4.00	$5.70
Fuel Tax	$0.65	$0.15	$1.05	$0.28
Fuel Efficiency	23	19	34	100
Fueling Station Availability	Within 5 miles	Within 15 miles	Within 5 miles	4-hr Home Charge Only

Which option would you prefer for your vehicle ownership in **2010**?

- I Will KEEP My Current Vehicle
- I Will BUY a Gasoline Vehicle that runs on Gasoline
- I Will BUY an Alternative Fuel Vehicle that runs on Alternative Fuel
- I Will BUY a Diesel Vehicle that runs on Diesel Fuel
- I Will BUY a Electric Vehicle
- I Will BUY a Plug-In Hybrid Electric Vehicle
- I Will SELL My Current Vehicle and NOT REPLACE It

Taxes and Fees Survey

Start Over (Testing)

A survey about customer preference for alternative fuel vehicles in relation to taxes and fees

Question 5.

The following vehicle taxes and fees are available in **2012**:

	Current Vehicle	New Gasoline Vehicle	New Hybrid Vehicle	New Electric Vehicle
Registration Fee (per year)	$77	$84	$70	$84
Income Tax Deduction	$0	$0	$1000	$2500
Toll Cost	Normal Price	Normal Price	25% less than Normal Price	25% less than Normal Price
Miles Traveled Fee	**$90** per 1,000 miles traveled	**$30** per 1,000 miles traveled	**$60** per 1,000 miles traveled	**$60** per 1,000 miles traveled

Which option would you prefer for your vehicle ownership in **2012**?

- I Will KEEP My Current Vehicle
- I Will KEEP My Current Vehicle And Drive Less
- I Will BUY The Gasoline Vehicle And KEEP My Current Vehicle
- I Will BUY The Hybrid Vehicle And KEEP My Current Vehicle
- I Will BUY The Electric Vehicle And KEEP My Current Vehicle
- I Will BUY The Gasoline Vehicle And SELL My Current Vehicle
- I Will BUY The Hybrid Vehicle And SELL My Current Vehicle
- I Will BUY The Electric Vehicle And SELL My Current Vehicle
- I Will SELL My Current Vehicle and NOT REPLACE It

D. DISTRIBUTION OF HOUSEHOLDS

E. MAJOR C CODE FOR THE FORMULATION AND ESTIMATION

• data.h

```
#ifndef __DATA_H__
#define __DATA_H__
//#include "memory.h"
#ifdef __cplusplus
extern "C" {
#endif
#include <oratio/oratio.h>
typedef struct {
// individual ID
int id;
// socioeconomic variables for each individual
double *indiv;
} ind;
typedef struct {
// variable for current product in each time, for each person
double **current;
} cur;
// information for each potential product
typedef struct {
// decision variable, 0 or 1 at each time period to each product type for
each person
double **decision;
// static attributes for potential choice j in time t
double **stati;
} poten;
// Define the type of data used for the samplings.
typedef struct {
// number of individuals
int indivNum;
// number of time period
int time;
// number of individual variables
int numINDIVAR;
// number of current car variables
int numCURTVAR;
// number of static vars for potential choice
int numSTATIC;
// number of dynamic vars for potential choice
int numDYNAMIC;
// number of choice
int numch;
} glo;
```

```
//put all data files together
typedef struct {
ind* in;
cur* curr;
poten* pot;
glo* glonum;
double*** prob_matrix; // Probability matrix
Random *rand_seed; // Random seed
double** draw; //drawn from Random normal distribution
double ***err1; //error term for i,j,t
double *err2; //error term for i
double **p; //the random value (0 <= p <= 1) for calculating normal dynamic
variable y.
double vp; //the random value (0 <= p <= 1) for calculating MC v
double ***perror;//the random value (0 <= p <= 1) for calculating gumbel
error_ijt for different i,j,t
double *perrori;//the random value (0 <= p <= 1) for calculating gumbel
error_ijt for different i
double *y_int;//current gas prices(the varible will be seen as dynamic )
} Aldata;
//get number of parameters (dimension of the problem)
int get_dimension(Aldata* d);
//read data from four data structures and allocate memory
Aldata* format_data();
//generating gas prices for scenario tree
double** draw_random_y(Aldata* d, double** draw);
//mode in scenario tree
double*** calculate_mode(Aldata* d, double* x, double** y);
//mode that is correlated to current situation in each time
double** calculate_mode_real(Aldata* d, double* x);
//calculate log sum utility of three car alternatives
double*** calculate_v(Aldata* d, double* x, double** y, double vp);
//recursive process for scenario tree to calculate expectation utility E
double cal_E(Aldata* d, int t, int T, double *v, double current, int n,
double* x, int indiv);
//probability of buying cetern kind of car
double*** cal_probcar (Aldata* d, double* x);
//probability of not buying, PI0
double** cal_prob (Aldata* d, double* x, double** y);
// functions of reading four .txt data files
void read_new_indiv(glo* paraNB, ind* paraIN);
void read_new_current(glo* paraNB, cur* paraC, ind* paraIN);
void read_new_poten(glo* paraNB, ind* paraIN, poten* paraP);
void read_new_choice(glo* paraNB, ind* paraIN, poten* paraP);
// function of reading coefficient data files
void read_new_para(double *x, glo* paraNB);
```

```
// functions of allocating memories to four structures and their elements
glo* getGlo();
ind* getIn(glo *glonum);
cur* getC(glo *glonum);
poten* getP(glo *glonum);
//memory allocation functions for probability matrix
double*** c_malloc_P(glo* paraNB);
//function of allocating memory to error terms
double*** c_malloc_e(glo* paraNB);
// function of allocating memory to the array of utility U[i][j][t]
double*** c_malloc_u(glo* paraNB);
double**** c_malloc_uy(glo* paraNB);
// function of allocating memory to the array of summumation of exp(U[i][j])
for each person i
double* c_malloc_w(glo* paraNB);
//function of allocating memory to v_itj (the dimension of j is for 1000
draws)
double*** c_malloc_v(glo* paraNB);
// free functions
void free_ind (ind* in, glo* paraNB);
void free_cur (cur* c, glo* paraNB);
void free_poten (poten* p, glo* paraNB);
void free_uy(double**** u, glo* paraNB);
void free_glo (glo* g);
void free_w(double* W);
void free_v(double*** v, glo* paraNB);
void free_u(double*** u, glo* paraNB);
void free_p(double*** u, glo* paraNB);
void free_err (Aldata *d);
// Function to help in calculating t-stats
double amlet_t_statistics(int n, double *theta, double *hypothetical, double
**I, double alpha, double *t);
//function to inverse matrix
void op_matrix_inverse(const enum CBLAS_ORDER Order, const enum CBLAS_UPLO
Uplo, double *I, int npar);
//function to print out matrix
void nt_matrix_print(FILE *out, char *name, double **A, int m, int n);
#ifdef __cplusplus
}
#endif
#endif
```

- data.c

```
#include "data.h"
#include <stdlib.h>
#include <string.h>
#include <stdio.h>
#include <oratio/oratio.h>
//read data from four data structures and allocate memory
Aldata* format_data(){
Aldata *d;
int i,j,t,n;
d= malloc(sizeof(Aldata));
d->glonum = getGlo();
d->in = getIn(d -> glonum) ;
d->pot = getP(d -> glonum);
d->curr = getC(d -> glonum);
read_new_indiv(d->glonum, d->in);
read_new_current(d->glonum, d->curr, d->in);
read_new_poten(d->glonum, d->in, d->pot);
read_new_choice(d->glonum, d->in, d->pot);
d->prob_matrix = c_malloc_P(d -> glonum);
d->rand_seed = ran_random();
d->y_int=malloc(d->glonum->time*sizeof(double));
/*
Place the current dynamic variables into the array.
They are calculated by normal diffusion process with
the same factors as function draw_random_y, initial
y is 3.5
*/
d->y_int[0] = 3.50;
d->y_int[1] = 3.27;
d->y_int[2] = 3.20;
d->y_int[3] = 3.25;
d->y_int[4] = 3.43;
d->y_int[5] = 3.42;
d->y_int[6] = 3.24;
d->y_int[7] = 3.23;
d->y_int[8] = 3.41;
d->y_int[9] = 3.63;
d->y_int[10] = 3.86;
d->y_int[11] = 3.95;
d->p=nt_matrix_new(d->glonum->time+1, 20);
for (t=0; t<d->glonum->time; t++){
for (n=0; n<20; n++){
d->p[t][n] = ran_random_get_val(d->rand_seed );
}
}
d->vp = ran_random_get_val(d->rand_seed );
```

```
d->perror=c_malloc_e(d -> glonum);
for(i = 0; i < d->glonum ->indivNum; i++) {
for(t = 0; t < d->glonum->time+2; t++) {
for(j = 0; j < d->glonum ->numch+1; j++) {
d->perror[i][j][t] = ran_random_get_val(d->rand_seed );
}
}
}
d->perrori=malloc(d->glonum ->indivNum*sizeof(double));
for(i = 0; i < d->glonum ->indivNum; i++) {
d->perrori[i] = ran_random_get_val(d->rand_seed );
}
d->draw=nt_matrix_new(d->glonum->time, 20);
for (t=0; t<d->glonum->time; t++){
for (n=0; n<20; n++){
d->draw[t][n] = st_normal_icdf(d->p[t][n], 0, 16);
}
}
d->err1=c_malloc_e(d->glonum);
d->err2=malloc(d->glonum ->indivNum*sizeof(double));
for(i = 0; i < d->glonum ->indivNum; i++) {
for(t = 0; t < d->glonum->time+2; t++) {
d->err2[i]=st_gumbel_icdf(d->perrori[i], 0, 1 );
for(j = 0; j < d->glonum ->numch+1; j++) {
d->err1[i][j][t]=st_gumbel_icdf(d->perror[i][j][t], 0, 1 );
}
}
}
free_p(d->perror, d->glonum);
free(d->perrori);
return d;
}
//check number of parameters (dimension of the problem)
int get_dimension(Aldata* d){
int l0= d->glonum->numch-1;//asc
int l1= d->glonum->numINDIVAR;//indiv var
int l2= d->glonum->numSTATIC;//static var
int l3= d->glonum->numDYNAMIC;//dynamic var
int l4= d->glonum->numCURTVAR;//current var
return l0+l1+l2+l3+l4;
}
/*
**
* get a glo variable
*/
glo* getGlo() {
glo *glonum;
```

```
glonum = (glo *)malloc(sizeof(glo));
glonum->indivNum = 200;
glonum->time=12;
glonum->numINDIVAR=2;
glonum->numCURTVAR=1;
glonum->numSTATIC=2;
glonum->numDYNAMIC=1;
glonum->numch=3;
return glonum;
}
/**
* get a ind variable
*/
ind* getIn(glo *glonum) {
ind* individual = malloc((glonum->indivNum) * sizeof(ind));
return individual;
}
/**
* get a cur variable
*/
cur* getC(glo *glonum) {
int i;
cur *curr;
int ttnumCUR=(glonum->time+2) * glonum->numCURTVAR;
curr = (cur *)malloc(sizeof(cur));
curr->current = malloc( glonum->indivNum * sizeof(double *));
for(i=0;i<glonum->indivNum;i++) {
curr->current[i]=malloc(ttnumCUR * sizeof(double));
}
return curr;
}
/**
* get a Potential variable
*/
poten* getP(glo *glonum) {
int i;
poten *pot;
int ttnumSTA = (glonum->time+2) * glonum->numch * glonum->numSTATIC;
int ttnumDEC = glonum->time *( glonum->numch+1);
pot = (poten *)malloc(sizeof(poten));
pot->stati = malloc(glonum->indivNum * sizeof(double *));
for(i = 0;i < glonum->indivNum;i++) {
pot->stati[i]=malloc(ttnumSTA * sizeof(double));
}
```

```
pot->decision = malloc(glonum->indivNum * sizeof(double));
for(i = 0;i < glonum->indivNum;i++) {
pot->decision[i]=malloc(ttnumDEC * sizeof(double));
}
return pot;
}
void read_new_indiv(glo* paraNB, ind* paraIN) {
FILE *inn;
FILE *out;
int i=0, j=0;
// read the indivX.txt file
inn=fopen("indivX.txt","r");
out=fopen("oput1.txt","w");
if (inn == NULL) {
printf ("File could not be opened\n");
exit(-1);
}
for(i = 0; i < paraNB ->indivNum; i++) {
paraIN[i].indiv = malloc( (paraNB ->numINDIVAR) * sizeof(double));
// read the individual's index number
fscanf(inn, "%d", &paraIN[i].id);
for (j=0;j<(paraNB->numINDIVAR);j++) {
// read the x variables for this individual
fscanf(inn, "%lg", &paraIN[i].indiv[j]);
// printf("%d\n", in[i].x[j]);
}
//fscanf(inn,"\n");
}
for(i=0;i<paraNB->indivNum;i++) {
// print out the person's index number
fprintf(out,"%d", paraIN[i].id);
for (j=0;j<(paraNB->numINDIVAR);j++) {
// print out the x variables for each individual
fprintf(out, "%lg", paraIN[i].indiv[j]);
}
fprintf(out,"\n");
}
fclose(inn);
fclose(out);
}
void read_new_current(glo* paraNB, cur* paraC, ind* paraIN) {
FILE *inn;
FILE *out;
int i=0, j=0;
```

```
int ttnumCURT = paraNB->time * paraNB->numCURTVAR;
inn=fopen("current.txt","r");
out=fopen("oput2.txt","w");
if (inn == NULL) {
printf ("File could not be opened\n");
exit(-1);
}
for(i=0;i<paraNB->indivNum;i++) {
fscanf(inn, "%d", &paraIN[i].id);
for (j=0;j<ttnumCURT;j++) {
fscanf(inn, "%lg", &paraC->current[i][j]);
}
}
for(i=0;i<paraNB->indivNum;i++) {
fprintf(out,"%d", paraIN[i].id);
for (j=0;j<ttnumCURT;j++) {
fprintf(out,"%lg",paraC->current[i][j]);
}
fprintf(out,"\n");
}
fclose(inn);
fclose(out);
}
void read_new_poten(glo* paraNB, ind* paraIN, poten* paraP) {
FILE *inn;
FILE *out;
int i=0, j=0;
int ttnumSTAT = (paraNB->time+2) * paraNB->numch * paraNB->numSTATIC;
inn=fopen("poten.txt","r");
out=fopen("oput3.txt","w");
if (inn == NULL) {
printf ("File could not be opened\n");
exit(-1);
}
for(i = 0;i < paraNB->indivNum;i++) {
fscanf(inn, "%d", &paraIN[i].id);
for (j=0;j<ttnumSTAT;j++) {
fscanf(inn, "%lg", &paraP->stati[i][j]);
}
}
```

```
for(i=0; i<paraNB->indivNum; i++) {
fprintf(out,"%d", paraIN[i].id);
for (j=0; j<(ttnumSTAT); j++) {
fprintf(out,"%lg",paraP->stati[i][j]);
}
fprintf(out,"\n");
}
fclose(inn);
fclose(out);
}
void read_new_choice(glo* paraNB, ind* paraIN, poten* paraP) {
FILE *inn;
FILE *out;
int i=0, j=0;
int ttnumDEC = paraNB->time * (paraNB->numch+1);
inn=fopen("choice.txt","r");
out=fopen("oput4.txt","w");
if (inn == NULL) {
printf ("File could not be opened\n");
exit(-1);
}
for(i=0; i< paraNB->indivNum; i++) {
fscanf(inn, "%d", &paraIN[i].id);
for (j = 0;j<(ttnumDEC);j++) {
fscanf(inn, "%lg", &paraP->decision[i][j]);
}
}
for(i=0;i<paraNB->indivNum;i++) {
fprintf(out,"%d", paraIN[i].id);
for (j=0;j<(ttnumDEC);j++) {
fprintf(out,"%lg",paraP->decision[i][j]);
}
fprintf(out,"\n");
}
fclose(inn);
fclose(out);
}
void read_new_para(double *x, glo* paraNB){
FILE *inn;
int i;
int j = 0;
float f;
```

```
inn=fopen("parady.txt","r");
if (inn == NULL) {
printf ("File could not be opened\n");
exit(-1);
}
// Reads the ASCs
for(i=0; i< paraNB->numch-1; i++) {
fscanf(inn, "%f", &f);
x[j++] = f;
}
// Reads the individual specific parameters
for(i=0; i< paraNB->numINDIVAR; i++) {
fscanf(inn, "%f", &f);
x[j++] = f;
}
fscanf(inn, "\n");
// Reads the static parameters
for(i=0; i< paraNB->numSTATIC; i++) {
fscanf(inn, "%f", &f);
x[j++] = f;
}
// Reads the dynamic parameters
for(i=0; i< paraNB->numDYNAMIC; i++) {
fscanf(inn, "%f", &f);
x[j++] = f;
}
// Reads the current product parameters
for(i=0; i< paraNB->numCURTVAR; i++) {
fscanf(inn, "%f", &f);
x[j++] = f;
}
}
double*** c_malloc_P(glo* paraNB){
double ***P;
int i,j;
P = (double***)malloc(paraNB ->indivNum*sizeof(double**));
for(i = 0; i < paraNB->indivNum; i++) {
P[i] = (double**)malloc(4 *sizeof(double*));
for(j=0;j<4;j++){
P[i][j]=(double*)malloc(paraNB->time*sizeof(double));
}
}
return P;
}
double*** c_malloc_e(glo* paraNB){
double ***e;
int i,j;
```

```
e = (double***)malloc(paraNB ->indivNum*sizeof(double**));
for(i = 0; i < paraNB->indivNum; i++) {
e[i] = (double**)malloc(4 *sizeof(double*));
for(j=0;j<4;j++){
e[i][j]=(double*)malloc((paraNB->time+2)*sizeof(double));
}
}
return e;
}
double*** c_malloc_u(glo* paraNB){
double ***U;
int i,j;
U = (double***)malloc(paraNB ->indivNum*sizeof(double**));
for(i = 0; i < paraNB->indivNum; i++) {
U[i] = (double**)malloc(paraNB->numch *sizeof(double*));
for(j=0;j<paraNB->numch;j++){
U[i][j]=(double*)malloc(paraNB->time*sizeof(double));
}
}
return U;
}
double**** c_malloc_uy(glo* paraNB){
double ****U;
int i,j,t;
U = (double****)malloc(paraNB ->indivNum*sizeof(double***));
for(i = 0; i < paraNB->indivNum; i++) {
U[i] = (double***)malloc(paraNB->numch *sizeof(double**));
for(j=0;j<paraNB->numch;j++){
U[i][j]=(double**)malloc((paraNB->time+1)*sizeof(double*));
for(t=0; t<paraNB->time+1; t++){
U[i][j][t]=(double*)malloc(20*sizeof(double));
}
}
}
return U;
}
double*** c_malloc_v(glo* paraNB){
double ***v;
int i,j;
v = (double***)malloc(paraNB ->indivNum*sizeof(double**));
for(i = 0; i < paraNB->indivNum; i++) {
v[i] = (double**)malloc((paraNB->time+1) *sizeof(double*));
for(j=0;j<paraNB->time+1;j++){
v[i][j]=(double*)malloc(20*sizeof(double));
}
}
```

```
return v;
}
void free_ind(ind* in, glo* paraNB){
int i;
for(i=0;i<paraNB->indivNum;i++) {
free(in[i].indiv);
}
free(in);
}
void free_cur(cur* curr, glo* paraNB){
int i;
for(i=0;i<paraNB->indivNum;i++){
free(curr->current[i]);
}
free(curr->current);
free(curr);
}
void free_poten(poten* p, glo* paraNB){
int i;
for(i=0;i<paraNB->indivNum;i++){
free(p->stati[i]);
free(p->decision[i]);
}
free(p->stati);
free(p->decision);
free(p);
}
void free_err (Aldata *d) {
free_p(d->err1, d->glonum);
free(d->err2);
}
void free_glo (glo* g){
free(g);
}
void free_v(double*** v, glo* paraNB){
int i, t;
for(i = 0; i < paraNB ->indivNum; i++) {
for(t=0; t<paraNB->time+1; t++) {
free(v[i][t]);
}
free(v[i]);
}
free(v);
}
void free_u(double*** u, glo* paraNB){
```

```
int i, j;
for(i = 0; i < paraNB ->indivNum; i++) {
for(j=0; j<paraNB->numch; j++) {
free(u[i][j]);
}
free(u[i]);
}
free(u);
}
void free_uy(double**** u, glo* paraNB){
int i, j, t;
for(i = 0; i < paraNB ->indivNum; i++) {
for(j=0; j<paraNB->numch; j++) {
for (t=0; t<paraNB->time+1; t++){
free(u[i][j][t]);
}
free(u[i][j]);
}
free(u[i]);
}
free(u);
}
void free_p(double*** u, glo* paraNB){
int i, j;
for(i = 0; i < paraNB ->indivNum; i++) {
for(j=0; j<paraNB->numch+1; j++) {
free(u[i][j]);
}
free(u[i]);
}
free(u);
}
```

• cal_v.c

```
#include <math.h>
#include <float.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include "data.h"
#include <oratio/oratio.h>
/*
U is utility of each i, j, t;
v is utility of each i, t, gumbel distributed, vit is randomly generated
r is mode of v
j=0, choice gas vehicle, u=asc+y*beta_y
j=1, choice hybrid vehicle, u=asc+indiv*beta_indiv+static1*beta_sta1+y*beta_y
j=2, choice electrical vehicle, u=indiv*beta_indiv+static2*beta_sta2
*/
/*make draws for y from normal distribution;
parameters are from JM's calibration;
There is a concept of scenario tree. From each current gas price, one
scenario tree is generated with a two-time-period
expansion which means the respondent can imagine all the possible situations
happen for the next two time periods when standing at
current time.
y[t][n] is gas price in scenario trees which is dynamic variable.
From the root of the tree, there are two levels of gas prices generated. From
every price, four hypothetical prices are generated.
The root price is gas price at current time period, y_int[t]. For example,
from root price y_int[0], four prices at time 1
in the scenario tree are generated; from each of the four prices at time 1,
another four prices at time 2 are generated seperately.
Therefore, for each current time period, total 20 gas prices will be
generated. In the function, the 20 prices are put in one array,
but this array is divided by two levels.
From the current price y_int[t] at time t, four prices y[t+1][0], y[t+1][1],
y[t+1][2], y[t+1][3] at time t+1 are generated first;
then, from y[t+1][0],we have y[t+1][4],y[t+1][5],y[t+1][6],y[t+1][7]
generated;from y[t+1][1],we have y[t+1][8],y[t+1][9],y[t+1][10],y[t+1][11]
generated;
from y[t+1][2],we have y[t+1][12],y[t+1][13],y[t+1][14],y[t+1][15]
generated;from y[t+1][3],we have y[t+1][16],y[t+1][17],y[t+1][18],y[t+1][19]
generated;
*/
/*
y_int[t] is 3.XX, but the calibration function is only adapted to 3XX, so i
put d->y_int[t]*100 for the function;
therefore, y[t][n] generated will be 3XX, but the utility function will need
3.xx, so i divide y[t][n] by 100 in the end.
*/
double** draw_random_y(Aldata* d, double **draw){
int t, n;
double **y;
```

```
y=nt_matrix_new(d->glonum->time+1, 20);
for (t=0; t<d->glonum->time; t++){
for (n=0; n<4; n++){
y[t+1][n]=0.9757*d->y_int[t]*100+4.49+draw[t][n];
y[t+1][4*n+4+0] = 0.9757*y[t+1][n]+4.49+draw[t][4*n+4+0];
y[t+1][4*n+4+1] = 0.9757*y[t+1][n]+4.49+draw[t][4*n+4+1];
y[t+1][4*n+4+2] = 0.9757*y[t+1][n]+4.49+draw[t][4*n+4+2];
y[t+1][4*n+4+3] = 0.9757*y[t+1][n]+4.49+draw[t][4*n+4+3];
}
}
for (t=1; t<d->glonum->time+1; t++){
for (n=0; n<20; n++){
y[t][n]=y[t][n] /100.0;
}
}
return y;
}
/*calculate mode r[i][t] for the current time period with y_int[t] and all
other variables correlated
r_it=sum(exp(U_ijt));
j=0,choice gas vehicle, u=asc+y*beta_y
j=1,choice hybrid vehicle, u=asc+indiv*beta_indiv+static1*beta_sta1+y*beta_y
j=2, choice electrical vehicle, u=indiv*beta_indiv+static1*beta_sta1
*/
double** calculate_mode_real(Aldata* d, double* x){
int i=d->glonum->indivNum;
int j=d->glonum->numch;
int t;
int k, l, m;
double sum=0;
int l0= d->glonum->numch-1;
int l1= d->glonum->numINDIVAR;
int l2= d->glonum->numch;
int l3= d->glonum->numDYNAMIC;
double **r_real;
r_real= nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
double ***U=c_malloc_u(d->glonum);
// calculate utility for three choices, from time 0
```

```
for(i = 0; i < d->glonum ->indivNum; i++) {
for(t=0; t<d->glonum->time; t++) {
j=0;
sum = 0.0;
sum+=x[j];
for (m=0; m<l3;m++ )
{
sum+=d->y_int[t]*x[l0+l1+l2-1+m];
}
U[i][j][t] = sum;
j=1;
sum = 0.0;
sum+=x[j];
for (k=0; k<l1;k++ )
{
sum+=d->in[i].indiv[k]*x[l0+k];
}
for (l=0; l<d->glonum->numSTATIC-1;l++ )
{
sum+=(d->pot->stati[i][ (d->glonum->numSTATIC)*j+l+t*d->glonum-
>numSTATIC*l2])*x[l0+l1+l];
}
for (m=0; m<l3;m++ )
{
sum+=d->y_int[t]*x[l0+l1+l2-1+m];
}
U[i][j][t] = sum;
j=2;
sum = 0.0;
for (k=0; k<l1;k++ )
{
sum+=d->in[i].indiv[k]*x[l0+k];
}
for (l=0; l<d->glonum->numSTATIC-1;l++ )
{
sum+=(d->pot->stati[i][ (d->glonum->numSTATIC)*j+l+t*d->glonum-
>numSTATIC*l2])*x[l0+l1+(d->glonum->numSTATIC-1)*(j-1)+l];
}
U[i][j][t] = sum;
}
```

```
}
for(t=0; t<d->glonum->time; t++) {
for(i = 0; i < d->glonum ->indivNum; i++) {
sum=0;
for (j = 0;j<d->glonum->numch;j++) {
sum+=exp(U[i][j][t]);
}
r_real[i][t]= log(sum);
}
}
free_u(U,d->glonum);
return r_real;
}
/*calculate mode for the scenario tree, r[i][t][n], n means the position in
the tree;
in order to get mode, utilities need to be calculated, UY[i][j][t][n];
utilities for each time period have two levels,
UY[i][j][t][n],n=0,1,2,3 are in the first level with correlated variables at
time t and y[t][n],
n=0,1,2,3; UY[i][j][t][n],n=4...19 are in the second level with correlated
variables at time t+1
and y[t][n],n=4...19;
so for the mode r[i][t][n], n=0,1,2,3 are in the first level; n=4...19 are in
the second level.
*/
double*** calculate_mode(Aldata* d, double* x, double** y){
int i=d->glonum->indivNum;
int j=d->glonum->numch;
int t;
int k, l, m, n;
double sum=0;
double ***r;
int l0= d->glonum->numch-1;
int l1= d->glonum->numINDIVAR;
int l2= d->glonum->numch;
int l3= d->glonum->numDYNAMIC;
r= c_malloc_v(d->glonum);
double ****UY = c_malloc_uy(d->glonum);
for(i = 0; i < d->glonum ->indivNum; i++) {
for(t=1; t<d->glonum->time+1; t++) {
//calculate utilities for the first level of scenario tree, n=0,1,2,3
for(n=0; n<4; n++){
j=0;
```

```
sum = 0.0;
sum+=x[j];
for (m=0; m<l3;m++ )
{
sum+=y[t][n]*x[l0+l1+l2-1+m];
}
UY[i][j][t][n] = sum;
j=1;
sum = 0.0;
sum+=x[j];
for (k=0; k<l1;k++ )
{
sum+=d->in[i].indiv[k]*x[l0+k];
}
for (l=0; l<d->glonum->numSTATIC-1;l++ )
{
sum+=(d->pot->stati[i][ (d->glonum->numSTATIC)*j+l+t*d->glonum-
>numSTATIC*l2])*x[l0+l1+(d->glonum->numSTATIC-1)*(j-1)+l];
}
for (m=0; m<l3;m++ )
{
sum+=y[t][n]*x[l0+l1+l2-1+m];
}
UY[i][j][t][n] = sum;
j=2;
sum = 0.0;
for (k=0; k<l1;k++ )
{
sum+=d->in[i].indiv[k]*x[l0+k];
}
for (l=0; l<d->glonum->numSTATIC-1;l++ )
{
sum+=(d->pot->stati[i][ (d->glonum->numSTATIC)*j+l+t*d->glonum-
>numSTATIC*l2])*x[l0+l1+(d->glonum->numSTATIC-1)*(j-1)+l];
}
UY[i][j][t][n] = sum;
}
//calculate utilities for the second level of scenario tree, n=4,...19
for(n=4; n<20; n++){
j=0;
sum = 0.0;
```

```
sum+=x[j];
for (m=0; m<l3;m++ )
{
sum+=y[t][n]*x[l0+l1+l2-1+m];
}
UY[i][j][t][n] = sum;
j=1;
sum = 0.0;
sum+=x[j];
for (k=0; k<l1;k++ )
{
sum+=d->in[i].indiv[k]*x[l0+k];
}
for (l=0; l<d->glonum->numSTATIC-1;l++ )
{
sum+=(d->pot->stati[i][ (d->glonum->numSTATIC)*j+l+(t+1)*d->glonum-
>numSTATIC*l2])*x[l0+l1+(d->glonum->numSTATIC-1)*(j-1)+l];
}
for (m=0; m<l3;m++ )
{
sum+=y[t][n]*x[l0+l1+l2-1+m];
}
UY[i][j][t][n] = sum;
j=2;
sum = 0.0;
for (k=0; k<l1;k++ )
{
sum+=d->in[i].indiv[k]*x[l0+k];
}
for (l=0; l<d->glonum->numSTATIC-1;l++ )
{
sum+=(d->pot->stati[i][ (d->glonum->numSTATIC)*j+l+(t+1)*d->glonum-
>numSTATIC*l2])*x[l0+l1+(d->glonum->numSTATIC-1)*(j-1)+l];
}
UY[i][j][t][n] = sum;
}
}
}
for(i = 0; i < d->glonum ->indivNum; i++) {
```

```
for(t=1; t<d->glonum->time+1; t++) {
for(n=0; n<20; n++){
sum=0;
for (j = 0;j<d->glonum->numch;j++)
{
sum+=exp(UY[i][j][t][n]);
}
r[i][t][n]= log(sum);
}
}
}
free_uy(UY,d->glonum);
return r;
}
/*
v is randomly drawn from gumbel distribution with mode r_itn, also in the
scenario tree;
n means the position of v in the tree;
v[i][t][n], n=0,1,2,3 are in the first level; n=4...19 are in the second
level
*/
double*** calculate_v(Aldata* d, double* x, double** y, double vp){
int t,i,n;
double*** v= c_malloc_v(d->glonum);
double*** r=calculate_mode(d, x, y);
for(i = 0; i < d->glonum ->indivNum; i++) {
for(t=1; t<d->glonum->time+1; t++) {
for(n=0; n<20; n++){
v[i][t][n]= st_gumbel_icdf(vp, r[i][t][n], 1 );
}
}
}
free_v(r, d->glonum);
return v;
}
inline double max(double a, double b) {
if (a<b) {
```

```
return b;
}
return a;
}
/*p_ijt is probability of purchasing the j kind of car;
p is calculated through traditional logit way;
err1 is error term for i,j,t; err2 is error term for i,t;
*/
double*** cal_probcar (Aldata* d, double* x){
int i=d->glonum->indivNum;
int j=d->glonum->numch;
int t;
int k, l, m;
double sum=0;
double ***err1, *err2;
int l0= d->glonum->numch-1;
int l1= d->glonum->numINDIVAR;
int l2= d->glonum->numch;
int l3= d->glonum->numDYNAMIC;
int T=d->glonum->time;
double ***P =c_malloc_u(d->glonum);
double **w=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
double ***U = c_malloc_u(d->glonum);
err1 = d->err1;
err2 = d->err2;
for(t=0; t<T; t++) {
for(i = 0; i < d->glonum ->indivNum; i++) {
j=0;
sum = 0.0;
sum+=x[j];
for (m=0; m<d->glonum->numDYNAMIC;m++ )
{
sum+=d->y_int[t]*x[l0+l1+l2-1+m];
}
U[i][j][t] = sum+err1[i][j][t]+err2[i];
j=1; //choice hybrid vehicle,
u=asc+indiv*beta_indiv+static1*beta_sta1+y*beta_y
sum = 0.0;
sum+=x[j];
for (k=0; k<l1;k++ )
{
```

```
sum+=d->in[i].indiv[k]*x[l0+k];
}
for (l=0; l<d->glonum->numSTATIC-1;l++ )
{
sum+=(d->pot->stati[i][ (d->glonum->numSTATIC)*j+l+t*d->glonum-
>numSTATIC*l2])*x[l0+l1+(d->glonum->numSTATIC-1)*(j-1)+l];
}
for (m=0; m<l3;m++ )
{
sum+=d->y_int[t]*x[l0+l1+l2-1+m];
}
U[i][j][t] = sum+err1[i][j][t]+err2[i];
j=2; //choice electrical vehicle, u=indiv*beta_indiv+static1*beta_sta1
sum = 0.0;
for (k=0; k<l1;k++ )
{
sum+=d->in[i].indiv[k]*x[l0+k];
}
for (l=0; l<d->glonum->numSTATIC-1;l++ )
{
sum+=(d->pot->stati[i][ (d->glonum->numSTATIC)*j+l+t*d->glonum-
>numSTATIC*l2])*x[l0+l1+(d->glonum->numSTATIC-1)*(j-1)+l];
}
U[i][j][t] = sum+err1[i][j][t]+err2[i];
}
}
for(t=0; t<d->glonum->time; t++) {
for(i = 0; i < d->glonum ->indivNum; i++) {
sum=0;
for (j = 0;j<d->glonum->numch;j++)
{
sum+= exp(U[i][j][t]);
}
w[i][t]=sum;
for (j = 0;j<d->glonum->numch;j++)
{
P[i][j][t]= exp(U[i][j][t])/w[i][t];
}
```

```
}
}
nt_matrix_free(w);
free_u(U,d->glonum);
return P;
}
/*recursive process for calculating E_AVE;
n is the position in the tree
*/
double cal_E(Aldata* d, int t, int T, double *v, double current, int n,
double* x, int indiv){
int i;
double e_ave;
double ***err1, *err2;
double c;
err1 = d->err1;
err2 = d->err2;
int l0= d->glonum->numch-1;
int l1= d->glonum->numINDIVAR;
int l2= d->glonum->numch;
int l3= d->glonum->numDYNAMIC;
// Base case to cover the last time period
/*
if (T < t)
return 0;
*/
c = current* x[l0+l1+l2-1+l3]+err1[indiv][3][t]+err2[indiv];
if (t==T) // Base Case
return max(v[n], c);
else // Recursive Step
e_ave=0;
for(i=0; i<4; i++){
e_ave+=cal_E(d, t+1, T, v, current+0.5, (4+4*n)+i, x, indiv);// go further to
reach the second level in the tree
}
e_ave=e_ave/4;
return max(v[n], c+e_ave);
}
```

```
/*PI0,PI1 is probability of buying and not buying;
C_it is utility payoff when not buying, =indiv*beta_indiv+mile*beta_mile;
PI0 = F(v<W), Wit=Cit+Eit+1, E_it+1=max(v_it+1, C_it+1+E_it+2);
E_AVE, the average expectation at each time period from 8 children
excpectations;
*/
double** cal_prob (Aldata* d, double* x, double** y){
double **PI0,**PI1, **C, **W, **E_AVE;
double sum;
double ***err1, *err2;
int t, t2, i, n;
int l0= d->glonum->numch-1;
int l1= d->glonum->numINDIVAR;
int l2= d->glonum->numch;
int l3= d->glonum->numDYNAMIC;
int T=d->glonum->time;
err1 = d->err1;
err2 = d->err2;
C=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
W=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
E_AVE=nt_matrix_new(d->glonum-> indivNum, d->glonum->time+1);
PI0=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
PI1=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
double ***v=calculate_v(d,x,y,d->vp);
double **r_real=calculate_mode_real(d,x);
for(i = 0; i < d->glonum ->indivNum; i++) {
for(t=0; t<d->glonum->time; t++) {
sum = 0.0;
sum+=d->curr->current[i][t]* x[l0+l1+l2-1+l3];
C[i][t] = sum+err1[i][3][t]+err2[i];
}
}
// calculate expecations E_AVE in the first level of the tree in a recursive
way;
for(i = 0; i < d->glonum ->indivNum; i++) {
for (t=0; t<T; t++){
t2 = t+2;
/*
```

```
if(t2 == T)
t2 = T-1;
else {
if(t2 == T+1)
t2 = 0;
}
*/
E_AVE[i][t+1]=0;
for(n=0; n<4; n++){
E_AVE[i][t+1]+= cal_E(d, t+1, t2, v[i][t+1], d->curr->current[i][t]+0.5, n, x,
i);
}
E_AVE[i][t+1]=E_AVE[i][t+1]/4;
W[i][t]= C[i][t]+E_AVE[i][t+1];
}
}
//calculate probabilities of postponing PI0 with reservation utility W, mode
r
for(i = 0; i < d->glonum ->indivNum; i++) {
for(t=0; t<d->glonum->time; t++) {
PI0[i][t]=st_gumbel_cdf(W[i][t], r_real[i][t], 1);
PI1[i][t]=1-PI0[i][t];
}
}
nt_matrix_free(C);
nt_matrix_free(W);
nt_matrix_free(E_AVE);
free_v(v,d->glonum);
nt_matrix_free(r_real);
nt_matrix_free(PI1);
return PI0;
}
```

- LL.c

```
#include <math.h>
#include <float.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include "data.h"
#include <ophelia/nlp.h>
#include <ophelia/nlp_collection.h>
#include <oratio/oratio.h>
//#include "tstat.h"
/*PB is the product specific purchase probability (j-0,1,2);
PB=PI1*P
j=3 PB is the probability of not buy
*/
double fLL(double* x, int n, void* data) {
Aldata* d = (Aldata*) data;
double ***PB = d->prob_matrix;
double** y= draw_random_y(d, d->draw);
double ***P =cal_probcar(d,x);
double **PI0=cal_prob (d,x,y);
int i, j, t;
double ***ch, LL;
ch=c_malloc_P(d->glonum);
for(i = 0; i < d->glonum ->indivNum; i++){
for (t=0; t<d->glonum->time; t++) {
for (j = 0;j<d->glonum->numch+1;j++) {
ch[i][j][t]=d->pot->decision[i][t*4+j];
}
}
}
for(i = 0; i < d->glonum ->indivNum; i++){
for (t=0; t<d->glonum->time; t++) {
for (j = 0;j<d->glonum->numch;j++)
{
PB[i][j][t]= (1-PI0[i][t])*P[i][j][t];
}
PB[i][j][t]= PI0[i][t];
}
}
LL=0;
```

```
for(i = 0; i < d->glonum ->indivNum; i++) {
for (t=0; t<d->glonum->time; t++) {
for (j = 0;j<d->glonum->numch+1;j++)
{
LL+=ch[i][j][t]*log(PB[i][j][t]);
}
}
}
// printf("Log likelihood:");
//printf(" %f", LL);
//printf("\n");
free_p(ch, d->glonum);
nt_matrix_free(PI0);
nt_matrix_free(y);
free_u(P, d->glonum);
return -LL/(200*12);
}
/* optimization;
H, preallocated array of size btr->n*btr-btr->n, if the hessian is needed;
I, Hessian matrix;
I1, inversed Hessian matrix;
*/
int btr_unconstrained_opt(NTLog *log, BTR *b, Aldata* d)
{
double **H;
double *t, *h;
int n=get_dimension(d);
int i;
double tol=0; //sets tolerance and scale for hessian derivation
double scale[n];
int s;
FILE *out;
out=fopen("matrix.txt","w");
t=malloc(n*sizeof(double));
double work[100*(b->n)];
//b->retro = 1;
H = nt_matrix_new(n, n);
nt_matrix_identity(n, n, *H, n);
nt_log_subsection(log, "optim of Log likelihood");
nlp_btr_init(b, n, 0);
```

```
/* Starting point */
// b->x[0] = -1;
read_new_para(b->x, d->glonum);
b->printer = btr_print_iteration;
nlp_btr(b, (C_GENERIC)fLL, NULL, d, log->f, H, work);
//derive hessian matrix
for(s = 0; s < n; ++s)
scale[s]=1.0;
h = malloc(n*sizeof(double));
nt_derive_hess_cd((C_GENERIC)fLL, b->x, H, h, tol, scale, n, NULL, work,
(void*) d);
op_matrix_inverse(CblasRowMajor, CblasUpper, *H, n);
double bugfound[n];
int bugindex = 0;
for(bugindex=0; bugindex<n; ++bugindex)
bugfound[bugindex] = 0;
amlet_t_statistics(n, b->x, bugfound, H, 0.05, t);
//Print out inversed hessian matrix
nt_matrix_print(out, "matrix", H, n, n);
// Print out the t-statistics
printf("t:");
for(i=0; i < n; i++) {
printf(" %f", t[i]);
}
return 0;
}
int main(int argc, char **argv) {
BTR *b = malloc(sizeof(BTR));
NTLog *log;
Aldata *d = (Aldata*) format_data();
int n=get_dimension(d);
nlp_btr_init(b, n, 0);
log = nt_log_new(NULL);
btr_unconstrained_opt(log, b, d);
```

```
nt_log_free(log);
nlp_btr_free(b);
free_ind(d->in, d->glonum);
free_cur(d->curr, d->glonum);
free_poten(d->pot, d->glonum);
free_err(d);
free_glo(d->glonum);
free(d);
return 0;
}
;
```

BIBLIOGRAPHY

Emmanuel Abbe, Michel Bierlaire, and Tomer Toledo. Normalization and correlation of cross-nested logit models. *Transportation Research Part B: Methodological*, 41(7):795–808, 2007.

Victor Aguirregabiria and Pedro Mira. Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, forthcoming, 2009.

Staffan Algers, Andrew Daly, and Staffan Widlert. The stockholm mode system - travel to work. *presented to World Conference on Transportation Research, Yokohama*, 1989.

AVV. Variabilisatie van autokosten en de aanbodzjide van de mobiliteitmarkt. Rotterdam, Novermber 2000. eindrapport, AVV.

Moshe Ben-Akiva and Maya Abou-Zeid. Hybrid choice models: from static to dynamic. In *Oslo Workshop on Valuation Methods in Transport Planning*, March 19-20 2007.

Moshe Ben-Akiva and Steven Lerman. *Discrete Choice Analysis: Theory and application to travel demand*. The MIT Press, 1985.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890, 1995.

Chandra R. Bhat and Vamsi Pulugurta. A comparison of two alternative behavioural choice mechanisms for household auto ownership decisions. *Transportation Research Part B: Methodological*, 32(1):61C7, 1998.

David Brownstone, David S. Bunch, and Kenneth Train. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, 34(5):315–338, 2000.

David S. Bunch, David Brownstone, and Thomas F. Golob. A dynamic forecasting system for vehicle markets with clean-fuel vehicles. *World Transport Research*, 1: 189–203, 1996.

Kenneth Button, Ndoh Ngoe, and John Hine. Modelling vehicle ownership and use in low-income countries. *Journal of Transport Economics and Policy*, 27, 1993.

Juan Esteban Carranza. Consumer heterogeneity, demand for durable goods and the dynamics of quality. *Meeting Papers from Society for Economic Dynamics*, (47), 2006.

Charisma Farheen Choudhury. *Modeling Driving Decisions with Latent Plans*. PhD thesis, Department of Civil and Environmental Engineering, MIT, September 2007.

Chaushie Chu. *Structural issues and sources of bias in residential location and travel choice models*. PhD thesis, Northwestern University, 1981.

Chaushie Chu. A paired combinational logit model for travel demand analysis. *Proceedings of Fifth World Conference on Transportation Research*, 4:295–309, 1989.

J. S Cramer and A. Vos. Een model voor prognoses van het personenauto park. University of Amsterdam, 1985. Amsterdam: Interfaculty of Actuarial Science en Econometrics.

Carlos Daganzo, Fernando Bouthelier, and Yosef Sheffi. Multinomial probit and qualitative choice: A computationally efficient algorithm. *Transportation Science*, 11:338–358, 1977.

Andrew J. Daly. Estimating choice models containing attraction variables. *Transportation Research*, B16:5–15, 1982.

Joyce Dargay and Dermot Gately. Income's effect on car and vehicle ownership worldwide: 1960-2015. *Transportation Research Part A: Policy and Practice*, 33 (2), 1999.

Joyce M. Dargay and Petros C. Vythoulkas. Car ownership in rural and urban areas: a pseudo panel analysis. *(London: ESRC Transport Studies Unit, Centre for Transport Studies, University College London)*, 1999a.

Joyce M. Dargay and Petros C. Vythoulkas. Estimation of a dynamic car ownership model; a pseudopanel approach. *Journal of Transport Economics and Policy*, 33 (3):287–302, 1999b.

Van den Broecke/Social Research. De mogelijke groei van het personenautobezit tot 2010. Technical report, Report for PbIVVS. (Amsterdam: BSR).

Edited by Randolph W. Hall. *HANDBOOK OF TRANSPORTATION SCIENCE, Second Edition*. Kluwer, 2003.

Gerard De Jong James Fox, Andrew Daly, Marits Pieters, and Remko Smit. Comparison of car ownership models. *Transport Reviews*, 24:379–408, 2004.

Carol C. S. Gilbert. A duration model of automobile ownership. *Transportation Research Part B: Methodological*, 26(2):97–111, 1992.

Vladislav Golounov, Benedict Dellaert, and Harry J P Timmermans. A dynamic lifetime utility model of car purchase behavior using revealed preference consumer panel data. Washington, DC, USA., 2001. Paper presented at the 81st Annual Meeting of the Transportation Research Board.

Brett R. Gordon. Estimating a dynamic model of demand for durable goods. *Unpublished manuscript*, Unpublished manuscript, 2006.

Gautam Gowrisankaran and Marc Rysman. Dynamics of consumer demand for new durable goods. *Working Paper Series*, 2007.

Mark Hanly and Joyce M. Dargay. Car ownership in great britaina panel data analysis. *ESRC Transport Studies Unit, University College London.*, 2000.

HCG. Resource papers for landelijk model. 2, 1989.

HCG. Sydney car ownership models. Report 9009-3B, 2000.

HCG and TOI. A model system to predict fuel use and emissions from private travel in norway from 1985 to 2025. The Netherlands, 1990. Hague Consulting Group.

David A. Hensher and William H. Greene. Choosing between conventional, electric and lpg/cng vehicles in single-vehicle households. Gold Coast, Australia, 2000. Paper presented at IATBRC200.

David A. Hensher and Tu Ton. Tresis: a transportation, land use and environmental strategy impact simulator for urban areas. *Transportation*, 29(4):439–457, 2002.

David A. Hensher, Peter O. Barnard, Nariida C. Smith, and Frank W. Milthorpe. Dimensions of automobile demand; a longitudinal study of automobile ownership and use. North-Holland, Amsterdam, 1992.

Moshe Hirsh, Joseph N. Prashker, and Moshe Ben-Akiva. Day-of-the-week models of shopping activity patterns. *Transportation Research Record*, 1085:63–69, 1986.

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

Chieh hua Wen and Frank Koppelman. The generalized nested logit model. *Transportation Research Part B*, 35:627–641, 2001.

Robert A. Johnston. The urban transportation planning process. *University of California Davis*, 2003.

Gerard De Jong. *Some joint models of car ownership and car use.* PhD thesis, Faculty of Economic Science and Econometrics, University of Amsterdam, 1989a.

Gerard De Jong. Simulating car cost changes using an indirect utility model of car ownership and car use. Brighton, UK, 1989b. Paper presented at PTRC SAM 1989, PTRC.

Gerard De Jong. An indirect utility model of car ownership and car use. *European Economic Review*, 34:971–985, 1991.

Gregory K.Ingram and Zhi Liu. Motorization and the provision of roads in countries and cities. In *Policy Research Working Paper*, 1842. 1997. (Washington, DC: World Bank).

Ryuichi Kitamura. A panel analysis of household car ownership and mobility. In *Proceedings of the Japan Society of Civil Engineers*, volume 383, pages 13–27, 1987.

Ryuichi Kitamura and DAVID S. BUNCH. Heterogeneity and state dependence in household car ownership: a panel analysis using ordered-response probit models with error components. *Transportation and traffic theory*, 8 Reprint n.52:477–496, 1990.

Ole Kveiborg. Forecasting developments of the car fleet in the altrans model. Trondheim, Norway, March 2001. Paper presented to the Nordic Research Network on Modelling Transport, Land-Use and the Environment, 3rd Workshop.

William H.K Lam, Zhi-Chun Li, Hai-Jun Huang, and S.C.Wong. Modeling time-dependent travel choice problems in road networks with multiple user classes and multiple parking facilities. *Transportation Research Part B*, 40:368–395, 2006.

Uzi Landau, Joseph N. Prashker, and Moshe Hirsh. The effect of temporal constraints on household travel behavior. *Environment and Planning*, A13:435–448, 1981.

K.U Leuven. Standard and poor's dri auto-oil ii cost-effectiveness: Study description of the analytical tools tremove 1.1. *Second Draft, Working Document, K.U. Leuven en Standard and Poor's DRI*, February 1989.

Yu-Hsin Liu and Hani S. Mahmassani. Dynamic aspects of commuter decisions under advanced traveler information systems: modeling framework and experimental results. *Transportation Research Record*, 1645:111–119, 1998.

Szabolcs Lorincz. Persistence effects in a dynamic discrete choice model. application to low-end computer servers. *JOB MARKET PAPER*, October 2005.

Michael Maness. Modeling vehicle ownership decisions in maryland: A preliminary stated-preference survey and model, December 2010.

Charles F. Manski. Analysis of equilibrium automobile holdings in israel with aggregate discrete choice models. *Transportation Research Part B: Methodological*, 17(5):373–389, 1983.

Charles F. Manski and Leonard Sherman. An empirical analysis of household motor vehicle holdings. *Transportation Research Part A: Policy and Practice*, 14(5/6): 349–366, 1980.

Daniel L. McFadden. Modeling the choice of residential location. *Transportation Research Record*, 672:72–77, 1978.

Daniel L. McFadden and Kenneth Train. Mixed mnl models of discrete response. *Journal of Applied Econometrics*, 15:447–470, 2000.

Oleg Melnikov. Demand for differentiated durable products: The case of the u.s. computer printer market. *Yale University*, 2000.

Hendrik Jan Meurs. A panel data analysis of travel demand. Groningen: Groningen University, 1991.

Hendrik Jan Meurs. A panel data switching regression model of mobility and car ownership. *Transportation Research Part A: Policy and Practice*, 27(6):461–476, 1993.

Harikesh Nair. Intertemporal price discrimination with forward-looking consumers: Application to the us market for console video games. *Quantitative Marketing and Economics*, in press, 2007.

Agostino Nobile, Chandra R. Bhat, and Eric I. Pas. A random effects multinomial probit model of car ownership choice. *Research Paper (Amherst, MA: Duke University and University of Massachusetts)*, 1996.

Matthew Page, Gerard Whelan, and Andrew Daly. Modelling the factors which influence new car purchasing. Cambridge, UK, 2000. Paper presented at the European Transport Conference 2001, PTRC.

Ariel Pakes. Patents as options: Some estimates of the value of holding european patent stocks. *Econometrica*, 54:755–785, 1986.

Andrea Papola. Some development of the cross-nested logit model. *Proceedings of the 9th IATBR Conference*, 2000.

Jeppe Rich and Otto Anker Nielsen. A microeconomic model for car ownership, residence and work location. Cambridge, UK, 2001. Paper presented at the European Transport Conference 2001, PTRC.

John Rust. Optimal replacement of gmc buses: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033, 1987.

John Rust. *Numerical Dynamic Programming in Economics*. Revised November 1994 draft for Handbook of Computational Economics, 1994.

Kenneth Small. A discrete choice model for ordered alternatives. *Econometrica*, 55 (2):409–424, 1987.

Nariida C. Smith, David A. Hensher, and Neil Wrigley. A dynamic discrete choice sequence model: Method and an illustrative application to automobile transactions. In *Working Paper (Sydney: Macquarie University)*, 1989.

Inseong Song and Pradeep K. Chintagunta. A micromodel of new product adoption with heterogeneous and forward-looking consumers: An application to the digital camera category. *Quantitative Marketing and Economics*, 1:371–407, 2003.

J.S. Tanner. Methods of forecasting kilometers per cars. *Transport and Road Research Laboratory,Department of the Environment and of Transport, Crowthorne, Berkshire*, 1981.

Kenneth Train. Consumers' responses to fuel efficient vehicles. *Transportation*, 8: 237–258, 1979.

Kenneth Train. *Qualitative Choice Analysis: Theory, Econometrics and an Application to Automobile Demand*. MIT Press, Cambridge, MA:, 1986.

Kenneth Train. *Discrete Choice Methods with Simulation*. The MIT PressCambridge University Press, 2002a.

Kenneth E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, September 18 2002b.

Gerard Whelan. Methodological advances in modelling and forecasting car ownership in great britain. Cambridge, UK, 2001. Paper presented at the European Transport Conference 2001, PTRC.

Gerard Whelan, Mark Wardman, and Andrew Daly. Is there a limit to car ownership growth? an exploration of household saturation levels using two novel approaches. Cambridge, UK, 2000. Paper presented at the European Transport Conference 2000, PTRC.

Kenneth I. Wolpin. An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political Economy*, 92:852–874, 1984.

Kenneth I. Wolpin. Estimating a structural search model: The transition from school to work. *Econometrica*, 55:801–818, 1987.