A dynamic formulation for car ownership modeling

Cinzia Cirillo and Renting Xu

Department of Civil & Environmental Engineering
University of Maryland, College Park, MD 20742, USA

December 1, 2009

Abstract

Discrete choice models are commonly used in transportation planning and modeling, but their theoretical basis and their applications have been mainly developed in a static context. We propose to develop an estimation technique for analyzing the impact of technological change on the dynamics of consumer demand. The proposed research presents a dynamic model of consumer demand that explicitly accounts for consumers’ expectations of future product quality and the market evolution. The timing of consumers’ purchases is formalized as an optimal stopping problem where the agent must decide on the optimal time of purchase. This model frame will be further improved by modeling the choice from a set of differentiated products whose quality changes stochastically over time. The framework proposed is developed in the context of car ownership.

1 Introduction

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. The multinomial logit (MNL) (Ben-Akiva and Lerman[1]) model has been the most widely used structure for modeling discrete choices in travel behavior analysis. The random components of the utilities of the different alternatives in the MNL model are assumed to be independently and identically distributed (IID) with a type I extreme-value (or Gumbel) distribution. Nested logit (NL) model (Daly [2]) relaxes in part MNL model assumptions; it is derived from the generalized extreme value (GEV) model.
(McFadden[3]) and in particular it allows dependence between the utilities included in a common group (or nest). 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 [4]), the paired combinatorial logit (PCL) model (Chu [5, 6]) and the cross-nested logit (CNL) model (Wen and Koppelman[7]; Abbe, et al. [8]; Papola [9]). Non-closed form discrete choice models as Probit (Daganzo [10]) and Mixed logit (McFadden and Train [11]) 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 the formulation of changes occurring in the system.

Recently, in transportation researchers are trying to incorporate dynamics effects in their models. Moshe Ben-Akiva and Maya Abou-Zeid [12] have 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 (2007). Choudhury [13] in her MIT PhD thesis has 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 of the external conditions. Therefore, in transportation the development of dynamic discrete choice models has not been as comprehensive as in economics or marketing. The present paper provides a review of current car ownership models and proposes a dynamic framework based on discrete choice models. The formulation developed here explicitly accounts for change in state over time and for industry evolution. Finally, we present our conclusions and the avenues for future research opportunities.
2 Existing car ownership applications

Car users’ preferences are usually modeled with demand models, using one of the two possible forms: aggregate or disaggregate. Aggregate car ownership models are mainly of three types: (1) time series models (Tanner[14], Dargay and Gately [15]), (2) cohort models (Algers et al.[16]) and (3) car market models (TREMOVE [17]). Aggregate models no longer appear in academic journals but are still used in practice; their major limitation is the impossibility to model vehicle type and use; they also usually include limited socio-demographic variables. In this paper we refer to aggregate models, because they can be dynamic over time (with the exception of car market models) and they require limited amounts of data. In practice, they could be used just to predict a total number of cars in a future year (especially medium- to long-run), which would then be used as a starting point in other, more detailed 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 [18]), Norway (HCG and TOI [19]), Sydney (Hensher et al.[20]), US (Mannering and Winston [21]). Their success is due to their behavioral foundations and to the possibility of including a large number of policy variables, as well as car types and use. Car ownership decisions can also be linked to a range of other travel choices allowing impact on car ownership of variable cost, public transport cost and congestion price to be represented. Pseudo-dynamics models are possible but rare in the literature because of the main problem being the equilibrium assumption if applied for shorter time period (less than 5-10 years). Large data sets are required, but usually available in travel surveys (i.e. number of cars per household and type): data on vehicles that are not marketed would require special data collection from surveys on hypothetical scenarios (i.e. stated preference Whelan et al.[22]). Dynamic models are introduced to estimate short and long run effects; they are able to determine whether a household will do a vehicle transaction. Duration models for the time between vehicle transactions (and the type of transaction: disposal, replacement, acquisition, also scrappage) have in general been used to explain the total number of cars (Bhat and Pulugurta [23]). Their data requirement is in general very heavy, panel or retrospective data are required. Transactions involving new car types (Alternative-fuel vehicles and/or automated vehicles) need Stated Preference (SP) data to be modeled (Brownstone et al.[24]; Hensher and Greene [25]). Pseudo-panel models have been recently proposed to circumvent the need for panel data and their associated problems. A pseudo panel is an artificial panel based on (cohort) averages of repeated cross-sections.
In Dargay and Vythoulkas [26] the dependent variable is the number of cars in the household; the variable now indicates the average number of cars for that particular cohort. The explanatory variables are socio-economics characteristics of the household; income, the number of adults, the number of children, metropolitan and rural areas and the generation effect of the household. Price indices for car purchase costs, car running costs and public transport fares are added to the set of explanatory variables. In their real case study the authors found that long term elasticities are almost 3 times as large as the short term elasticities, which indicates a lot of dynamics in car ownership. One important limitation of pseudo-panel is that averaging over cohorts transforms (discrete) values of variables into cohort means, thereby losing information about the individual. An exhaustive model of the demand for personal-use cars and light trucks is described in Train [27]. The model consists of a system of sub-models that separately describe the number of vehicles owned, the class and vintage of each vehicle and the miles traveled in each vehicle. The model was estimated on the sample of households that constitute the National Transportation Survey (NTS) augmented with data on the characteristics of makes and models. Consumer-level models of vehicle choice have been developed during the past two decades to explain the erosion of US automobile manufacturers’ market to the advantage of Japanese and European manufacturers. (Mannering and Winston [28]; Train and Winston [29]) Dynamic in car ownership choice, both at intertemporal dimension (resistance to change in ownership levels due to uncertainty of financial position) and intra-temporal dimensions (acquired taste for a certain lifestyle) have often been estimated on the Dutch National Mobility Panel. (Kitamura and Bunch [30], Nobile et al. [31]) More recently (Golob et al. [32]) household vehicle usage has been studied to forecast future vehicle emissions, specifically including the potential gains from alternative fuel vehicles; Stated Preference (SP) data regarding Electric Vehicle usage were collected on a mail-back survey in California. In a later paper, Revealed Preference (RP) and two waves of SP data have been collected in California and used to estimate alternative fuel vehicles demand. SP data remained critical for obtaining information about attributes not available in the marketplace, but pure SP models gave implausible forecasts, hence the use of joint models. The study mainly focused on heterogeneity in taste, but the difference between heterogeneity and state dependency was not discussed. (Brownstone et al. [24])
3 Summary of dynamic model in Economics

Dynamic discrete choice models have been firstly developed in economics and related fields. In dynamic discrete choice structural models, agents are forward looking and maximize expected inter-temporal payoffs; the consumers get to know the rapidly evolving nature of product attributes within a given period of time and different products are supposed to be available on the market. 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.

John Rust [33] did the pioneer work of dynamic models, formalizing the optimal stopping problem and estimating the optimal stopping time to replace a used bus engine. It is a single agent problem describing the decision of time to make one purchase over a set of products with homogenous 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. Berry, Levinsohn and Parkes [34] - 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 [35] 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. In the Melnikov’s framework, the products are heterogeneous which is different from Rust’s bus engine model while consumers are homogeneous; error terms are in fact assumed to be independently distributed across consumers, products and time periods; furthermore, the purchase is only made once in the consumers’ lifetime. In addition, the parameters of the static problem part are estimated separately from the dynamic part; the participation probability of a consumer is directly obtained from observing the number of purchases in the total market. The estimation of dynamic discrete choice models is computationally costly because the solution of the fixed point problem as defined by Rust is required on all points along the estimation algorithm. In conclusion a three-step method was used to solve the estimation problem. Szabolcs Lorincz [36] extended Melnikov’s optimal stopping problem with a persistent effect. This persistence means that cus-
tomers who already had a product may choose to upgrade it, (i.e. upgrade the operating systems). 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. These dynamic economic models are generally applied to evaluate price and elasticities, intertemporal substitution and the welfare gains from industry innovations.

In 2006, Carranza [37] examined digital cameras market and proposed a logit utility model with one time purchase and fully heterogeneous consumers. 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 [38] 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.

In economics, dynamic models are used to study a variety of problems. Retirement plans are represented by an optimal stopping problem where agents are supposed to decide whether to continue to work or to retire and collect social security pension benefits. Occupational choices and career decision problems over time are also well suited for dynamic modeling (i.e. choosing between staying at home, attending school or entering the job market).

In transportation, dynamic applications of discrete choice model mainly focus on the evolution of individuals’ previous plans and actions but not consider the changing external conditions, such as price and attributes of the targets under study. Therefore, we believe that the development of dynamic discrete choice models in transportation needs to be extended. Here we focus on possible applications of dynamic discrete choice models to car ownership
for short and medium-term planning (number of cars and type to own or to purchase) and in particular on the potential of advanced technology (rapidly changing over time) on individual preferences and market evolution.

4 Car ownership formulation

4.1 General consumer stopping problem

We consider a consumers set \( I = \{1, \ldots, M\} \), where each consumer \( i \in I \) 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:

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

‘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. The decision process halts when he/she reaches state 1. 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. 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 as answered in the survey.

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 parameters of the problem 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 \( j \in J_t \) and obtain a terminal period payoff \( u_{ijt} \),

where \( J_t = \{1, \ldots, J_t\} \) 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. $\theta_i$ and $\alpha_i$ are parameters vectors for $x_{it}$ and $q_{it}$ respectively.

It is here assumed that the choice set $J_i$ is consistent in each time period $t$, so we can drop the subscript $t$ from $J_i$ and $J_t$, and keep $J$ and $J$ respectively.

The payoff $u_{ijt}$ is expressed as a random utility function

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

where

- $x_{it}, \theta_i \in \mathbb{R}^Q$ are defined as above;
- $d_j \in \mathbb{R}^K$ is a vector of static attributes for potential choice $j$ and $\gamma_i$ is a vector of parameters related to $d_j$;
- $y_{jt} \in \mathbb{R}^H$ 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. $\lambda_i$ is a vector of parameters related to $y_{jt}$.
- $\epsilon_{ijt}$ is an individual-specific random term depending of $i$, the product $j$ and the time period $t$. We assume the random terms $\epsilon_{ijt}$ are components of $J$-dimensional random vectors $\zeta_{it} (i = 1, \ldots, M, t = 0, \ldots, T)$, which are independent and identically-distributed amongst individuals and periods, and have zero means (so for convenience, we will sometimes drop the subscripts $i$ and $t$). We also require that $\zeta_{it}$ follows the generalized extreme value (GEV) distribution, characterized by the cumulative joint distribution function $F_{\zeta}^{\epsilon_1, \ldots, \epsilon_J}$ 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 we set $\kappa = 1$;
\(iii\) $\lim_{a_j \to -\infty} G(a_1, \ldots, a_J) = \infty, \forall j = 1, \ldots, J$;
\(iv\) for any distinct sequence $(j_1, \ldots, j_k)$, $\partial^k G / \partial a_{j_1} \ldots \partial a_{j_k}$ is greater than 0 if $k$ is odd and less than 0 if $k$ is even.

\(^1\)This allows us to summarize car consumption and current fuel price into one attribute.
Besides, we can write $G_j(\cdot)$ as the first partial derivative of $G(\cdot)$ with respect to $j^{th}$ argument. We further assume that we can rewrite equation (1) with error acting in an additive way:

$$u_{ijt} = V_{ijt} + \epsilon_{ijt},$$

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

Relying on McFadden seminal paper [39], $\zeta$ follows a multivariate extreme value distribution. An example of a quite general $G$ function is

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

(2)

where $B_n \subseteq \{1, \ldots, J\}$, $\bigcup_{n=1}^{N} B_n = \{1, \ldots, J\}$, and $0 \leq \delta_n < 1$. We can therefore see each $B_n$ ($n = 1, \ldots, N$) as a nest, with possible overlappings between the nests. $\delta_n$ can be interpreted as an index of the similarity of the unobserved attributes in $B_n$. The choice probabilities for the function (2) satisfy

$$P_i = \sum_{n=1}^{N} P[i \mid B_n]P[B_n]$$

$$= \frac{\sum_{i \in B_n} e^{\frac{V_i}{1-\delta_n}} \left( \sum_{j \in B_n} e^{\frac{V_j}{1-\delta_n}} \right)^{1-\delta_n}}{\sum_{n=1}^{N} \left( \sum_{k \in B_n} e^{\frac{V_k}{1-\delta_n}} \right)^{1-\delta_n}},$$

(3)

where

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

In the special case $G(a_1, \ldots, a_J) = \sum_{j=1}^{J} a_j$, we have $F_\zeta(\epsilon_1, \ldots, \epsilon_J) = F(\epsilon_1), \ldots, F(\epsilon_J)$, so $\epsilon_j$ are all independent, and $\delta_n = 0$ ($n = 1, \ldots, N$). When all alternatives are available, probabilities (3) simplify to usual multinomial logit probabilities.
The two-step decision process is, at each period, first to choose the product $j^*$ that maximizes utility (1) from $J$; then, the consumer decides to buy or to postpone the purchase until the optimal time period $\tau$, that is the time when the consumer decides to buy instead of postponing. The consumer deciding to buy or postpone is the optimal stopping problem at time $t$:

$$D(u_{it}, \ldots, u_{i,t}, c_{it}, t) = \max_{\tau} \left\{ \sum_{k=t}^{\tau-1} \beta^{k-t} c_{it} + \beta^{\tau-t} E \left[ \max_{j \in J} u_{ij\tau} \right] \right\}$$  \hspace{1cm} (4)

where

- $\beta$ 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 $v_{it} = \max_{j \in J} u_{ijt}$. According to the previously described assumption about $\epsilon_{ijt}$, $v_{it}$ is Gumbel distributed with a scale factor equals to 1 since we assume in property (ii) that $G(a)$ is homogenous of degree one. In other terms, we have

$$F_v(z; r_{it}) = e^{-e^{-(z-r_{it})}},$$  \hspace{1cm} (5)

$$f_v(z; r_{it}) = e^{r_{it}} e^{(e^{-(z-r_{it})})},$$

$$f_v(z; r_{it}) = -e^{-(z-r_{it})} F_v(z; r_{it}),$$

where $r_{it}$ is the mode of the distribution of $v_{it}$, that is

$$r_{it} = \ln G(e^{V_{1it}}, \ldots, e^{V_{Jit}}).$$  \hspace{1cm} (6)

Later, we will write $r_{it}$ as $r_{it}(y_{it})$ 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) into:

$$D(v_{it}, c_{it}) = \max \{ v_{it}, c_{it} + \beta E[D(v_{i,t+1})] \}$$  \hspace{1cm} (7)

### 4.2 Utility formulation

The Bellman equation (7) becomes:

$$D(v_{it}, c_{it}) = \max \{ v_{it}, c_{it} + \beta E[D(v_{i,t+1}(y_{j,t+1}, c_{i,t+1})) | y_{jt}] \}$$

10
This is a standard regenerative optimal stopping problem as described in Section 4.1. The stopping set is given by:

$$\Gamma(y_{jt}) = \{ v_{it} \mid v_{it} \geq c_{it} + \beta E[D(\cdot) \mid y_{jt}] \}$$  \hspace{1cm} (8)

We define the reservation utility level $W(y_{jt})$ as

$$W(y_{jt}) = c_{it} + \beta E[D(v_{i,t+1}(y_{jt}, t+1, c_{i,t+1})) \mid y_{jt}]$$  \hspace{1cm} (9)

and consider the optimal policy

$$v_{it} \quad \text{if} \quad v_{it} \geq W(y_{jt})$$

$$W(y_{jt}) \quad \text{otherwise}.$$  

The reservation utility can be integrated as:

$$W(y_{jt}) = c_{it} + \beta \int_{\infty}^{-\infty} \int_{\mathbb{R}^d} \max(v_{it}, W(z)) dF(v|z) d\Phi(z|y_{jt}),$$  \hspace{1cm} (10)

where, from (5) and (6),

$$F(v|z) = e^{-e^{-e^{-e^{-v_{it} - r(t)}}}},$$

and

$$d\Phi(z|y_{jt}) = \phi[L^{-1}(y_{jt})(z - \mu(y_{jt}))]dz,$$

and $\phi(z|y_{jt})$ is the density of a standard multivariate normal distribution. Using (10), we can rewrite (7) as

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

There is a fixed point problem to solve here and we will refer to Rust’s nested fixed point algorithm. As presented in (8), the consumer $i$ will buy some product at time $t$ only when $v_{it} > W(y_{jt})$. The probability of postponing the purchase until the next period can therefore be written as:

$$\pi_{0it}(y_{jt}) \stackrel{def}{=} P[v_{it} \leq W(y_{jt})] = P[\text{postpone} \mid s_{it} = 0, y_{jt}]$$

$$= F_{v}(W(y_{jt}), y_{jt}) = e^{-e^{-e^{-W(y_{jt}) - r(t)}}}.$$  

The hazard rate of the product adoption is $h(y_{jt}) = P[\text{buy} \mid s_{it} = 0, y_{jt}] = 1 - \pi_{0it}(y_{jt})$, and the product-specific purchase probability is:

$$\pi_{ijt}(y_{jt}) \stackrel{def}{=} P[U_{ijt} \geq U_{ikt}, \forall k \neq j, u_{ijt} \geq W(y_{ijt})]$$

$$= P[U_{ijt} \geq W(y_{ijt}) \mid U_{ijt} \geq U_{ikt}, \forall k \neq j]P[U_{ijt} \geq U_{ikt}, k \neq j]$$

$$= P[v_{it} \geq W(y_{ijt})]P[U_{ijt} \geq U_{ikt}, k \neq j]$$

$$= h(y_{jt}) e^{V_{ijt}} G_j(e^{V_{ijt}}, \ldots, e^{V_{ijt}}) \frac{G(e^{V_{ijt}}, \ldots, e^{V_{ijt}})}{G(e^{V_{ijt}}, \ldots, e^{V_{ijt}})}.$$  \hspace{1cm} (11)
As introduced in Section 4.1, $G_j(\cdot)$ is the partial derivative of $G(\cdot)$ with respect to $j$th argument.

### 4.3 Industry evolution

As expressed in Section 4.1, $y_{jt}$ represents the evolution of the product attributes and the market environment. $y_{jt}$ here is assumed to follow a normal diffusion process:

$$y_{j,t+1} = \mu(y_{jt}) + L(y_{jt})\nu_{j,t+1},$$

where

- $\nu_{jt}$ ($j = 1, \ldots, J$, $t = 1, \ldots, T$) are i.i.d. multivariate standard normal random vectors, $\mathcal{N}(0, I)$, where $I$ denotes the identity matrix;
- $\mu(y_{jt}) : \mathbb{R}^H \to \mathbb{R}^H$ and $L(y_{jt}) : \mathbb{R}^{H \times H} \to \mathbb{R}^{H \times H}$ 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 $y_{j,t+1}$; $\Sigma(y_{jt})$ 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 $y_{j,t+1}$ therefore follows a multivariate normal distribution where $L$ is the Cholesky factor of the variance-covariance matrix $\Sigma$ and is lower triangular. The vector $\mu$ is the expected value of $y_{j,t+1}$. $y_{j,t+1}$ can for instance be specified as a random walk with drift $\eta_j$, i.e. $y_{j,t+1} = \psi_j y_{jt} + \eta_j + L\nu_{j,t+1}$. For simplicity, we then assume that $\psi_j$ and $\eta_j$ are the same over all the alternatives. Therefore, $\mu(y_{jt}) = \psi y_{jt} + \eta$ and $L(y_{jt}) = L$.

### 4.4 Objective function and parameters to estimate

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

- $\theta$, a vector of stationary consumer preference parameters related to individual attributes $x_{it}$.
- $\gamma$, a vector of parameters related to attributes for potential choice $d_j$.
- $\lambda$, a vector of parameters related to dynamic attributes of product $j$, $y_{jt}$.
• $\beta$, the discount factor.

• $\alpha$, a vector of parameters for characteristics of current owned car $q_{it}$.

Recall also that $\mu$ is the expected value of $y_{jt,t+1}$ and $L$ is the Cholesky factor of variance-covariance matrix $\Sigma$. If we assume that $y_{jt,t+1}$ follows a random walk with drift $\eta$, i.e. $y_{jt,t+1} = \psi y_{jt} + \eta + L\nu_{jt,t+1}$, and we have to know the values of $\psi$, $\eta$ and $L$. $\nu_{jt,t+1}$ follows a multivariate normal distribution, so we set $f(y_{jt})$ as the standard (multivariate) probability density. Assuming that $y_{jt,t+1}$ follows a random walk with drift $\eta$, i.e. $y_{jt} = \psi y_{jt-1} + \eta + L\nu_{jt}$, and we have to know the values of $\psi$, $\eta$ and $L$.

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

$$
\mathcal{L}(\psi, \eta, L) = \prod_{j=1}^{J} \prod_{t=1}^{T} f(\hat{y}_{jt}).
$$

Note that in a revealed stated stated preferences survey, we can use the collected $y_{jt}$, while in a stated preference survey, we need to generate a series of $y_{jt}$ following the current period $y_{jt,0}$, i.e. $y_{jt,1}, y_{jt,2}, \ldots$. The random walk assumption allows us to easily construct (independent) scenarios but recursively drawing from $y_{jt,1}, y_{jt,2}, \ldots, y_{jt}$.

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

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

(13)

The decision probability is presented as:

$$
P_{it}[\text{decision}] = P_{it}[\text{decision}, s_{it} = 0] + P_{it}[\text{decision}, s_{it} = 1]
= P_{it}[\text{decision} | s_{it} = 0]P[s_{it} = 0] + P_{it}[\text{decision} | s_{it} = 1]P[s_{it} = 1]
$$

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

• if the individual reports consider to buy, $s_{it} = 0$, therefore $P[s_{it} = 0] = 1$ and $P[s_{it} = 1] = 0$, and
  $$
P_{it}[\text{decision}] = P_{it}[\text{decision} | s_{it} = 0];
$$

• if the individual reports not consider to buy, $s_{it} = 1$, therefore $P[s_{it} = 0] = 0$ and $P[s_{it} = 1] = 1$, and
  $$
P_{it}[\text{decision}] = P_{it}[\text{decision} | s_{it} = 1].
$$
Under this last condition, the interviewee 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 (13) becomes \( L(\theta, \gamma, \lambda, \beta, \alpha) = 0 \). Thus, the whole system collapses (indicating a problem in the underlying dataset). As a result, in car ownership example, the complete likelihood function is:

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

where \( V = \{(i, t) | i \in 1, \ldots, M, t \in 1, \ldots, T \text{ and } s_{it} = 0\} \).

5 Conclusions

Dynamic models based on optimal stopping problem, stochastic growth process and technological changes are new in transportation demand analysis and particularly in car ownership modeling. Although we recognize that many of the behaviors that we model are dynamic, still the majority of the models estimated by transportation analysts are static in nature. Our current car ownership models are incapable to predict when confronted with technological improvements affecting car performances and related costs.

In this period, characterized by the concerns about increasing fuel cost we are often incapable to predict the demand for smaller and less consuming vehicles. In this paper we have proposed to develop an estimation technique for analyzing the impact of technological change on the dynamics of consumer demand for vehicles. The proposed research presents a dynamic model of consumer demand that explicitly accounts for consumers’ expectations of future vehicle quality and consumers’ outflow from the car market, arising endogenously from their purchase decisions. The timing of consumers’ purchases is formalized as an optimal stopping problem where the agent must decide on the optimal time of purchase. This model frame also allows to model the choice from a set of differentiated vehicles whose quality changes stochastically over time. Understanding the dynamics in car-ownership states and plans, and explaining factors that influence actions is the main challenge proposed in this paper. This cannot be done without introducing strong methodological innovations in discrete choice modeling and by developing new concepts and techniques adapted to solve the complexity of the econometric problem.
Car ownership is not the only possible application in transportation, discrete choices to be modeled in revenue management schemes or activity based frameworks are heavily affected by the temporal component. We would like to conclude by saying that possible impediments to the acceptance of discrete choice models in transportation are the data requirements (panel data) and the relative computational complexity in their application. A number of multi period surveys and continuous travel diaries are now available to researchers and the successful analysis conducted in recent years confirm that panel data are reliable and free from bias due to fatigue effects. The implementation of dynamic programming concepts in current transportation software will encourage the use of discrete choice models, while their success will certainly depend on the accuracy of short-medium term and long-term forecasts.

References


[27] Train K. Qualitative choice analysis: Theory, econometrics and an application to automobile demand. The MIT Press, Cambridge MA.


