ESSAYS ON DYNAMIC DISCRETE CHOICE MODELS

by

Jongyearn Lee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2012

Doctoral Committee:

Professor Daniel A. Ackerberg, Co-Chair
Assistant Professor Yoonseok Lee, Co-Chair
Professor Brian P. McCall
Assistant Professor Ying Fan

To my family

ACKNOWLEDGEMENTS

My deepest gratitude is to the members of my dissertation committee. I am particularly grateful to my co-chairs, Dr. Dan Ackerberg and Dr. Yoonseok Lee. Their patience, encouragement, supervision, and support from the preliminary to the concluding level enabled me to develop an understanding of the subject. Without their numerous advice and guidance, this humble achievement would not be accomplished. I also want to give my heart-felt gratitude to Dr. Brian McCall and Dr. Ying Fan for their insightful feedbacks and suggestions. I am heavily indebted to them for their continuous encouragement and guidance.

I would like to acknowledge Dr. Tilman Börgers, Dr. Bob Barsky, and Dr. Stephan Lauermann for numerous discussions and lectures on related topics that helped me organize my primitive idea.

Many friends and colleagues I met in Ann Arbor have been and will be an invaluable asset of my life. They have always been willing to help me with all their hearts. I will never forget delightful "couple's nights" with Sungwoo and Jinhyung that provide me vitality. Staff members at Rackham Information & Technology Services have supported me tremendously. I am particularly grateful to Maia Bergman for her permanent generosity and support.

Most importantly, none of this would have been possible without the love and patience of my family. My family to whom this dissertation is dedicated to, has been a constant source of love, concern, support and strength all these years.

"*Gamsahamnida*!"

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Essays on Dynamic Discrete Choice Models

by

Jongyearn Lee

Co-Chairs: Daniel A. Ackerberg and Yoonseok Lee

This dissertation presents three topics in dynamic discrete choices.

Chapter 1 sets up a dynamic model of demand to identify consumers preferences for "newness" of products in a new durable goods market, namely golf drivers market. Forward-looking heterogeneous consumers with preferences for newness of products decide when and what to purchase. The model also accounts for the fact that the market is highly subject to seasonal fluctuations. Using the aggregated data from the US golf drivers market the model succeeds at identifying consumers preference for newness of products when the seasonality and quality differences are controlled for. Experiments with different assumptions are performed to confirm the robustness of the model. Finally, a counterfactual analysis of a merger scenario is carried out to see the effect of consumers preferences for newness on the volume of sales.

Chapter 2 is a joint essay with Yoonseok Lee. This chapter finds that the state legislation decision on the mandatory motorcycle-helmet-use law is affected by the neighboring states' choices. It turns out that such a social interaction is one of the key factors in this decision making, whereas the fatality rate from motorcycle-related accidents is not so. Using the U.S. state level panel data, the analysis is conducted

by developing a mixed proportional hazard model with grouped data, which allows for possible cross sectional duration dependence. Though this analysis does not give an answer to a long-debated issue whether to introduce the mandatory motorcycle-helmet-use law, it explains a behavioral aspect of the legislative decision making procedure (i.e. social interactions) and empirically shows how the proximity between agents affects the decision making.

Chapter 3 deals with a dynamic panel data model where the dependent variable is latent while only its ranking among individuals is observable at each time period. It sets up a dynamic panel data model where the latent dependent variable is characterized by its ranking in the previous period and current exogenous variables along with individual heterogeneity. In order to overcome a small sample problem when the number of individuals is large and the number of time periods is small, it uses the explosion property of logit models with ranked data. As an application, this study applies the econometric model to the panel ranking data of the best states for business announced by Forbes magazine. It finds a significant relation between lagged ranking, along with selective covariates, and current business environment.

CHAPTER I

When Consumers Are Fascinated by Brand-New Models: A Case of US Golf Drivers Market

1.1 Introduction

In the market of golf drivers, consumers seem to have strong preferences for "newness" of products. Another stylized fact of the golf drivers market is that the market is highly subject to seasonal fluctuations. This chapter attempts to identify consumers' preferences for newness and measure the amount of newness premium when the seasonality and quality difference are controlled for.

Then, what is newness? The definition we use in this study is the status of being the latest model among its own brand.[1] It is distinguished from the age of a product defined as the time elapsed from the model's first inception in the market. Consumers do not prefer a product just launched last month to one released two months ago simply because the former is introduced a month later than the latter. Rather they compare all the latest models of several brands available in the market if they care about new models much. In many circumstances, the newness of a product does not necessarily mean a better quality, e.g. moving manufacturing site from US to China for cost reduction that can yield lower quality. Moreover, the best seller is

[1] Throughout the chapter, we use firm and brand interchangeably. They are considered identical but are used to represent the circumstances appropriately.

not always the best product.

Strong preference for the newness of a product can be explained by the prestige and image effect. As in Stigler and Becker (1977) and Becker and Murphy (1993) we can consider the prestige effects of consuming a new good. As Becker and Murphy suggested we think of the effect of introducing a new model in a characteristic sense. Having preferences for search and experience characteristics as described in Stigler (1961), consumers also have preferences for "introducing characteristics" and subsequently "newness characteristics" including when was the last time the brand launched a new model and how frequently a brand introduces new models. Beyond the prestige effect, to a small degree, the snob effect and the Veblen effect contribute to the strong preference for newness. Observing fairly fast drops in price over time, they play a role in explaining why consumers, or early adopters, want to purchase the just launched drivers even though they are relatively expensive, instead of waiting until the price falls sufficiently. Consumers prefer to use newer driver models because they are different from those commonly used/preferred, e.g. your golf buddies envy your new driver's exclusive look. Some consumers buy a new model because they think it serves as a means of attaining or maintaining their social status.

Golf drivers have rich taste aspects in a horizontal sense. Almost all observable characteristics are taste characteristics including bona fide taste characteristics, e.g. hitting sound, loft of head, length and stiffness of shaft, and feel of grip, as well as many other characteristics inherently having trade-offs between them, e.g. distance the driver carries a ball, accuracy, and forgiveness[2]. A difference in quality among products still exists at least over time, a challenge to an econometrician that requests to find an observable and discernable (to the econometrician) quality measure in a vertical sense.[3]

[2]Forgiveness is a rough measure of golf clubs representing how good of a shot a golfer will get when she misses. Due to the manufacturing reasons, it is well-known that the more forgiving a club is, the less the distance is obtained.

[3]Handbags and shoes of luxury brands have same features as described. Markets of automobiles

For the time frame of data, 2005-2009, head size of a driver can serve as an effective quality measure: it is observable for all models and it generally grows over time encompassing overall performance improvement and raised cost. Most importantly, it lets the consumers' preference for newness be identified in the model by absorbing all aspects of quality.[4]

In dealing with durable goods, recent literature extends models with demand-side dynamics to explore consumers' optimal timing problem. Consumers face intertemporal trade-offs: they compare the value of purchasing a product today to what it is expected later. Initiated by Melnikov (2001), a stream of literature adopt a logit specification to derive the expected utility of product choice in a simple closed form including Song and Chintagunta (2003), Gowrisankaran and Rysman (2011), Carranza (2010), Zhao (2008), and Conlon (2010). Many of studies analyze high-tech industries including digital cameras, video games, and LCD TVs. A stylized fact in these industries is a declining price path over time, which motivates a dynamic modeling of demand.

On the supply side, however, firms' dynamic pricing decisions are not fully exploited except Nair (2007), Zhao (2008), and Conlon (2010) among others. Nair models a serially correlated price process of forward-looking firms, Zhao deals with a dynamic Euler equation approach originated in Berry and Pakes (2000) to derive the optimality condition for pricing, and Conlon sets up forward-looking firms' pricing in response to demand state. However, they all rely on assumptions that eliminates the inter-temporal influences between competitors to simplify the complicated problem with strategic behaviors of firms.

---

or TVs share many facets with the golf drivers market but the quality of products is easily observable and rated both by consumers and an econometrician. A Universal Serial Bus (USB) flash drive is an example of goods with opposite traits. It has well-established grade of quality, i.e. storage capacity, has negligible prestige effect, and little preference on the newness of products. Consumers generally care about the specification of product only.

[4]See the Appendix A for a detailed explanation as to why almost all observable characteristics of a driver are viewed in a horizontal sense, and why the size of each model's head can serve as a quality measure in the time frame of dataset analyzed: 2005-2009.

The model in this study extends approaches in previous studies mentioned above. It is distinguished from them by explicitly modeling consumers' preference for newness of products which impact firms' pricing and introducing decision. Observed dynamic behavior of the market, this study models uniquely the transition of cyclical seasonality. In the model, consumers are forward-looking and heterogeneous.

The remainder of the chapter is organized as follows. Section 1.2 describes the market for golf drivers to emphasize the stylized fact of consumers' preference for newness of products. Section 1.3 presents the model for dynamic demand designed to explain the market behavior when consumers strongly appreciate the newness of products. Section 1.4 discusses how to estimate the model with simulations. Section 1.5 discusses the estimation results, experiments, and a counterfactual analysis. Finally, section 1.6 is devoted to conclusions.

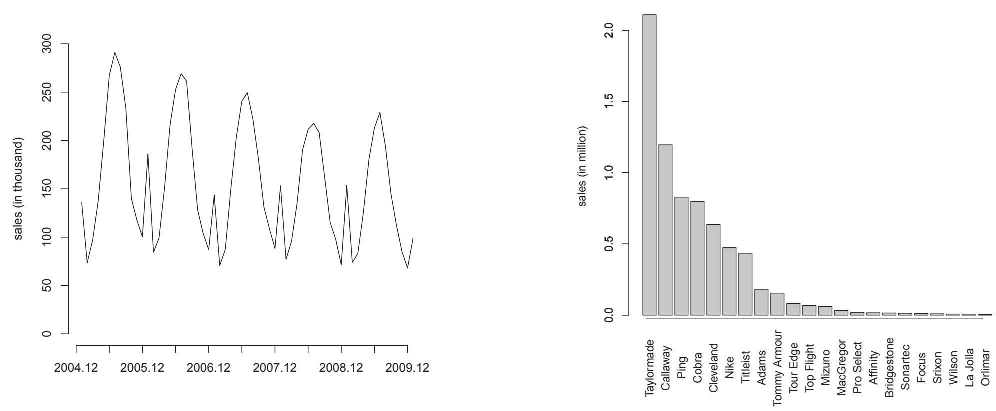
1.2 Golf Drivers Market and Preference for Newness

In this section, we discuss the observed dynamic behavior of the market and address empirical questions in regard to consumers' strong preferences for newness of products.

Figure 1.1 (a) exhibits the time path of total sales of drivers in US market from 2005 to 2009.[5] Apparently, strong seasonality in the volume of sales exists. Each summer shows high volume of sales as it is the high season for playing golf. December depicts a small but sharp peak in each year. It mainly comes from low price by big year-end sale as described below. The figure also depicts a downward long-term trend of sales throughout the five-year span, suggesting evidence of the US economy going slow toward the late-2000s. Figure 1.1 (b) is each brand's total sales ordered by

[5]Monthly model-by-model sales data are obtained from a market research company specializing in golf industry. They collect actual sales data from approximately 600 green grass pro shops and 250 off-course shops, including stores from national chains, national franchises, individual owners with multiple locations and individual owner single-unit stores.

volume of sales. Seven major brands occupy 90.5 percent of total sales in this period, where top two players take 48 percent of the total. The fact that leading brands dominate the market sales suggests that the brand-related prestige effect exits in this market. Hence we will focus on seven dominant brands when we deal with brand-specific valuation of consumers.



(a) Monthly Total Sales (b) Total Sales by Brand

Figure 1.1: Total Sales of Drivers: 2005-2009

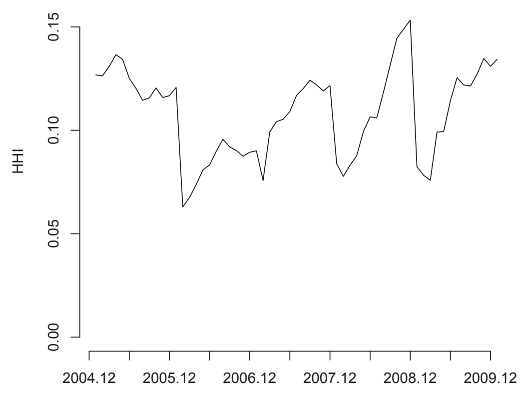


Figure 1.2: Herfindahl-Hirschman Index: 2005-2009

Figure 1.2 plots the Herfindahl-Hirschman index during the sample period to show the level of market concentration. Although it seems that the market is concentrated to a few top selling firms in Figure 1.1 (b), the US golf drivers market in this period is either unconcentrated or moderately concentrated.

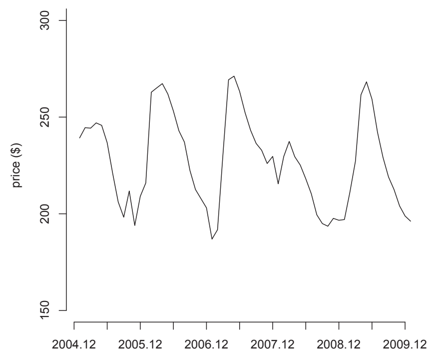


Figure 1.3: Monthly Average Price of Drivers

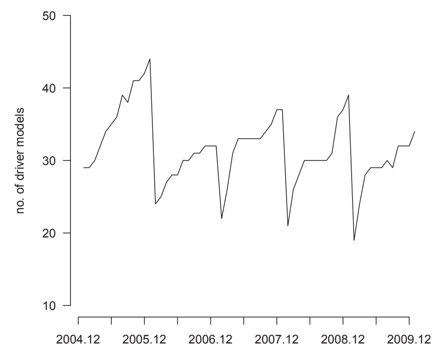


Figure 1.4: Number of Models Available in a Month

Figure 1.3 is the average price of all drivers available in the US market, weighted by their sales.[6] Generally, prices are high during the golf season (in the summer) and lower in the winter, around December in particular. The high average price in the summer is obtained because companies launch new models in the beginning of the golf season. On the other hand, the low average price in December accounts for the year-end discount and sale, which leads to high volume of sales in December. Figure 1.4 plots the number of available models in each month. The number of models typically grows within each year, reflecting the fact that new models are launched in the middle of each year.

As a caveat, truncation issues exist in the data. First, the number of models seems

[6] Average price data are collected from various sources. Due to confidentiality, annual model-by-model average point-of-sales prices and monthly total average price over all available driver models in each month are provided by the company that allowed to use sales data. Collected prices are validated to match comparably what the company provided.

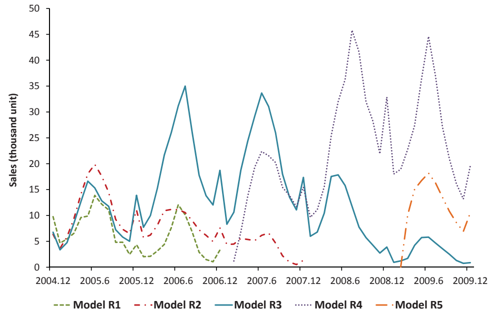


Figure 1.5: Sales of Drivers Launched by a Company

Figure 1.6: Average Price and Sales

lower than what exists in reality. It is because we only consider models identified in the data. Other models in the market are not traced since they exhibit too low sales, they are clone golf drivers, and so forth. However, we consider second-hand markets are separated from the new golf drivers market and hence clone and used drivers are not in our interest. Second, once time hits the end of each year, the number of models drops in Figure 1.4. It is because the data for some models available in a year are no longer collected in the following year, when they are not successful in particular. The feature of the market, however, curtails the effect of truncation. Every year major brands launch new models and the older models' sales drop fast in response. Moreover, the data are collected across years for the same model when they have sufficient sales. Consequently, the data show small sales values at the end tails of models on which the tracking stopped before the end of sample period. See Figure 1.5 to find the low tail values of model R1 and model R2.

Figure 1.5 is a representative time trend of sales of drivers launched by a company. Model names are ordered by their launching date. While the seasonality shown in Figure 1.1 withstands, it shows an evidence of a cannibalization effect: when a new product is introduced, sales of older models are adversely affected and sales of the latest one drops fast in particular. Putting cannibalizing behavior aside, we also observe a downward movement of sales even without the interference of introductions, i.e. the aging effect. Together with fast declining price path of each driver, the intertemporal competition between new and old models poses an interesting empirical question: do consumers have strong preferences for a driver's newness?

Another empirical question to verify is if consumers depreciate a product over time or not. Figure 1.6 displays the time paths of average price and average sales since the inception of each model. The declining price path shows a typical dynamic behavior of a new durable goods market. Data show that the price of a model drops over time with varying rate depending on how the model is appreciated in the market. Generally, the price path declines fast in response to sluggish sales in early periods after introduction, and vice versa. The time path of sales also demonstrates a typical pattern of a new durable goods market: sales increase initially and decline generally as time elapses. Observing dynamic behavior of price and sales, many of previous literature deal with a product's age, defined as the time elapsed from the product's first inception in the market, in modeling demand. The assumption behind using the age as a product characteristic is that consumers feel a product less attractive as time elapsed from its first appearance in the market, leading to the significant drop in price and sales over time. Refer to Hui (2004) and Hitsch (2006) for recent applications that are close to our model.

Finally, Table 1.1 displays the estimation results of hedonic regression. The dependent variable is the price in log scale. The explanatory variables, newness and age are as defined above. The quality measure, head size, is in cc divided by 460. The dummy for golf season has value 1 for May-September and 0 for the rest of the year. Also, the premium of seven major brands are considered. With all highly significant estimated coefficients the hedonic regression results suggest consumers' do have preferences for newness. Moreover, the results suggest that it is worth while to examine the aging effect, products' quality difference, seasonality in demand, and heterogeneous brand premium.

Table 1.1: Hedonic Estimation Results

Variable	Coefficient	Std. Err.
(constant)	4.406	0.151
newness	0.116	0.022
age	-0.040	0.001
head size	1.201	0.151
season	0.023	0.021
Taylormade	0.474	0.029
Callaway	0.279	0.029
Ping	0.155	0.038
Cobra	-0.102	0.037
Cleveland	-0.131	0.037
Nike	0.220	0.034
Titleist	0.312	0.044

Note: The dependent variable is the price ($) in log scale. Newness variable is defined as an indicator that has value 1 if a product is the latest model of its brand, and 0 otherwise. Age is the time elapsed since the model is first introduced in the market, in year scale. Head size is in cc divided by 460.

The findings from the market suggest a dynamic modeling of demand. Consumers have intertemporal trade-offs when they make a purchasing decision. Consumers seem to care about newness of products when they make a purchase, an incentive to purchase a new product promptly. On the other hand, expecting price drops in the future and introduction of a new model that can fit better her taste, each consumer has an incentive to wait until the next period instead of paying more at this period.

1.3 The Model

1.3.1 Demand

We denote each product as j and its brand as b. Let $\mathcal{J}_t^b$ and $\mathcal{B}_t$ be the set of available products of brand b at time t and available brands at time t, respectively. Denote all available products in the market, regardless of their brands, at time t as $\mathcal{J}_t$ so that $\mathcal{J}_t = \cup_{b \in \mathcal{B}_t} \mathcal{J}_t^b$. We introduce the "outside" product, an option for each consumer not to buy any of available products. Without loss of generality, denote the outside brand and product to be $b = j = 0$.

Consumers make an intertemporal choice: purchasing decision is not only what to choose but when to buy. During each period, consumers who have not purchased a product decide whether to buy one or not among those available in the market. They get the chance to make a same decision in the following period if they decide not to buy any. If they buy one, they leave the market. This assumption is reasonable for a short sample period in a durable goods market: once they buy a durable good they use it for several periods. Returning a purchased product within a term is considered as non-purchase of a product. The benefit of this assumption exceeds the loss of omitting possibilities of repeated purchases: the utility obtained from purchasing a product is maintained throughout the lifetime, and consumers' behavior other than intertemporal purchasing decision is ruled out, e.g. upgrading a driver and reselling the purchased driver in a secondary used market. Even though the possibility of upgrading a driver is ruled out, we allow golfers' skill level or taste may vary over time. For example, a golfer generally prefers a driver with sharper accuracy to one with higher forgiveness as the skill level increases.

Consumers are aware of all available products' characteristics including specifications, prices, brand names, ages, and whether they are the brand's latest model or not. Consumers compare all available models based on their own taste, i.e. golfers test drivers and choose the one that meets their needs best.

Let Ω_{it} denote the set of all state variables affecting consumer i's purchase decisions at time t. Then we let $U(\Omega_{it})$ be the value function for a consumer i at state Ω_{it} that contains all relevant information regarding purchasing and timing decision. Also, let u_{ijt} be her lifetime utility given by product j purchased at time t. A consumer who has not purchased any product faces the decision problem according to the following value function:

$$U(\Omega_{it}) = \max\left\{ \max_{j \in \mathcal{J}_t} u_{ijt}, \;\; u_{i0t} + \beta \mathbb{E}[U(\Omega_{i,t+1})|\Omega_{it}] \right\}, \tag{1.1}$$

where β is the common discount factor shared by all consumers. So the consumer who has not owned a product chooses to purchase product j if and only if both the following conditions hold: 1) the expected overall lifetime utility she would get at time t by purchasing product j is the maximum of all those from $j' \in \mathcal{J}_t$, and 2) it is higher than a reservation value, $\beta \mathbb{E}[U(\Omega_{i,t+1})|\Omega_{it}]$, plus the utility generated from buying no product denoted as u_{i0t}.

Let d_{ijt} be the choice variable of consumer i having value 1 if she chooses product j at time t, and 0 otherwise. Accordingly, if $d_{ijt} = 1$ then $d_{ij't} = 0$ for all available $j' \neq j$ including the outside good. From the setting we have in the value function (1.1), if $d_{ijt} = 1$ for $j \neq 0$, consumer i receives u_{ijt} and leaves the market at time t. If $d_{i0t} = 1$, consumer i has another chance to make a purchasing decision at time $t + 1$. We then can rewrite the value function of consumer i as follows:

$$U(\Omega_{it}) = \max_{j \in \mathcal{J}_t \cup \{0\}} L_j U(\Omega_{it}), \tag{1.2}$$

where L_j is the alternative-specific operator defined by

$$\begin{aligned} L_j U(\Omega_{it}) &= u_{ijt}, \qquad \text{for all } j = 1, \ldots, |\mathcal{J}_t|, \\ L_0 U(\Omega_{it}) &= u_{i0t} + \beta \mathbb{E}[U(\Omega_{i,t+1})|\Omega_{it}, d_{i0t} = 1]. \end{aligned} \tag{1.3}$$

It is well-known that if a solution to the problem (1.2) exists then it is unique, e.g. Rust (1994).

The unobservable (to an econometrician) characteristics of products are separated into two groups: vertical and horizontal characteristics, denoted as ξ_{jt} and ζ_{jt}, respectively. The vertical characteristics summarize quality-related characteristics of a product, e.g. durability and maintenance cost. The horizontal characteristics include all features dependent upon consumers' taste: due to consumers' diverse taste or skill level most of product characteristics falls in this category, e.g. hitting distance, forgiveness, and controllability of a driver. In a horizontal manner, each consumer has her own optimal taste on products, denoted as ζ_{it} in a same taste space that ζ_{jt} lies on. Here ζ_{it} is unobservable to an econometrician as well. Let $d(\zeta_{it}, \zeta_{jt})$ be the economic distance between consumer i's optimal taste and product j's taste characteristics. Then $\varepsilon_{ijt} := h(d(\zeta_{it}, \zeta_{jt}))$ is a decreasing function of $d(\zeta_{it}, \zeta_{jt})$ such that $h(0) = \infty$ and $h(\infty) = -\infty$. This setup allows us to transform consumers' heterogeneity in taste into a vertical measure in a Hotelling sense: each consumer pays the traveling cost which is proportional to the distance from her location, i.e. her optimal taste, to a shop, i.e. a product's characteristics.

We define the newness of product j at time t, denoted as n_{jt}, as an indicator variable equal to 1 if product j is the brand's latest model at time t and 0 otherwise. It is distinguished from a product's age: even though a model was launched long ago, it may still be the latest model made by its brand, and vice versa. We assume the lifetime utility generated by purchasing product j to be as follows:

$$\begin{aligned} &u_{ijt}(x_{jt}, n_{jt}, p_{jt}, Z_t, \xi_{jt}, \varepsilon_{ijt}; \theta^d) \\ &\qquad = \alpha_0 + x_{jt}\alpha_{xi} + \lambda_i n_{jt} + \alpha_{si} Z_t - \alpha_{pi} \log p_{jt} + \xi_{jt} + \varepsilon_{ijt}, \qquad \text{for all } i, j, t. \end{aligned} \tag{1.4}$$

Here, x_{jt} is a length-K (row) vector of observed characteristics of product j at time t, p_{jt} is the price of product j at time t, Z_t is the marketwise seasonality variable that accounts for golf season, and ε_{ijt} is a stochastic term for taste characteristics defined as above. All the parameters on demand side are summarized in θ^d, including α and λ. Verifying if $\lambda_i > 0$ is an important empirical question of this study. A positive estimate of λ_i represents the estimated amount of prestige effect that newness of a product brings to consumer i.

The way n_{jt} is set up (being independent of rivals' launching behavior in particular) resembles how consumers really choose the right driver for them. Many golfers are only interested in the latest models of all (or several) brands. They do not prefer a driver launched by brand A last month to one introduced by brand B three months ago simply because that was launched two months later. Within brand umbrella, however, consumers do take the newness into account.

Following Berry, Levinsohn, and Pakes (1995), hereafter BLP, we introduce the random utility setting:

$$\begin{aligned} & u_{ijt}(x_{jt}, n_{jt}, p_{jt}, Z_t, \xi_{jt}, \varepsilon_{ijt}; \theta^d) \\ & \quad = \alpha_0 + x_{jt}\alpha_x + \lambda n_{jt} + \alpha_s Z_t - \alpha_p \log p_{jt} + \xi_{jt} + \\ & \quad\quad \sum_{k=1}^{K} \sigma_{\alpha_x k}\nu_{\alpha_x ik} x_{jkt} + \sigma_\lambda \nu_{\lambda i} n_{jt} + \sigma_{\alpha_s}\nu_{\alpha_s i} Z_t - \sigma_{\alpha_p}\nu_{\alpha_p i}\log p_{jt} + \varepsilon_{ijt}, \end{aligned} \tag{1.5}$$

such that

$$\begin{aligned} & \alpha_{xi}^k = \alpha_x^k + \sigma_{\alpha_x}^k \nu_{\alpha_x i}^k \quad \text{for all } k = 1, \ldots, K, \\ & \alpha_{si} = \alpha_s + \sigma_{\alpha_s}\nu_{\alpha_s i}, \quad \alpha_{pi} = \alpha_p + \sigma_{\alpha_p}\nu_{\alpha_p i} \quad \text{and} \quad \lambda_i = \lambda + \sigma_\lambda \nu_{\lambda i}, \end{aligned}$$

where $(\nu_{\alpha_x i1}, \ldots, \nu_{\alpha_x iK}, \nu_{\alpha_s i}, \nu_{\alpha_p i}, \nu_{\lambda i})$ are unobserved consumer heterogeneity. The utility from not buying any of the products, u_{i0t}, is similarly given by

$$u_{i0t} = \sigma_0 \nu_{0i} + \varepsilon_{i0t}, \qquad \text{for all } i, t. \tag{1.6}$$

We set $\sigma_0 = 0$, which is equivalent to normalizing the utility from the outside good to zero.

Let us decompose the lifetime utility into two parts: one is the same for all consumers (mean utility, δ_{jt}) and the other is consumer-specific term varying by consumers' taste (μ_{ijt}.) Then we can rewrite the random utility setup in formula (1.5) as

$$u_{ijt} = \delta_{jt}(x_{jt}, n_{jt}, Z_t, p_{jt}; \theta_1^d) + \mu_{ijt}(x_{jt}, n_{jt}, Z_t, p_{jt}; \theta_2^d) + \varepsilon_{ijt}, \tag{1.7}$$

where

$$\delta_{jt} := \alpha_0 + x_{jt}\alpha_x + \lambda n_{jt} + \alpha_s Z_t - \alpha_p \log p_{jt} + \xi_{jt}$$

and

$$\mu_{ijt} := \sum_{k=1}^{K} \sigma_{\alpha_x k}\nu_{\alpha_x ik} x_{jkt} + \sigma_\lambda \nu_{\lambda i} n_{jt} + \sigma_{\alpha_s}\nu_{\alpha_s i} Z_t - \sigma_{\alpha_p}\nu_{\alpha_p i}\log p_{jt}.$$

Accordingly, the parameters in demand side θ^d are also decomposed into two parts: the product-specific demand parameters along with seasonality, $\theta_1^d = (\alpha, \lambda)$, and the consumer-specific ones, $\theta_2^d = (\sigma_\alpha, \sigma_\lambda)$.

To separate out the effect of the seasonality, we consider another decomposition of life time utility.

$$u_{ijt} = \delta_{jt}^0(x_{jt}, n_{jt}, p_{jt}; \theta_1^d) + \mu_{ijt}^0(x_{jt}, n_{jt}, p_{jt}; \theta_2^d) + (\alpha_s + \sigma_{\alpha_s}\nu_{\alpha_s i})Z_t + \varepsilon_{ijt}, \tag{1.8}$$

where

$$\delta_{jt}^0 := \delta_{jt} - \alpha_s Z_t \qquad \text{and} \qquad \mu_{ijt}^0 := \mu_{ijt} - \sigma_{\alpha_s}\nu_{\alpha_s i} Z_t.$$

Notice that the seasonality affects all available products equally.

Solving the general dynamic programming problem (1.2) is very difficult. It is almost impossible to solve the transition probability of the state space Ω_{it} precisely if its dimension is big. We therefore assume the followings, to specify the utility in a computationally tractable way:

Assumption I.1 (transformation of uniform taste shocks)**.** *Assume that* ζ_{it} *and* ζ_{jt}

are independent and are uniformly distributed on interval (0,1). Let h *be a monotone continuous function on (0,1) such that* $h(x) = -\log(-\log(1-x)^2)$.

Assumption I.1 gives that ε_{ijt} are distributed *iid* according to Type I extreme value distribution over i, j.[7] Along with assuming ε_{i0t} follows same distribution as ε_{ijt}, it is well-known that the difference of two ε_{ijt}'s follows the logistic distribution: refer to Anderson, de Palma, and Thisse (1992) pp. 39-40. The logit specification renders the probability that a consumer will purchase any product, or participate in the market, does not depend on which product will be purchased. Rather, it is only determined by the log-sum term also known as the "inclusive value":

$$\begin{aligned} r_{it} &:= \log \sum_{j \in \mathcal{J}_t} \exp(\delta_{jt} + \mu_{ijt}) \\ &= \alpha_{si} Z_t + \log \sum_{j \in \mathcal{J}_t} \exp(\delta_{jt}^0 + \mu_{ijt}^0) \\ &= \alpha_{si} Z_t + r_{it}^0, \end{aligned} \tag{1.9}$$

[7] The sketch of proof is as follows: Let $\Xi_{(1)} := \min\{\zeta_{it}, \zeta_{jt}\}$ and $\Xi_{(2)} := \max\{\zeta_{it}, \zeta_{jt}\}$ where ζ_{it} and ζ_{jt} are independent and uniformly distributed in (0,1). Also, let the distance between ζ_{it} and ζ_{jt} be $D := |\zeta_{it} - \zeta_{jt}| = \Xi_{(2)} - \Xi_{(1)}$. Then the density for joint order statistic is obtained as

$$f_{\Xi_{(1)}, \Xi_{(2)}}(x_1, x_2) = 2.$$

See Casella and Berger (2001) pp.233-234 for the proof of above joint density. To find the probability of having $\Xi_{(1)}$ and $\Xi_{(2)}$ within some interval d, we need to integrate over all (permissible) starting positions of x_1. The density for D is then given by

$$f_D(d) = \int_0^{1-d} f_{\Xi_{(1)}, \Xi_{(2)}}(x_1, x_1 + d) dx_1 = 2(1-d).$$

Since $d = 1 - \exp\left(-\frac{1}{2}\exp(-x)\right)$ where $x = h(d)$, we obtain the density for ε_{ijt} using the change of variable technique,

$$\begin{aligned} f(x) &= f_D(h^{-1}(d)) \left| \frac{1}{h'(h^{-1}(d))} \right| \\ &= 2\exp\left(-\frac{1}{2}\exp(-x)\right) \left| \frac{1}{2}\exp(-x)\exp\left(-\frac{1}{2}\exp(-x)\right) \right| \\ &= \exp(-x)\exp\left(-\exp(-x)\right) \end{aligned}$$

which is the density of standard Type I extreme value distribution.

where $r_{it}^0 := \log \sum_{j \in \mathcal{J}_t} \exp(\delta_{jt}^0 + \mu_{ijt}^0)$.

As discussed in Rust (1994), we assume the well-known conditional independence to make the transition probability of each consumer's state space, $\mathbb{P}(\Omega_{i,t+1}|\Omega_{it})$, computationally tractable by reducing the dimension of Ω_{it}:

Assumption I.2 (conditional independence)**.** *Assume that the demand-side state space is partitioned into observable and unobservable components, $\Omega_{it} = (r_{it}, \epsilon_{it})$, and the unobserved state variable is time specific and does not affect future states.*

In Assumption I.2, ϵ_{it} denotes the state variable observed by consumer i but unobserved by an econometrician. We then have

$$\mathbb{P}(\Omega_{i,t+1}|\Omega_{it}) = \mathbb{P}(r_{i,t+1}, \epsilon_{i,t+1}|r_{it}, \epsilon_{it}) = \mathbb{P}(r_{i,t+1}|r_{it}) \cdot \mathbb{P}(\epsilon_{i,t+1}). \tag{1.10}$$

Song and Chintagunta (2003), for example, rely on the same assumption in dealing with new product adoption of heterogeneous and forward-looking consumers. Using the relation between the (overall) inclusive value, r_{it}, and the inclusive value net of seasonality, r_{it}^0, in (1.9), we can rewrite the transition probability (1.10) as

$$\begin{aligned} \mathbb{P}(\Omega_{i,t+1}|\Omega_{it}) &= \mathbb{P}(r_{it+1}^0, Z_{t+1}|r_{it}^0, Z_t) \cdot \mathbb{P}(\epsilon_{it+1}) \\ &= \mathbb{P}(r_{it+1}^0|r_{it}^0) \cdot \mathbb{P}(Z_{t+1}|Z_t) \cdot \mathbb{P}(\epsilon_{it+1}). \end{aligned} \tag{1.11}$$

We assume that the transition of seasonality $\mathbb{P}(Z_{t+1}|Z_t)$ is given deterministically to all consumers. Then what remains to understand is the transition of the observable inclusive value net of seasonality r_{it}^0 over time. To facilitate the computation, we make an assumption on the transition of inclusive value proposed by Melnikov (2001):

Assumption I.3 (Markov property)**.** *The inclusive value net of seasonality r_{it}^0 follows a 1st-order Markov process.*

Assumption I.3 makes r_{it}^0 the sufficient statistic for the distribution of $r_{i,t+1}^0$: the distribution of future inclusive value $r_{i,t+1}^0$ depends only on the current value r_{it}^0 and does not depend on any past values of r_{is}^0 for all $s < t$. Assumption I.3 is rationalized when many products are available in the market. When there are sufficiently large number of products in the market, the effect of an individual firm's pricing and introducing decision on the inclusive value is negligible. Many previous studies in dynamic demand models rely on this type of assumption in reducing computational burden: see Hendel and Nevo (2006), Carranza (2010), Zhao (2008), Gowrisankaran and Rysman (2011), and Conlon (2010) for applications.

Under Assumptions I.2 and I.3, an active consumer who is still in the market does not have to keep track of all past behaviors in the market. Rather, she makes a purchasing decision based on the realized current inclusive value since $\Omega_{it} \equiv r_{it}$. (see Aguirregabiria and Mira, 2010, for details.)[8] The unobservable state variable ϵ_{it} that does not contribute to the transition of state can be handled in a time specific manner. Denote the expected continuation value as

$$\begin{aligned} \mathbb{U}_{i,t+1}^0(\Omega_{it}) &:= \mathbb{E}[U(\Omega_{i,t+1})|\Omega_{it}, d_{i0t} = 1] \\ &= \mathbb{E}[U(r_{i,t+1})|r_{it}, d_{i0t} = 1] = \mathbb{U}_{i,t+1}^0(r_{it}). \end{aligned} \tag{1.12}$$

By virtue of the extreme value specification, we have the solution to the dynamic

[8] A more realistic assumption would be consumers update their beliefs on the probability of introduction and product characteristics of future models in a Bayesian manner, e.g. Jiang, Manchandab, and Rossi (2009). Bayesian approach, however, is not within the scope of this study. Furthermore, Bayesian updating may not be reliable with fairly short sample periods.

programming problem (1.2) in a closed form as follows:

$$
\begin{aligned}
&\mathbb{U}^0_{i,t+1}(r_{it}) \\
&\quad = \mathbb{E}[U(r_{i,t+1})|r_{it}, d_{i0t} = 1] \\
&\quad = \mathbb{E}\left[\left.\max_{j\in\mathcal{J}_t\cup\{0\}} L_j U(r_{i,t+1})\right| r_{it}, d_{i0t} = 1\right] \\
&\quad = \sum_{j\in\mathcal{J}_t} \left(\delta_{j,t+1} + \mu_{ij,t+1} + \mathbb{E}[\varepsilon_{ij,t+1}|r_{it}, d_{i0t} = 1]\right) \mathbb{P}(d_{ij,t+1} = 1|r_{it}, d_{i0t} = 1)
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \left(\beta\mathbb{E}[U(r_{i,t+2})|r_{it}, d_{i0,t+1} = 1] + \mathbb{E}[\varepsilon_{i0,t+1}|r_{it}, d_{i0t} = 1]\right) \mathbb{P}(d_{i0,t+1} = 1|r_{it}, d_{i0t} = 1) \\
&= \sum_{j\in\mathcal{J}_t} \left(\delta_{j,t+1} + \mu_{ij,t+1} + \gamma - \log\mathbb{P}(d_{ij,t+1} = 1|r_{it}, d_{i0t} = 1)\right) \mathbb{P}(d_{ij,t+1} = 1|r_{it}, d_{i0t} = 1) \\
&\quad + \left(\beta\mathbb{E}[U(r_{i,t+2})|r_{it}, d_{i0,t+1} = 1] + \gamma - \log\mathbb{P}(d_{i0,t+1} = 1|r_{it}, d_{i0t} = 1)\right) \\
&\qquad \times \mathbb{P}(d_{i0,t+1} = 1|r_{it}, d_{i0t} = 1) \\
&= \gamma + \sum_{j\in\mathcal{J}_t} \left(\delta_{j,t+1} + \mu_{ij,t+1} - \log\mathbb{P}(d_{ij,t+1} = 1|r_{it}, d_{i0t} = 1)\right) \mathbb{P}(d_{ij,t+1} = 1|r_{it}, d_{i0t} = 1) \\
&\quad + \left(\beta\mathbb{E}[U(r_{i,t+2})|r_{it}, d_{i0,t+1} = 1] - \log\mathbb{P}(d_{i0,t+1} = 1|r_{it}, d_{i0t} = 1)\right) \\
&\qquad \times \mathbb{P}(d_{i0,t+1} = 1|r_{it}, d_{i0t} = 1)
\end{aligned} \tag{1.13}
$$

where $\gamma \simeq 0.5772$ is Euler's constant, see for example Eckstein and Wolpin (1989) and Ali (2008) for related discussions. The transition probability for $j \neq 0$ is

$$
\begin{aligned}
\mathbb{P}(d_{ij,t+1} = 1|\Omega_{it}, d_{i0t} = 1) &= \mathbb{P}(d_{ij,t+1} = 1|r_{it}, d_{i0t} = 1) \\
&= \left(\frac{\exp(\widetilde{r}_{i,t+1})}{\exp(\widetilde{r}_{i,t+1}) + \exp(\beta\mathbb{U}^0_{i,t+2})}\right)\left(\frac{\exp(\delta_{j,t+1} + \mu_{ij,t+1})}{\exp(\widetilde{r}_{i,t+1})}\right) \\
&= \frac{\exp(\delta_{j,t+1} + \mu_{ij,t+1})}{\exp(\widetilde{r}_{i,t+1}) + \exp(\beta\mathbb{U}^0_{i,t+2})}
\end{aligned} \tag{1.14}
$$

whereas the transition probability for the outside good is

$$\mathbb{P}(d_{i0,t+1} = 1|\Omega_{it}, d_{i0t} = 1) = \frac{\exp(\beta \mathbb{U}^0_{i,t+2})}{\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2})}, \tag{1.15}$$

where $\widetilde{r}_{i,t+1}$ is the predicted inclusive value of consumer i after she observed the realized inclusive value at time t, r_{it}. The realized $r_{i,t+1}$ and the conjectured $\widetilde{r}_{i,t+1}$ are not necessarily equal. In particular, they are different when new models are introduced and/or old models are discontinued at time $t+1$, i.e.

$$\begin{aligned}
\widetilde{r}_{i,t+1} &= \log\Bigg(\sum_{j \in \mathcal{J}_t \cap \mathcal{J}_{t+1}} \exp(\delta_{j,t+1} + \mu_{ij,t+1}) + \sum_{j \in \mathcal{J}_t \setminus \mathcal{J}_{t+1}} \exp(\delta_{j,t+1} + \mu_{ij,t+1}) \Bigg) \\
&\neq \log\Bigg(\sum_{j \in \mathcal{J}_t \cap \mathcal{J}_{t+1}} \exp(\delta_{j,t+1} + \mu_{ij,t+1}) + \sum_{j \in \mathcal{J}_{t+1} \setminus \mathcal{J}_t} \exp(\delta_{j,t+1} + \mu_{ij,t+1}) \Bigg) = r_{i,t+1},
\end{aligned}$$

if $\mathcal{J}_t \setminus \mathcal{J}_{t+1} \neq \emptyset$ and/or $\mathcal{J}_{t+1} \setminus \mathcal{J}_t \neq \emptyset$.

Then we have

$$\begin{aligned}
\mathbb{U}^0_{i,t+1} &= \gamma + \sum_{j \in \mathcal{J}_t} \left(\log\left[\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2}) \right] \right) \frac{\exp(\delta_{j,t+1} + \mu_{ij,t+1})}{\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2})} \\
&\quad + \left(\log\left[\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2}) \right] \right) \frac{\exp(\beta \mathbb{U}^0_{i,t+2})}{\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2})} \\
&= \gamma + \frac{\log\left[\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2}) \right]}{\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2})} \left(\sum_{j \in \mathcal{J}_t} \exp(\delta_{j,t+1} + \mu_{ij,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2}) \right) \\
&= \gamma + \frac{\log\left[\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2}) \right]}{\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2})} \left(\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2}) \right) \\
&= \gamma + \log\left[\exp(\widetilde{r}_{i,t+1}) + \exp(\beta \mathbb{U}^0_{i,t+2}) \right].
\end{aligned} \tag{1.16}$$

From the specification we made, the market share of product j at time t is obtained by aggregating the individual probability of choices such that

$$\begin{aligned} s_{jt} &= \int \mathbb{P}(d_{ijt} = 1 | r_t) dG_i^{\nu}(\nu) \\ &= \int \left(\frac{\exp(\delta_{jt} + \mu_{ijt})}{\exp(r_{it}) + \exp(\beta \mathbb{U}_{i,t+1}^0)} \right) dG_i^{\nu}(\nu), \end{aligned} \tag{1.17}$$

where G_i^{ν} denotes the distribution of ν_i. Accordingly, the demand for product j at time t is obtained by

$$q_{jt}\big(x_{jt}, \xi_{jt}, p_{jt}; \theta^d\big) = M_t s_{jt} = M_t \int \mathbb{P}(d_{ijt} = 1 | r_t) dG_i^{\nu}(\nu), \tag{1.18}$$

where M_t is the total number of active consumers in the market who have not purchased a product. It is obtained by taking the exogenous number of potential consumers, namely M_t^0, and subtracting those who have purchased at least a product in all previous periods.

1.3.2 Remarks on supply

We do not present the model for firms in this study. However, a few remarks are in order. First, firms' decisions relevant to our demand model includes pricing and introduction decision. Without firms' explicit introduction decisions, the model would not be able to perform a counterfactual analysis in view of the pace of model changes. Second, strategic behavior of firms has to be incorporated in the model. In making pricing and introducing decisions firms take rivals' decisions into account. In order to avoid complexity in dynamic modeling, previous literature rely on rather controversial assumptions that result in static and/or monopolistic pricing. Incorporating firms' heterogeneity and forward-looking behavior, assumptions in recent literature are still limited to depend solely on demand states, e.g. Zhao (2008) and Conlon (2010). Third, the challenges in modeling firms' dynamic behavior also includes the unknown cost structure. In most cases, cost information is not available to researchers and it is to be recovered. Dealing with both pricing and introduction decisions, the supply model should explain both marginal and introduction costs. Fourth, the choice of product characteristics is also an important issue in regard to firms' introduction decision.

1.4 Estimation

1.4.1 Seasonality

Let m_t be an indicator variable that has value 1 if t is a month in golf season, and 0 otherwise. In this study, the golf season is set from May to September. The seasonality variable is then the present discounted value of all future m_t's since u_{ijt} is the lifetime utility. For example, the present discounted value of all future m_t's in May is obtained as

$$\begin{aligned}
Z_t &= \sum_{\tau=t}^{\infty} \beta^{\tau-t} m_\tau \\
&= m_t + \beta m_{t+1} + \beta^2 m_{t+2} + \beta^3 m_{t+3} + \dots \\
&= (1+\beta+\beta^2+\beta^3+\beta^4) + (\beta^{12}+\beta^{13}+\beta^{14}+\beta^{15}+\beta^{16}) + \dots \\
&= (1+\beta+\beta^2+\beta^3+\beta^4) + \beta^{12}(1+\beta+\beta^2+\beta^3+\beta^4) + \dots \\
&= (1+\beta+\beta^2+\beta^3+\beta^4)(1+\beta^{12}+(\beta^{12})^2+\dots) \\
&= \frac{1+\beta+\beta^2+\beta^3+\beta^4}{1-\beta^{12}} =: B \qquad (1.19)
\end{aligned}$$

Similarly, one can easily calculate Z_t for all other months and the results are shown in Table 1.2. The months in bold represent the golf season.

Of course, we can think of another setup that depicts a small peak in each December in Figure 1.1 (a). However, the baseline model assumes the above to demonstrate

Table 1.2: Present Discounted Value of Seasonality

month	Z_t	month	Z_t	month	Z_t
January	$\beta^4 B$	**May**	B	**September**	$1+\beta^8 B$
February	$\beta^3 B$	**June**	$(1+\beta+\beta^2+\beta^3)+\beta^{11}B$	October	$\beta^7 B$
March	$\beta^2 B$	**July**	$(1+\beta+\beta^2)+\beta^{10}B$	November	$\beta^6 B$
April	βB	**August**	$(1+\beta)+\beta^9 B$	December	$\beta^5 B$

Note: Months in bold indicate the golf season. Z_t is the marketwise seasonality and β is the monthly discount factor same across all consumers. B is solely dependent to β. See equation (1.19) for its derivation.

the effect of golf season on seasonality. An experiment is performed in Section 1.5 that accommodates the year-end high demand to the seasonality.

1.4.2 Markov process of r_{it}^0

Under Assumption I.3, we introduce the AR(1) model, i.e. a Markov process, to specify the transition of the inclusive value net of seasonality as follows:

$$r_{i,t+1}^0 = \eta_{0i} + \eta_{1i} r_{it}^0 + \epsilon_{it}, \qquad \text{where } \epsilon \sim \mathcal{N}(0, \sigma_r^2). \tag{1.20}$$

Then as $t \to t+1$ the seasonality value is updated while $\widetilde{r}_{i,t+1}^0 = \mathbb{E}_{it} r_{i,t+1}^0$ is obtained from (1.20), i.e.

$$\widetilde{r}_{i,t+1} = \eta_{0i} + \eta_{1i} r_{it}^0 + (Z_{t+1} - Z_t)(\alpha_s + \sigma_s \nu_{is}). \tag{1.21}$$

1.4.3 Triple layers of estimation loops

To make the integration in (1.17) tractable, we take the simulation approach: we draw $(\nu_i^1, \ldots, \nu_i^K, \nu_{\lambda i}, \nu_{\alpha i})$, $i = 1, \ldots, N$ from the standard multivariate normal distribution. The approximated market share is then given by

$$\widetilde{s}_{jt}\big(x_{jt}, \xi_{jt}, Z_t, p_{jt}\big) = \frac{1}{N}\sum_{i=1}^{N} \varphi_{it}\left(\frac{\exp\left(\delta_{jt}+\mu_{ijt}\right)}{\exp(r_{it})+\exp\left(\beta\mathbb{U}^0_{i,t+1}\right)}\right) \tag{1.22}$$

where φ_{it} is the "density" of consumer i at time t. At $t = 1$, $\varphi_{i1} = 1$ and for $t > 1$ each consumer leaves the market with probability $\left(\frac{1}{1+\exp(\beta\mathbb{U}^0_{i,t+1}-r_{it})}\right)$ and thus $\varphi_{it} = \varphi_{i,t-1}\left(\frac{\exp(\beta\mathbb{U}^0_{i,t+1}-r_{it})}{1+\exp(\beta\mathbb{U}^0_{i,t+1}-r_{it})}\right)$.

For estimating demand, we adopt an algorithm similar to that used in BLP. The difference to BLP is that the demand equation (1.18) accounts for a nonconstant continuation values and a time-varying distribution of consumers' characteristics. The estimation algorithm has three loops to converge. 1) Given a pre-specified θ^d, the continuation value at steady state $\mathbb{U}^0_{i\infty}$ is obtained. Then all previous continuation values, $\mathbb{U}^0_{i,t+1}$, $t = 1, \ldots, T$, are successively obtained from (1.16) and (1.21). 2) Using the obtained continuation values $\mathbb{U}^0_{i,t+1}$, $t = 1, \ldots, T$, we calculate the simulated market share and match it with observed market share. 3) Finally, a new set of values of θ^d is searched. These three loops are run until they all converge. Below is the detailed descriptions of each estimation loop.

[***Outer-loop***] Since we do not have information beyond the sample period, we assume that $\mathbb{U}^0_{i,T+1} = \mathbb{U}^0_{i,T+2} = \ldots = \mathbb{U}^0_{i\infty}$. Given a pre-specified θ^d, we can obtain the stationary continuation value using

$$\begin{aligned}\mathbb{U}^0_{i\infty} &= \gamma + \log\left[\exp(\widetilde{r}_{i,T+1}) + \exp(\beta\mathbb{U}^0_{i\infty})\right]\\ &= \gamma + \log\left[\exp(\eta_{0i} + \eta_{1i}r^0_{iT} + (Z_{T+1}-Z_T)(\alpha_s + \sigma_s\nu_{is})) + \exp(\beta\mathbb{U}^0_{i\infty})\right].\end{aligned} \tag{1.23}$$

Using (1.16) and (1.21), all previous continuation values are successively calculated from $\mathbb{U}^0_{i,T+1} = \mathbb{U}^0_{i\infty}$. With the pre-specified θ^d and obtained $\mathbb{U}^0_{i,t+1}$ for $t = 1, \ldots, T$, the simulated market share (1.22) is calculated.

[***Middle-loop***] Given the predicted (or simulated) market share (1.22) with the pre-specified values of the demand parameters θ^d, each of implied mean utility level δ_{jt} is numerically obtained using the fixed point algorithm proposed in BLP. At each iteration *(iter)*, the value of the mean utility at time t $\delta_{jt}^{(iter)}$ is updated by

$$\delta_{jt}^{(iter+1)} = \delta_{jt}^{(iter)} + \log s_{jt}^{obs} - \log \widetilde{s}_{jt}(\delta_{jt}^{(iter)}; \theta^d) \tag{1.24}$$

where s_{jt}^{obs} is the observed market share whereas $\widetilde{s}_{jt}$ is the simulated share as specified in the equation (1.22). Under a certain regularity condition, the algorithm guarantees a unique solution by the contraction mapping theorem. Carranza (2010) shows $0 \leq \partial\beta\mathbb{U}_{it}^0/\partial r_{it} < 1$ is a sufficient condition for the regularity conditions to guarantee the algorithm has a unique interior fixed point.

[***Inner-loop***] After the middle-loop converges, the new values of parameters θ^d are searched by matching the new simulated share value $\widetilde{s}_{jt}(\delta_{jt}^{(iter+1)}; \theta^d)$ with the observed market share s_{jt}^0. In this step, efficient and consistent estimates are obtained by the two-step GMM method. To control the endogeneity of price, the product-specific demand parameters θ_1^d is estimated by 2SLS using adequate instruments including the average age of rivals' models and the average head size of them. The iterations are repeated until all three loops converge.

1.5 Estimation Results and Discussion

1.5.1 Baseline model

Monthly sales and average price data are collected model-by-model for 61 time periods: December 2004 to December 2009. Prices are in dollar unit and are deflated by the December 2009 Consumer Price Index (CPI) value. Total of 103 driver models across 22 brands are included to constitute total of 1,922 data points. The observed product characteristics are set to be $x_{jt} = (a_{jt}, hsize_j, DB_j)$, where a_{jt} is the age of driver j at time t in year scale, $hsize_j$ is the head size of driver j in cc divided

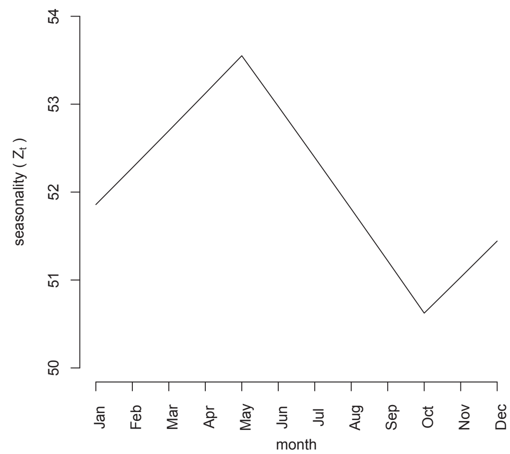


Figure 1.7: Seasonality in Each Month (Z_t)

Note: Z_t is the marketwise seasonality and the monthly discount factor β is set at 0.992.

by 460, the maximal size permitted by USGA as discussed in Section 1.2, and DB_j is a vector of brand dummies. Observed the total sales by brand shown in Figure 1.1 (b), the brand dummy vector includes top seven brands in sales: Taylormade, Callaway, Ping, Cobra, Cleveland, Nike, and Titleist. The indicator for newness n_{jt} is as defined in Section 1.3. The monthly discount factor is set at $\beta = 0.992$, which yields the seasonality Z_t in each month as shown in Figure 1.7.

Table 1.3 shows the descriptive statistics of product characteristics in x_{jt}. While the head size of drivers are clustered close to 460cc, the newness, age, and price have sufficient variability. Correlation coefficient between newness and age is -0.473 which shows natural negative but not too close correlation between them.

Table 1.4 displays the estimation results of the baseline model. The simulation of market share is performed by 10,000 random draws. In the mean utility portion of θ_1^d, the estimate for the newness (λ) is of ours special interest and is expected to

Table 1.3: Descriptive Statistics of Mean Utility Shifters

	Mean	Std. Dev.	Observations
newness (n_{jt})	0.463	0.499	1,922
age (a_{jt})	1.237	0.896	1,922
head size ($hsize_j$)	0.960	0.066	1,922
price (p_{jt})	208.76	118.95	1,922

Note: Newness variable is defined as an indicator that has value 1 if a product is the latest model of its brand, and 0 otherwise. Age is the time elapsed since the model is first introduced in the market, in year scale. Head size is in cc divided by 460. Price is in dollar unit.

be significantly positive if consumers have strong preference for newness, i.e. they are willing to pay more for the latest model over the same brand's outdated ones. We expect the estimate of parameter for age (α_a) to be negative if consumers in fact depreciate a model over time. Also, the parameter for head size (α_h) is anticipated to have a positive estimate when the consumers do observe the bigger driver head as a higher quality. Each estimate of brand dummy parameters (α_b) is expected to have positive values if consumers care and are willing to pay more for the drivers made by a major brand assigned to the dummy. In general they would have some order comparable to the ranking of aggregated sales shown in Figure 1.1 (b) but not necessarily. Finally, the estimate of price parameter (α_p) has to be positive as the way it is formulated.

The signs of $\widehat{\lambda}$, $\widehat{\alpha}_a$, $\widehat{\alpha}_h$, and $\widehat{\alpha}_p$ are obtained highly significantly as expected. Most importantly, the estimate $\widehat{\lambda}$ measures the amount of prestige effect effectively. The estimated $\widehat{\alpha}_b$'s are overall ordered by the ranking in total sales except Titleist with a comparable value to Callaway. It suggests that the top four brands in the US drivers market are Taylormade, Ping, Callaway, and Titleist in the sense of brand premium that consumers recognize. Also, the second tier group consists of Cobra, Cleveland, and Nike in the same sense.

The demand parameters of consumer heterogeneity, θ_2^d, are precisely estimated for

Table 1.4: Estimation Results: Baseline Model

Variable		Baseline Model Estimate	Std. Err.
Mean Utility (δ)			
(constant)	α_0	6.945	1.677
newness	λ	0.327	0.064
age	α_a	-0.825	0.153
head size	α_h	1.065	0.528
seasonality	α_s	1.452	0.026
Taylormade	$\alpha_b^{taylormade}$	3.177	0.164
Callaway	$\alpha_b^{callaway}$	2.275	0.112
Ping	α_b^{ping}	2.377	0.105
Cobra	α_b^{cobra}	1.682	0.095
Cleveland	$\alpha_b^{cleveland}$	1.641	0.097
Nike	α_b^{nike}	1.630	0.107
Titleist	$\alpha_b^{titleist}$	2.265	0.144
price	α_p	1.450	0.311
Consumer Heterogeneity (μ)			
newness	σ_λ	0.894	0.243
age	σ_a	0.029	0.010
head size	σ_h	0.092	0.043
seasonality	σ_s	0.299	0.127
price	σ_p	0.296	0.036

Note: Mean utility δ_{jt} is same across all consumers. All parameters are expected to have positive signs except age. Consumer heterogeneity is individual-specific variations on variables in the mean utility except brand dummies.

drivers' age, head size, newness, and price. The heterogeneity for brand dummies σ_b are not included in estimation. The estimation results with σ_b are obtained almost identical to those in Table 1.4 while only $\widehat{\sigma}_b$ are all small and insignificant at 5% level, meaning no significant evidence can be found that consumers' heterogeneous perception on brand premium is diverse.

Table 1.5 exhibits the calculated dollar values of product characteristics and brand dummies for a representative consumer, i.e. consumers' heterogeneity is ignored. Each dollar value is the percentage change in price that makes consumers remain indifferent before and after a certain change in a variable when other things are equal.

Table 1.5: Dollar Values of Characteristics: Baseline Model

Variable		Change in Variable	% Changes in Price
Product Characteristics			
newness	λ	$n_{jt} : 0 \to 1$	0.2530
age	α_a	$a_{jt} : a \to a + \frac{1}{12}$	-0.0486
head size	α_h	$hsize_j : h \to h + \frac{10}{460}$	0.0161
Brand Dummies			
Taylormade	$\alpha_b^{taylormade}$		7.9445
Callaway	$\alpha_b^{callaway}$		3.8017
Ping	α_b^{ping}		4.1516
Cobra	α_b^{cobra}		2.1899
Cleveland	$\alpha_b^{cleveland}$		2.1010
Nike	α_b^{nike}		2.0776
Titleist	$\alpha_b^{titleist}$		3.7687

Note: The numbers in the rightmost column are percent changes in price induced by the change in corresponding variables. Brand values are compared to minor brands with zero values of all brand dummies. In calculating the dollar values, a representative consumer is considered by ignoring consumers' heterogeneity.

Having newness solely induces 25.3 percent increase in price. A driver's price should fall by 4.9 percent in each month to compensate the aging effect when other things are kept equal. An increase of 10cc in head size corresponds to 1.6 percent increase in price. Among the product characteristics, the premium of newness stands out in its dollar value. The second part of Table 1.5 shows that having one of the major brand names induces price increase to a great extent. All of seven major brands reveal that more than 200 percent of increase in price compared to minor brands. It is largely because consumers do care the brand name when they choose a driver, i.e. consumers believe that brand name signals the quality.

1.5.2 Experiments

In this subsection we perform three experiments. Model I and II experiment with different consumers' behavioral assumptions, while model III adopts an approximation method where the assumptions on consumers' economic behavior remains same as the baseline model.

Model I: Myopic consumers

When consumers are assumed to behave myopically, we force $\beta = 0$ for all i and t. Then the model becomes equivalent to the static BLP model. Note that the influence of seasonality becomes dichotomous, i.e. $Z_t = m_t$. First set of parameter estimates in Table 1.7 displays the estimation results under the assumption of myopic consumers. Overall, parameters in mean utility are estimated with the expected signs. The impact of seasonality is obtained much smaller than the baseline model.

Model II: Quasi-hyperbolic discounting of seasonality with December shock

The dynamic movement of sales in Figure 1.1 (a) shows a consistent cyclical pattern: high demand in summer and small peak in December. To incorporate this pattern in the model, we make two behavioral assumptions: 1) consumers have time-inconsistent preferences, i.e. quasi-hyperbolic discounting of seasonality, and 2) consumers have high holiday demand at year-end. In other words, $m_t = 1$ when t is in December, where all other m_t values remain the same as above. The assumption of quasi-hyperbolic discounting is justified when consumers are in fact present-biased, i.e. consumers reveal strong tendency to care the current seasonality state. Let π be an additional discount factor that represents the dynamic inconsistency. The second assumption is by high holiday demand and/or companies' promotions at year-end. For example, Callaway has the "Preferred Retailer Program" that offers the year-end rebates and discounts for participating retailers, part of which in turn transfers to consumers. The present discounted value of all future m_t's in May is then obtained as

$$
\begin{aligned}
Z_t &= m_t + \pi \sum_{\tau=t+1}^{\infty} \beta^{\tau-t} m_\tau \\
&= (1 + \pi\beta + \pi\beta^2 + \pi\beta^3 + \pi\beta^4) + \pi\beta^7 + (\pi\beta^{12} + \pi\beta^{13} + \pi\beta^{14} + \pi\beta^{15} + \pi\beta^{16}) + \pi\beta^{19} \dots \\
&= (1 + \pi\beta + \pi\beta^2 + \pi\beta^3 + \pi\beta^4) + \frac{\pi\beta^{12}(1 + \beta + \beta^2 + \beta^3 + \beta^4)}{1-\beta^{12}} + \frac{\pi\beta^7}{1-\beta^{12}} \\
&= (1 + \pi\beta + \pi\beta^2 + \pi\beta^3 + \pi\beta^4) + \pi\beta^{12}B + \frac{\pi\beta^7}{1-\beta^{12}}, \qquad (1.25)
\end{aligned}
$$

where B is defined as in (1.19). Similarly, all other Z_t values are calculated in Table 1.6.

Table 1.6: The Present Discounted Value of Seasonality

month	Z_t
January	$\pi\beta^4 B + \frac{\pi\beta^{11}}{1-\beta^{12}}$
February	$\pi\beta^3 B + \frac{\pi\beta^{10}}{1-\beta^{12}}$
March	$\pi\beta^2 B + \frac{\pi\beta^{9}}{1-\beta^{12}}$
April	$\pi\beta B + \frac{\pi\beta^{8}}{1-\beta^{12}}$
May	$(1 + \pi\beta + \pi\beta^2 + \pi\beta^3 + \pi\beta^4) + \pi\beta^{12} B + \frac{\pi\beta^{7}}{1-\beta^{12}}$
June	$(1 + \pi\beta + \pi\beta^2 + \pi\beta^3) + \pi\beta^{11} B + \frac{\pi\beta^{6}}{1-\beta^{12}}$
July	$(1 + \pi\beta + \pi\beta^2) + \pi\beta^{10} B + \frac{\pi\beta^{5}}{1-\beta^{12}}$
August	$(1 + \pi\beta) + \pi\beta^{9} B + \frac{\pi\beta^{4}}{1-\beta^{12}}$
September	$1 + \pi\beta^{8} B + \frac{\pi\beta^{3}}{1-\beta^{12}}$
October	$\pi\beta^7 B + \frac{\pi\beta^{2}}{1-\beta^{12}}$
November	$\pi\beta^6 B + \frac{\pi\beta}{1-\beta^{12}}$
December	$\pi\beta^5 B + 1 + \frac{\pi\beta^{12}}{1-\beta^{12}}$

Note: Months in bold indicate the high demand: golf season and year-end. Z_t is the marketwise seasonality, β is the monthly discount factor is same across all consumers, and π is an additional discount factor which discounts future utility relative to current period utility. B is solely dependent to β.

Calibrating $\pi = 0.9$ along with $\beta = 0.992$, we have the time path of seasonality during the sample period as shown in Figure 1.8. Note that the level of Z_t is compa-

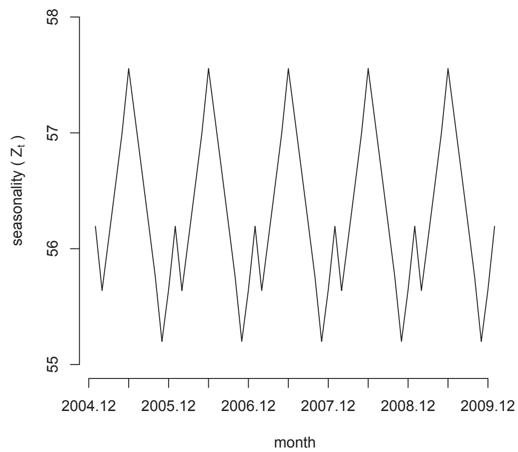


Figure 1.8: Time Path of Seasonality (Z_t)

Note: Z_t is the marketwise seasonality. The monthly discount factor β is set at 0.992 and an additional discount factor for dynamic inconsistency π is calibrated at 0.9.

rable to what we have in the baseline model shown in Figure 1.7 and the path also resembles the total sales in Figure 1.1 (a).

The second part of Table 1.7 shows the estimation results of Model II. As expected, they are much closer to the estimation results of the baseline model than Model I. Compared to the baseline model, almost all magnitudes of estimates are greater in Model II.

Model III: Approximation of continuation value a la Carranza (2010)

In this experiment, we adpot the approximation approach in obtaining the continuation value in (1.3) proposed by Carranza (2010). In order to facilitate the calculation, one can approximate the integral given by equation (1.17) in the following way: First, since we find the fact that $\mathbb{U}^0_{i,t+1}$ is solely dependent to r_{it}, we specify that

$$\beta\widetilde{\mathbb{U}}^0_{i,t+1}(r_{it};\epsilon_{it}) = \eta_0 + \eta_1\epsilon_{it} + \eta_2 r_{it} + \eta_3 r_{it}\epsilon_{it}, \tag{1.26}$$

where the unobservable component of state variables, $\epsilon_{it} \sim \mathcal{N}(0,1)$. We set $\eta_0 \equiv 0$ to identify the constant in the mean utility. Second, we replace $\mathbb{U}^0_{i,t+1}$ by $\widetilde{\mathbb{U}}^0_{i,t+1}$ in equation (1.17) and draw ϵ_l, $l = 1,\ldots,L$ from $\mathcal{N}(0,1)$. For each draw of ϵ_l, we draw $(\nu_n^1,\ldots,\nu_n^K,\nu_{\lambda n},\nu_{\alpha n})$, $n = 1,\ldots,N$ from the standard multivariate normal distribution. The approximated market share is then given by

$$\widetilde{s}_{jt}\big(x_{jt},\xi_{jt},Z_t,p_{jt}\big) = \frac{1}{L}\sum_{l=1}^{L}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi_t^{(n,l)}\left(\frac{1}{1+\exp\left(\beta\widetilde{\mathbb{U}}^{(n,l)}_{t+1} - r_t^{(n)}\right)}\right)\left(\frac{\exp\left(\delta_{jt}+\mu_{jt}^{(n)}\right)}{\exp\left(r_t^{(n)}\right)}\right)\right] \tag{1.27}$$

where

$$\begin{aligned}
\mu_{jt}^{(n)} &= \sum_{k=1}^{K}\sigma_k\nu_n^k x_{jt}^k + \sigma_\lambda\nu_{\lambda n}n_{jt} - \sigma_\alpha\nu_{\alpha n}\log p_{jt},\\
r_t^{(n)} &= \log\sum_{j\in\mathcal{J}_t}\exp\left(\delta_{jt}+\mu_{jt}^{(n)}\right),\\
\beta\widetilde{\mathbb{U}}^{(n,l)}_{t+1} &= \eta_0 + \eta_1\epsilon_l + \eta_2 r_t^{(n)} + \eta_3 r_t^{(n)}\epsilon_l,
\end{aligned}$$

and $\varphi_t^{(n,l)}$ is the "density" of consumer n at time t for the lth draw of ϵ. At $t = 1$, $\varphi_1^{(n,l)} = 1$ and for $t > 1$ each consumer leaves the market with probability $\left(\frac{1}{1+\exp(\beta\widetilde{\mathbb{U}}^{(n,l)}_{t+1}-r_t^{(n)})}\right)$ and thus $\varphi_t^{(n,l)} = \varphi_{t-1}^{(n,l)}\left(\frac{\exp(\beta\widetilde{\mathbb{U}}^{(n,l)}_{t+1}-r_t^{(n)})}{1+\exp(\widetilde{\mathbb{U}}^{(n,l)}_{t+1}-r_t^{(n)})}\right)$. The estimation strategy is same as what we have in section 1.4. Under a certain regularity condition, the algorithm guarantees a unique solution by the contraction mapping theorem. Carranza (2010) shows $0 \le \partial\beta\widetilde{\mathbb{U}}^0_{i,t+1}/\partial r_{it} < 1$ is a sufficient condition for the regularity conditions to guarantee the algorithm has a unique interior fixed point.

The last two columns of Table 1.7 show the estimation results of Model III. Though this approximation approach is less structural than our model, the relative values of estimates are obtained similarly to the baseline model and Model II. Noticeably, the condition $0 \le \partial\beta\widetilde{\mathbb{U}}^0_{i,t+1}/\partial r_{it} < 1$ is satisfied on average with $\widehat{\eta}_2 = 0.393$ and $\widehat{\eta}_3 = 0.06$.

Table 1.7: Estimation Results: Experiments

		Model I		Model II		Model III	
Variable		Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Mean Utility (δ)							
(constant)	α_0	5.547	0.451	7.025	1.179	8.261	1.713
newness	λ	0.480	0.067	0.363	0.065	0.573	0.079
age	α_a	-1.101	0.157	-0.853	0.153	-1.283	0.186
head size	α_h	1.677	0.541	1.206	0.528	2.566	0.639
seasonality	α_s	0.254	0.047	1.461	0.034	0.503	0.055
Taylormade	$\alpha_b^{taylormade}$	3.484	0.168	3.187	0.164	3.743	0.198
Callaway	$\alpha_b^{callaway}$	2.444	0.116	2.287	0.113	2.591	0.137
Ping	α_b^{ping}	2.445	0.110	2.384	0.106	2.512	0.129
Cobra	α_b^{cobra}	1.644	0.100	1.683	0.096	1.629	0.118
Cleveland	$\alpha_b^{cleveland}$	1.570	0.102	1.621	0.098	1.519	0.120
Nike	α_b^{nike}	1.770	0.111	1.640	0.108	1.914	0.131
Titleist	$\alpha_b^{titleist}$	2.405	0.149	2.287	0.145	2.548	0.176
price	α_p	2.028	0.317	1.547	0.311	2.524	0.374
Consumer Heterogeneity (μ)							
newness	σ_λ	0.123	0.060	0.850	0.359	0.121	0.065
age	σ_a	0.104	0.033	0.031	0.012	0.109	0.086
head size	σ_h	0.102	0.115	1.780	0.632	0.307	0.112
seasonality	σ_s	0.135	0.127	0.524	0.236	0.132	0.148
price	σ_p	0.100	0.014	0.483	0.207	0.103	0.026
Continuation Value of Utility ($\mathbb{U}^0$)							
Idiosyncracy	η_1					0.359	0.061
Inclusive value	η_2					0.393	0.041
Interaction	η_3					0.060	0.025

Note: Model I assumes myopic consumers. Model II adopts two behavioral assumptions, consumers have dynamic inconsistency for seasonality and they have high demand in golf season and in December. Model III is the baseline model with approximated continuation value.

Table 1.8: Dollar Values of Characteristics: Experiments

Variable		% Changes in Price Model I	Model II	Model III
Product Characteristics				
newness	λ	0.2670	0.2645	0.2549
age	α_a	-0.0442	-0.0449	-0.0415
head size	α_h	0.0181	0.0171	0.0223
Brand Dummies				
Taylormade	$\alpha_b^{taylormade}$	4.5731	6.8469	3.4060
Callaway	$\alpha_b^{callaway}$	2.3372	3.3857	1.7914
Ping	α_b^{ping}	2.3388	3.6695	1.7054
Cobra	α_b^{cobra}	1.2494	1.9681	0.9068
Cleveland	$\alpha_b^{cleveland}$	1.1688	1.8515	0.8254
Nike	α_b^{nike}	1.3936	1.8867	1.1347
Titleist	$\alpha_b^{titleist}$	2.2736	3.3857	1.7443

Note: The numbers are percent changes in price induced by the change in corresponding variables. Brand values are compared to minor brands with zero values of all brand dummies. In calculating the dollar values, a representative consumer is considered by ignoring consumers' heterogeneity.

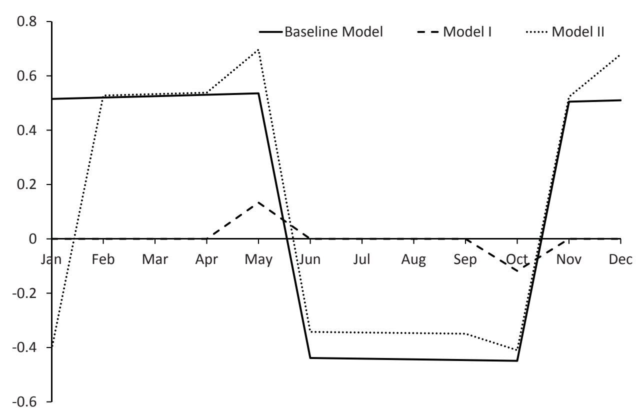


Figure 1.9: Dollar Values of Seasonality

Note: Dollar value represents the adjusted percent price to make a representative consumer equally satisfied as the previous month when other things are held equal. Positive value means that a representative consumer is willing to pay more in the corresponding month compared, *ceteris paribus*, to the previous month, and vice versa.

Table 1.8 shows the dollar values calculated in the same way as in Table 1.5. In Model I that assumes myopic consumers, the effect of newness is the highest with 26.7 percent increase in price, while all other dynamic models result in less than 26.5 percent increases in price. Model II resembles the baseline model but less brand effect is obtained. Model III suggests even lesser brand effect.

Figure 1.9 is the dollar values of seasonality for the baseline model, Model I, and Model II. Dollar value represents the adjusted percent price to make a representative consumer equally satisfied as the previous month when other things are held equal. Positive value hence means that a representative consumer is willing to pay more in the corresponding month compared, *ceteris paribus*, to the previous month, and vice versa. Baseline model shows a dichotomous behavior such that a representative consumer consistently is willing to pay around 50 percent more in the off season, i.e. from November to May, and 44 percent less in the peak season, i.e. from June to October, as time is increased by a month. In other words, during the off season, consumers are willing to pay more as the golf season comes closer. It agrees with what companies reveal in regard to seasonality, e.g. Callaway Golf Company Annual Report 2010 states that

> "*[t]he Company's business is subject to seasonal fluctuations (16p.) [B]ecause of this seasonality, a majority of the Company's sales and most, if not all, of its profitability generally occurs during the first half of the year (33p.)*"

Model II exhibits similar pattern to the baseline model but higher peaks are observed in May and December and lowest dollar values are obtained in January and October. Deviations in May and October are due to the assumption of hyperbolic discounting while those in December and January are as a result of additional seasonality shock given in December. The myopic model, Model I, suggests much smaller dollar value of seasonality. Due to the assumption of myopic consumers, the influence of shifting one month forward is only effective in May and October, the beginning of

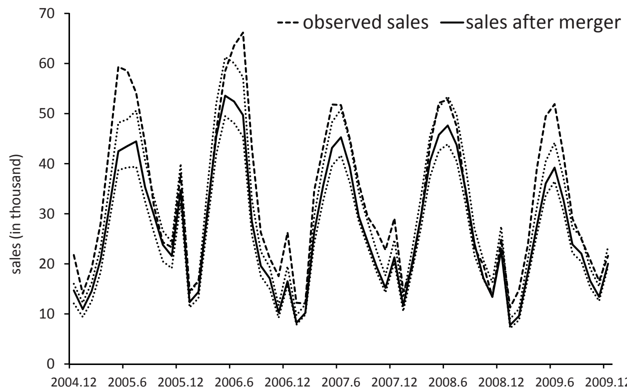


Figure 1.10: Sales of Callaway and Ping: Observed and Counterfactual

Note: Dashed line is the observed total sales of Callaway and Ping and solid line is the counterfactual total sales when they are merged. Dotted lines denote the 95% confidence interval of the counterfactual total sales.

golf season and off season, respectively.

From the results found in Table 1.7 and Figure 1.9, the misspecification of the static model results in the effects of newness and seasonality are smoothed out. It is because the denominator in equations (1.17) and (1.22) do not account for the heterogeneous continuation value of utility, i.e. $\exp(r_{it}) + \exp(\beta \mathbb{U}^0_{i,t+1})$ becomes $\exp(r_{it}) + 1$.

1.5.3 A simple counterfactual analysis: merger of Callaway and Ping

To see the impact of consumers' preferences for newness, we perform a counterfactual analysis with two firms ranked second and third in total sales shown in Figure 1.1 (b), namely Callaway and Ping. Consider they are merged before the sample period and do business under the brand of Callaway. The impact of merger is higher frequency of model changes. In our model, it yields the modification of newness values. In other words, consumers observe more frequent introductions of new Callaway models after merger and the existing models lose newness faster than before. We assume the pricing decisions of two firms remain same as observed. Figure 1.10 plots the observed total sales of Callaway and Ping, and the counterfactual total sales when they are assumed to be merged. Overall, the sales are reduced when they are merged. In particular, the loss in sales is more conspicuous when they have relatively high frequency of introductions, in the summer of 2005, 2006 and 2009. It suggests that the cannibalization effect in sales plays more significant role when the model changes are made faster than observed, i.e. optimal pace of model changes.

1.6 Concluding Remarks

In this study, we identified consumers' preferences for newness of products and measured the amount of prestige effect that the newness of a product brings. With strong preferences for newness of products, forward-looking heterogeneous consumers are modeled to behave time optimally in their purchasing decision. The deterministic transition of seasonality is also modeled. Estimation results show the evidence of strong preference for newness, depreciation with respect to age, and brand premium. Experiments with behavioral assumptions and approximation confirm the robustness of the results. A simple counterfactual analysis shows the negative effect on two firms' total sales when they are assumed to be merged. It is due to the faster loss of newness in some products after merger. While emphasizing on consumers' preferences for newness of products, this study omits modeling firms' pricing, introduction, and endogenous choice of product characteristics.

CHAPTER II

A Grouped Mixed Proportional Hazard Model with Social Interactions: The Passage of the Motorcycle-Helmet-Use Law

2.1 Introduction

The main purpose of this chapter[1] is to understand the decision making mechanism of state legislations and, in particular, to find an evidence that the coverage of motorcycle-helmet-use law, hereafter MHU law, in a state is spatially dependent on neighboring states' status or decisions.

To analyze such a social interaction aspect in state legislation decision making, we specify a discrete choice model with social interactions as Brock and Durlauf (2001a,b), which incorporates the social interaction term in the random utility maximization problem. Individual expectation to others' decision is the key component deriving the social interactions. In this particular example, however, each decision making can be only realized at certain times, such as only when there is a legislate meeting (i.e. timing friction in decision realization.) By introducing such timing friction in the discrete choice model, similarly as the job searching model of Lancaster (1979), we derive a hazard model with duration dependence. In particular, we

[1] This chapter is a joint essay with Yoonseok Lee.

introduce a grouped mixed proportional hazard (MPH) model (e.g. Kalbfleisch and Prentice, 1973; Prentice and Gloeckler, 1978) with social interactions, where the hazard rate is a function of other states' discrete choices (e.g. Carruthers, Guinnane, and Lee, 2012). The baseline hazard and the distribution of the unobserved heterogeneity are specified nonparametrically (e.g. Heckman and Singer, 1984; Meyer, 1990, 1995; Prentice and Gloeckler, 1978). The EM-algorithm is used to facilitate the estimation of the MPH model with social interactions (e.g. Lee, 2007).

Estimation results using U.S. state-level panel data from 1975 to 2006 show statistically significant interactions with neighboring states' decisions on the MHU law, whereas safety concern is found not to be important when the policy makers make decisions. Though this analysis does not give an answer to an issue whether introducing the mandatory MHU law is beneficial or not, it explains a behavioral aspect of the legislative decision making procedure in the context of social interactions and empirically shows how the proximity between agents affects the decision making.

The remainder of the chapter is organized as follows. Section 2.2 describes the main model, discrete choice with social interactions and timing friction in decision realization. Section 2.3 develops a mixed proportional hazard model with grouped data, that includes both the interaction term and the unobserved heterogeneity, and provides econometric foundation of the choice model. Section 2.4 introduces the EM algorithm as an estimation method. Section 2.5 delineates the data and discusses the estimated results. Section 2.6 concludes the chapter with some remarks. Technical details are provided in the Appendix B.

2.2 Discrete Choice with Social Interactions

2.2.1 Random utility maximization

We consider individuals $i = 1, \ldots, n$ who can switch their choices over time $t = 1, \ldots, T$. The binary choice is denoted by an indicator variable $d_i(t)$, which has support $\{-1, 1\}$. The observed characteristics of each individual i, possibly time-varying, are denoted as a $k \times 1$ vector $x_i(t)$. The unobservable independent random private utility (or a random shock) is denoted by $\varepsilon_i(t, d_i(t))$, which depends on the realized individual's choice and is independent of $x_i(t)$ for all i and t.

Similarly as Brock and Durlauf (2001a,b), we impose social interactions in the individual decision by assuming that the expected behavior of others influence each individual choice. Examples are spill-over effects, externalities and peer-effects. More precisely, we assume a twice-differentiable instantaneous individual utility function given by $V(d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t)))$, where $d_{-i}(t) = (d_1(t), \ldots, d_{i-1}(t), d_{i+1}(t), \ldots, d_n(t))$ denotes the vector of choices other than that of individual i and $\mu_i^e(d_{-i}(t))$ represents individual's belief concerning the choices of other agents. Given that individuals are myopic so that they only make choices by comparing current utilities without considering future paths of choices, each choice is described by solving

$$\max_{d_i(t) \in \{-1,1\}} V(d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t))) \tag{2.1}$$

for each t. Note that myopic behavior can be understood in the context of repeated search or infinite discount rate, and it well justifies the proportional hazard specification (e.g. Van den Berg, 2001).

We assume that the individual utility function can be represented as

$$\begin{aligned} V(d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t))) &= u\left(d_i(t), x_i(t)\right) + \varepsilon_i(t, d_i(t)) \\ &\quad + U_S\left(d_i(t), x_i(t), \mu_i^e(d_{-i}(t))\right), \end{aligned} \tag{2.2}$$

where $u\left(d_i(t), x_i(t)\right)$ and $U_S\left(d_i(t), x_i(t), \mu_i^e(d_{-i}(t))\right)$ are observable deterministic private and social utilities, respectively. Without the social utility U_S, (2.2) corresponds to the standard random utility function. We further let the deterministic private utility be linear as

$$u\left(d_i(t), x_i(t)\right) = d_i(t)u_1(x_i(t)) + u_2(x_i(t)), \tag{2.3}$$

which is without loss of generality since it coincides with the original utility function on the support of the individual choices $\{-1, 1\}$. (E.g. Brock and Durlauf, 2001b, p.3307.) The social utility possess a generalized quadratic conformity effect as

$$U_S\left(d_i(t), x_i(t), \mu_i^e(d_{-i}(t))\right) = -\frac{\alpha}{2}\mathbb{E}_{i,t}\left[\sum\nolimits_{j\neq i} w_{ij}(t)(d_i(t) - d_j(t))^2\right], \tag{2.4}$$

where α is an unknown scalar parameter and $\mathbb{E}_{i,t}[\cdot]$ denotes the conditional expectation of i at t given the values of $(d_i(t), x_i(t))$. In this specification, $\alpha w_{ij}(t) = \partial^2 V/\partial d_i(t)\partial\mathbb{E}_{i,t}[d_j(t)] = \partial^2 U_S/\partial d_i(t)\partial\mathbb{E}_{i,t}[d_j(t)]$ measures the strategic complementarity between individual choices and the expected choices of others (e.g. Brock and Durlauf, 2001b; Cooper and John, 1988). The term $w_{ij}(t)$ in (2.4) represents the interaction weight between agents i and j. For each t, we define $w_{ij}(t) = g(||\varphi_i(t), \varphi_j(t)||)$ for $i, j = 1, 2, \ldots, n$, where $w_{ii}(t) = 0$, $w_{ij}(t) = w_{ji}(t)$ and $||\cdot,\cdot||$ is a proxy of economic distance between i and j. More precisely, $||\cdot,\cdot||$ is a distance function of a pair of characteristics $\varphi_i(t)$ and $\varphi_j(t)$, and $g(\cdot)$ is a nonnegative and strictly increasing function with $g(0) = 0$. Note that $w_{ij}(t)$ are typically assumed to be fixed and deterministic for the identification purposes (e.g. Manski, 1993). We implicitly assume that there is no individual without a neighbor; the choice of neighborhood is also fixed and not endogenous.

For $d_i^2(t) = 1$ we can rewrite (2.4) as $\alpha\sum_{j\neq i} w_{ij}(t)(d_i(t)\mathbb{E}_{i,t}[d_j(t)] - 1)$, and obtain the values of expectations $\mathbb{E}_{i,t}[d_j(t)]$ by assuming that all the agents have rational expectations for each t:

$$\mathbb{E}_{i,t}\left[d_j(t)\right] = \mathbb{E}_t\left[d_j(t)|x_1(t),\ldots,x_n(t);\mathbb{E}_{k,t}d_\ell(t) \text{ for } k,\ell=1,\ldots,n\right]. \tag{2.5}$$

The solutions satisfying this self-consistent condition (e.g. Brock and Durlauf, 2001a) close the model, provided they exist. Uniqueness of the self-consistent equilibrium gives the identification condition in this framework. More precisely, from the standard random utility maximization, individual i chooses $d_i(t)$ over $-d_i(t)$ at time t if

$$V(d_i(t),x_i(t),\mu_i^e(d_{-i}(t)),\varepsilon_i(t,d_i(t))) \geq V(-d_i(t),x_i(t),\mu_i^e(d_{-i}(t)),\varepsilon_i(t,-d_i(t))), \tag{2.6}$$

whose probability is given by

$$\begin{aligned}
&\mathbb{P}\left\{V(d_i(t),x_i(t),\mu_i^e(d_{-i}(t)),\varepsilon_i(t,d_i(t))) \geq V(-d_i(t),x_i(t),\mu_i^e(d_{-i}(t)),\varepsilon_i(t,-d_i(t)))\right\}\\
&=\mathbb{P}\left\{\varepsilon_i(t,-d_i(t))-\varepsilon_i(t,d_i(t)) \leq 2\left[u_1(x_i(t))+\alpha\sum\nolimits_{j\neq i} w_{ij}(t)\mathbb{E}_{i,t}[d_j(t)]\right]d_i(t)\right\}\\
&=G\left(2\left[u_1(x_i(t))+\alpha\sum\nolimits_{j\neq i} w_{ij}(t)\mathbb{E}_{i,t}[d_j(t)]\right]d_i(t)\right),
\end{aligned} \tag{2.7}$$

where we assume that $\varepsilon_i(t,-1)-\varepsilon_i(t,1)$ is independent and identically distributed (i.i.d.) over i and t with $G(\cdot)$ being the distribution function that is symmetric around zero. Using this result, the self-consistent solution $\overline{\mu}(t)=\mathbb{E}_{i,t}[d_j(t)]$ satisfying (2.5) can be obtained by solving

$$\overline{\mu}(t) = H\left(u_1(x_i(t))+\alpha\sum\nolimits_{j\neq i} w_{ij}(t)\overline{\mu}(t)\right), \tag{2.8}$$

which is assumed to exist uniquely, where $H(y)=2G(2y)-1$ for some $y\in\mathbb{R}$.

Assumption II.1 (self-consistent equilibrium). *Given $H(\cdot)$, $u_1(x_i(t))$ and $\alpha\sum_{j\neq i} w_{ij}(t)$, there exists a unique self-consistent expectation $\overline{\mu}(t)$ satisfying (2.8) for each t.*

Note that the unique existence of the self-consistent equilibrium requires assumptions on the distribution $G(\cdot)$ as well as $u_1(x_i(t))$ and $\alpha \sum_{j \neq i} w_{ij}(t)$. For example, when G is logistic, $H(\cdot) = \tanh(\cdot)$ and thus the existence of self-consistent equilibrium in (2.8) follows immediately (e.g. Brock and Durlauf, 2001a, Proposition 1). Moreover, provided that $\alpha \sum_{j \neq i} w_{ij}(t) \leq 1$, which holds if $\alpha \leq 1$ under the row normalization assumption (i.e. $\sum_{j \neq i} w_{ij}(t) = 1$ for each i and t) in Assumption II.6, the equilibrium is unique from the properties of the $\tanh(\cdot)$ function. When $\alpha \sum_{j \neq i} w_{ij}(t) > 1$, on the other hand, the uniqueness can be obtained only when $|u_1(x_i(t))|$ is large enough (e.g. Brock and Durlauf, 2001a, Proposition 2).

2.2.2 Choice with timing frictions

Though we introduce the social interaction term in the choice mechanism, the random utility analysis in the previous subsection is rather standard since we assume that each individual is myopic and the utility maximization problem is solved for each t. There could be the cases, however, that the choice is not allowed for some t even though the random utility maximization tells so. A passage of a law is a good example: Let the agent i be the state legislature and the passage of a particular law is determined by solving the random utility maximization. For some cases, even when people want to pass the law soon, the number of legislate meetings is limited and thus there could be a exogenous timing friction in realizing the choice.

To incorporate such an idea into the framework, we suppose that an alarm clock is assigned to each agent i, where the alarms are independent over time. When the clock rings, the agent has an opportunity to revise her choice. The choice and the alarm are mutually independent, and the choice is assumed to be made when the clock rings or right before at $t^- = \lim_{\Delta \to 0}(t - \Delta)$ for $\Delta > 0$. More precisely, we let the occurrence of alarm follows the time-dependent (or non-homogeneous) Poisson process with rate $\rho_i(t) > 0$, so that the expected number of alarms in the time interval $[0, t)$ is given as

$\int_0^t \rho_i(s)ds$ for each i. We further impose a common factor structure on $\rho_i(t)$ as

$$\rho_i(t) = \lambda_0(t)v_i \tag{2.9}$$

with $\lambda_0(t) > 0$ and $v_i > 0$, where $\lambda_0(t)$ is the common time-dependent rate across individuals while the time-invariant (unobservable) heterogeneity v_i allows for variations over i.

If individual i followed the decision rule in (2.6), her revision of the choice is observed in the short time period $[t, t+\Delta)$ with $\Delta > 0$ if and only if (i) her alarm clock rings at t and (ii) the utility with the new choice exceeds that with the old one. Apparently, the probability of the evens (i) and (ii) are $\rho_i(t)\Delta = \lambda_0(t)v_i\Delta$ and $G(2[u_1(x_i(t)) + \alpha \sum_{j\neq i} w_{ij}(t)\mathbb{E}_{i,t}[d_j(t)]]d_i(t))$, respectively from (2.9) and (2.7). Since these events are assumed to be mutually independent, the probability of choice revision in $[t, t+\Delta)$ conditional on no revision occurred before t is given by their product and it yields the *hazard function* for the choice change of individual i:

$$\lambda_0(t)G\left(2[u_1(x_i(t)) + \alpha \sum\nolimits_{j\neq i} w_{ij}(t)\mathbb{E}_{i,t}[d_j(t)]]d_i(t)\right) v_i, \tag{2.10}$$

which is in the form of the mixed proportional hazard (MPH) model with the baseline hazard function $\lambda_0(t)$ and the unobservable heterogeneity v_i (e.g. Lancaster, 1990; Van den Berg, 2001).

We remark that knowing the characteristic of the choice-change-allowance process is crucial to correctly identify the choice behavior. For example, suppose we ignore the choice-change-allowance process and simply observe the choice behaviors at a fixed frequency. Then for any two identical consecutive choices of individual i, say $d_i(t) = d_i(t') = 1$ for $t < t'$, we cannot tell which scenario results in such observations among the followings: (i) $d_i(t') = 1$ because $V_i(t', 1) - V_i(t', -1) > 0$, where $V_i(t, d_i(t)) = V(1, x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t)))$, and the choice revision is allowed at t'; (ii) $d_i(t') = 1$

because the choice revision is not allowed at t' whether the sign of $V_i(t', 1) - V_i(t', -1)$ is positive or negative. Particularly when $V_i(t', 1) - V_i(t', -1) < 0$ but the individual cannot change her choice because it is not allowed at t', it violates the fundamentals of the standard random utility maximization problem.

2.3 Grouped MPH Models with Social Interactions

2.3.1 Semiparametric duration models with grouped data

Though the failure time is continuous, the standard panel data only provide observations on failure times aggregated up to discrete intervals (i.e. grouped duration data; e.g. Kalbfleisch and Prentice, 1973; Prentice and Gloeckler, 1978). To handle this discrepancy, we suppose failure times are grouped into intervals $B_s = [b_{s-1}, b_s)$ for $s = 1, 2, \ldots, S$ with $b_0 = 0$ and $b_S = \infty$ without loss of generality, where the length of each interval corresponds to the panel survey frequency. Survival to time b_s is the same as surviving until the s-th interval B_s (e.g. Sueyoshi, 1995) and the failure time of individual i in B_s are recorded as $\tau_i = s$. Since we usually deal with equi-spaced panel data, we simply let $b_t = t$ for all $t = 1, \ldots, T$ and consider T intervals: $[0, 1), [1, 2), \ldots, [T - 1, T)$ ignoring the last interval $[T, \infty)$ that is after the survey period and thus all durations lasting over T are naturally right censored. We further assume that covariates are at best recorded up to intervals and the values does not change in each interval $[t - 1, t)$. Following the standard notations of panel data, we simply rewrite $x_i(t)$, $w_{ij}(t)$, $d_i(t)$ as $x_{i,t}$, $w_{ij,t}$, $d_{i,t}$, respectively, in what follows.

In order to make the empirical analysis tractable, we further impose three more assumptions. First, we specify $u_1(x_{i,t}) = x'_{i,t}\beta$ for a $k \times 1$ parameter vector β as the standard discrete choice or the Cox's (1972) MPH models. Second, the interaction weight $w_{ij,t}$ is time invariant so that it is simply denoted as w_{ij}. Since the main analysis is based on the geographical proximity as a measure of the interaction weights, this assumption holds naturally. However, as long as $w_{ij,t}$ is predetermined, time varying weight can be considered. Third, knowing that the renewal of the choice is not highly frequent in the data set (only 42 revisions are made over 383 time periods,) we assume that the expectation for neighbors' current choices is equal to the previous choices (i.e. $\mathbb{E}_{i,t}[d_{j,t}] = d_{j,t-1}$ for all $j \neq i$ and t) similarly as Wallis (1980). Using the grouped duration, we thus define the hazard rate as

$$\lambda_i \left(t | x_{i,t}, \sum_{j \neq i} w_{ij} d_{j,t-1}, v_i \right) = \lambda_0(t) \phi \left(x'_{i,t} \beta + \alpha \sum_{j \neq i} w_{ij} d_{j,t-1} \right) v_i \tag{2.11}$$

from (2.10) for all $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where $\phi(y) = G(2y)$ if the hazard is from the choice -1 to 1, and $\phi(y) = 1 - G(2y)$ if it is from 1 to -1.

Heckman and Singer (1984) suggest that the distribution of the unobserved heterogeneity v_i be nonparametrically estimated in order to avoid any misspecification problem; as the number of mass points increases, discrete distributions can approximate any distribution arbitrarily well. The nonparametric estimator, however, is very sensitive to the assumed shape of the baseline hazard function $\lambda_0(t)$ (e.g. Trussell and Richards, 1985) especially with single-spell data. For a possible solution, Meyer (1990, 1995) proposes to use piecewise constant baseline hazard functions (e.g. Prentice and Gloeckler, 1978) as well as the Heckman-Singer approach. More precisely, we let the baseline hazard $\lambda_0(t)$ be piecewise constant:

$$\lambda_0(t) = \exp(\gamma_\ell) \quad \text{if } t \in [a_{\ell-1}, a_\ell), \tag{2.12}$$

where $1 = a_1 < \ldots < a_h = T$ (with $h \leq T$) is a subsequence of $t = 1, \ldots, T$. Such specification is useful especially when the hazard rate has much fluctuation or frequent peaks. It extracts common deterministic time trends from the covariates as the standard time effects in panel regressions. Note that γ_ℓ satisfies $\log \int_{a_{\ell-1}}^{a_\ell} \lambda_0(r)\, dr = \gamma_\ell$ for the continuous time case and the most flexible case is to let $a_t = t$ so that

$\gamma_t = \log \int_{t-1}^{t} \lambda_0 (r)\, dr$ for all t. For the unobserved heterogeneity, we assume v_i to be *i.i.d.* with a discrete distribution with m supports, whose density is given as

$$f(v) = \sum\nolimits_{j=1}^{m} p_j 1\{v = q_j\}, \tag{2.13}$$

where $\mathbb{E}v_i = 1$, $\sum_{j=1}^{m} p_j = 1$, $0 < p_j < 1$ and $0 < q_j < \infty$ for all $j = 1, 2, \ldots, m$. $1\{\cdot\}$ is the binary indicator.

2.3.2 Regularity conditions

Based on the specification (2.11), (2.12) and (2.13) in the previous section, we consider an MPH model given by

$$\lambda_i (t|v_i) = \exp(\gamma_t)\phi\left(x_{i,t}'\beta + \alpha \sum\nolimits_{j\neq i} w_{ij} d_{j,t-1}\right) v_i \tag{2.14}$$

for some known function $\phi : \mathbb{R} \to \mathbb{R}$, where v_i is *i.i.d.* with density given in (2.13) and independent of $z_{i,t} = (x_{i,t}', \sum_{j\neq i} w_{ij} d_{j,t-1})'$. Carruthers, Guinnane, and Lee (2012) use a similar method with the exponential link function as a particular example of $\phi(\cdot)$. The weighted sum of $d_{j,t-1}$ by the interaction weights w_{ij} can be interpreted as the average influence of other agents' past decisions on i that can be understood as the individual i's expectation on the others' behavior or a learning effect. We first assume the following conditions.

Assumption II.2 (failure time)**.** *The failure time $\tau_i > 0$ is independent across i conditional on $z_{i,t}$ and v_i for each t; τ_i is independent of the censoring time C_i for all i.*

Assumption II.3 (unobserved heterogeneity)**.** *The unobserved heterogeneity $v_i > 0$ is i.i.d. of finite mixture (2.13) with $\mathbb{E}v_i = 1$; v_i is independent of $z_{i,t}$ and C_i for all i and t.*

Assumption II.4 (baseline hazard)**.** *The baseline hazard $\lambda_0(t)$ is nonnegative and piecewise constant given by (2.12) with $a_t = t$ for all t.*

Heckman and Singer (1984) assume that the distribution of the censoring variable C_i is known and independent of the covariate $z_{i,t}$ to show the consistency of maximum likelihood estimators with nonparametric unobserved heterogeneity v_i. In our case, the censoring only occurs at the fixed time T when the panel survey is over, so those assumptions hold naturally. $\mathbb{E}v_i$ is usually normalized to one so that the expected hazard rate becomes the unconditional hazard rate with no unobserved heterogeneity: $\mathbb{E}[\lambda_i(t|v_i)|z_{i,t}] = \lambda_i(t)$, where $\lambda_i(t) = \exp(\gamma_t)\phi(x'_{i,t}\beta + \alpha \sum_{j \neq i} w_{ij} d_{j,t-1})$. Surely the finite mean condition of v_i restricts its tail behavior and it is necessary for identification (e.g. Van den Berg, 2001). The number of support of v is assumed to be finite and fixed, though identification can be obtained even with increasing number of supports (e.g. Heckman and Singer, 1984; Kiefer and Wolfowitz, 1956; Meyer, 1995). We assume further conditions for identifying λ_0, β, α, and the distribution of v.

Assumption II.5 (covariates)**.** *(i) $z_{i,t} \in \mathcal{Z}$ for an open set $\mathcal{Z}$ in $\mathbb{R}^{k+1}$ and $Z = (z_{1,1}, \ldots, z_{n,T})'$ is of full column rank. (ii) No element of $x_{i,t}$ is constant and at least one argument of $x_{i,t}$ is defined on the continuum. (iii) The regression function $\phi : \mathcal{Z} \to \mathbb{R}$ is nonlinear and differentiable on $\mathcal{Z}$.*

Ignoring the social interaction term, Assumptions II.3, II.4, and II.5 yield identifiably of our model (2.14) up to a constant multiplication as Elbers and Ridder (1982). In particular, Assumption II.3 is the same as Assumption 1 of Elbers-Ridder; Assumption II.4 satisfies Assumption 2 of Elbers-Ridder because $\int_0^t \lambda_0(r)dr = \sum_{s=1}^t \exp(\gamma_s)$ is an increasing function of $t \geq 0$; Assumption II.5 and the form of the proportional hazard function satisfy Assumption 3 of Elbers-Ridder. Adding the social interaction term $\sum_{j \neq i} w_{ij} d_{j,t-1}$ in $z_{i,t}$ does not change the identification result as long as Z is of full column rank and ϕ is nonlinear (e.g. Brock and Durlauf, 2001b). Note that the duration model we consider here does not depend on the durations of others directly; instead the durations are dependent indirectly by including the choice-dependent term $\alpha \sum_{j \neq i} w_{ij} d_{j,t-1}$ in the hazard rate. The identification in this case, therefore, can be obtained as the standard MPH models by considering $\alpha \sum_{j \neq i} w_{ij} d_{j,t-1}$ as another predetermined regressor, which becomes much simpler than checking the self-consistent condition under duration dependence like Brock and Durlauf (2001b, Ch.4.2). One remark is that there is no endogeneity issue in this case because the random private utility $\varepsilon_{i,t}$ in the previous section is assumed to be exogenous and each individual makes decisions myopically. That is, the current decision is based on the current values of observable covariates $z_{i,t}$ only and thus no simultaneity issue arises. The following condition is on the interaction weight w_{ij}, where we let W be the $n \times n$ matrix whose (i, j)-th element is w_{ij}.

Assumption II.6 (interaction weight matrix)**.** *(i) The interaction weight matrix W is predetermined and correctly specified; (ii) each element of W is nonnegative and all the diagonal elements are zero; (iii) W is row normalized, i.e. $\sum_{j=1}^{n} w_{ij} = 1$ for all i; (iv) W is independent of v_i for all i, and W is not in the range space of $X = (x_{1,1}, \ldots, x_{n,T})'$.*

It is important to assume that the interaction weight matrix W is predetermined and independent of v_i. Pre-specifying W outside the model is an easy way to obtain a predetermined and exogenous W, which prevents any identification problem as pointed by Manski (1993)−the reflection problem. The independence assumption also guarantees the absence of endogeneity problems since all the diagonal elements of W are zero by construction and Assumption II.3 holds. Keeping the interaction weight matrix W out of the covariate space prevents any possible multicollinearity problem between $x_{i,t}$ and $\sum_{j \neq i} w_{ij} d_{j,t-1}$ in the index structure. The row normalization condition is standard in spatial econometrics literature (e.g. Anselin and Bera, 1998), which prevents the weighted sum $\sum_{j \neq i} w_{ij} d_{j,t-1}$ from exploding under the in-filling asymptotics (i.e. increasing the number of observations within a fixed boundary.) It controls the degree of cross sectional dependence so that the standard M-estimator has proper asymptotic properties (e.g. Kelejian and Prucha, 1999; Lee, 2004).

2.4 Estimation via EM Algorithm

2.4.1 Log-likelihood function

We let $c_{i,t}$ be the binary censoring indicator equal to one if the duration of i is censored in the t-th interval $[t-1,t)$. Apparently, $c_{i,t}=1$ implies $c_{i,t+1}=1$ for all i and the censoring variable C_i corresponds to the smallest t that gives $c_{i,t}=1$.

We consider the individual i, who drops out of the sample in the T_i-th interval $[T_i-1,T_i)$ either by exiting the initial state ($c_{i,T_i}=0$) or by censoring ($c_{i,T_i}=1$). If the panel is balanced, $T_i=T$ for all i. If we ignore conditioning on the initial state, the conditional log-likelihood function on the unobserved v is then given by

$$\begin{aligned}\log L(\gamma,\theta|v) &= \sum_{i=1}^{n}\delta_i\log\left(1-\exp\left(-\exp\left(\gamma_{T_i}\right)\phi\left(x'_{i,T_i}\beta+\alpha W d_{T_i-1}\right)v_i\right)\right)\\ &\quad -\sum_{i=1}^{n}\sum_{s=1}^{T_i-1}\exp\left(\gamma_s\right)\phi\left(x'_{i,s}\beta+\alpha W d_{s-1}\right)v_i \end{aligned} \tag{2.15}$$

from (2.14) similarly as Prentice and Gloeckler (1978) and Meyer (1990, 1995), where $\delta_i=1-c_{i,T_i}$, $\gamma=(\gamma_1,\gamma_2,\ldots,\gamma_T)'$ and $\theta=(\beta',\alpha)'$. We also denote $Wd_{t-1}=\sum_{j\neq i}w_{ij}d_{j,t-1}$ with $d_{t-1}=(d_{1,t-1},\ldots,d_{n,t-1})'$ for $w_{ii}=0$. Integrating (2.15) over the distribution of v yields the unconditional log-likelihood given by

$$\log L(\gamma,\theta)=\sum_{j=1}^{m}p_j\log L\left(\gamma,\theta|v=q_j\right). \tag{2.16}$$

Note that, however, the ML estimation is not appropriate on the mixture model (2.16) since the parameters of the heterogeneity distribution are not guaranteed to lie on the interior of a compact set (e.g. Heckman and Singer, 1984; Lancaster, 1990, Chapter 8.4; Lee, 2000.) Moreover, many studies report that the ML estimation of mixture models has convergence problem when the models have both the piecewise constant baseline hazard and the finite mixture unobserved heterogeneity.

For estimation, we instead use the EM (Expectation-Maximization) algorithm (e.g. Dempster, Laird, and Rubin, 1977), which was originally invented to deal with inference in models imposing missing data. In our case, the unobserved heterogeneity v is essentially a problem of missing data. To fix this idea, we consider an alternative expression of the conditional log-likelihood function (2.15) for individual i as

$$\log L_i\left(\gamma, \theta | v_i = q_j\right) = \sum_{j=1}^{m} \eta_{ij} \log L_i^*\left(\gamma, \theta, q_j\right), \tag{2.17}$$

where $\eta_{ij} = 1\{v_i = q_j\}$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$ and

$$\begin{aligned} \log L_i^*\left(\gamma, \theta, q_j\right) &= \delta_i \log\left(1 - \exp\left(-\exp\left(\gamma_{T_i}\right)\phi\left(x_{i,T_i}'\beta + \alpha W d_{T_i-1}\right)q_j\right)\right) \\ &\quad - \sum_{s=1}^{T_i-1} \exp\left(\gamma_s\right)\phi\left(x_{i,s}'\beta + \alpha W d_{s-1}\right)q_j. \end{aligned} \tag{2.18}$$

Note that η_{ij} is unobservable and it can be viewed as missing data satisfying $\log f\left(v_i\right) = \sum_{j=1}^{m} \eta_{ij} \log p_j$ for the prior probabilities p_j $(j = 1, 2, \ldots, m)$ from (2.13). Hence the unconditional joint log-likelihood function of both the observed and the unobserved items is defined as

$$\begin{aligned} \log L(\Theta) &= \sum_{i=1}^{n} \left\{\log f\left(v_i\right) + \log L_i\left(\gamma, \theta | v_i\right)\right\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} \eta_{ij}\left\{\log p_j + \log L_i^*\left(\gamma, \theta, q_j\right)\right\}, \end{aligned} \tag{2.19}$$

where $\Theta = \left(\gamma', \theta', q_1, \ldots, q_m, p_1, \ldots, p_{m-1}\right)'$ is the complete parameter vector, provided that the known distribution function $G(\cdot)$ is free of additional unknown parameters.

Note that p_m is automatically determined from the restriction of probabilities, $p_m = 1 - \sum_{j=1}^{m-1} p_j$.

2.4.2 EM algorithm

The EM algorithm proceeds in two steps. The E-step calculates the conditional expectation of η_{ij} given observed data δ_i and $Z_i = \{z_{i,s} = (x'_{i,s}, Wd_{s-1})' : 0 \le s \le T_i\}$ and given the likelihood L_i^* evaluated at the current parameter estimates. It can be shown that the posterior probability of $v_i = q_j$ is derived as

$$\mathbb{E}\left(\eta_{ij}|\delta_i, Z_i\right) = \frac{p_j L_i^*\left(\gamma, \theta, q_j\right)}{\sum_{\ell=1}^{m} p_\ell L_i^*\left(\gamma, \theta, q_\ell\right)} \equiv \pi_{ij} \quad \text{for all } i = 1, \ldots, n \text{ and } j = 1, \ldots, m. \tag{2.20}$$

If we let $\widehat{\pi}_{ij}$ denote π_{ij} evaluated at the current parameter estimates, substituting $\widehat{\pi}_{ij}$ for η_{ij} in (2.19) gives the outcome of the E-step:

$$Q\left(\Theta\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \widehat{\pi}_{ij} \log p_j + \sum_{i=1}^{n}\sum_{j=1}^{m} \widehat{\pi}_{ij} \log L_i^*\left(\gamma, \theta, q_j\right). \tag{2.21}$$

Then, the M-step consists of maximizing $Q\left(\Theta\right)$ with respect to Θ, which only requires numerical maximization of $\log L_i^*\left(\gamma, \theta, q_j\right)$. Maximization with respect to p_j's has an explicit solution as $\widehat{p}_j = n^{-1}\sum_{i=1}^{n} \widehat{\pi}_{ij}$ from a general result in finite mixture models (e.g. Everitt and Hand, 1981). We iterate the entire E and M-steps until the estimates converge.

The initial values for the EM algorithm can be chosen from the ML estimates of the duration model without unobserved heterogeneity $\lambda_i\left(t\right) = \exp(\gamma_t)\phi(x'_{i,t}\beta + \alpha W d_{t-1})$. We then start the EM algorithm on the duration model with unobserved heterogeneity $\lambda_i\left(t|u_i\right) = \exp(\gamma_t + u_i)\phi(x'_{i,t}\beta + \alpha W d_{t-1})$, where $v_i = \exp(u_i)$, using the first step ML estimates $(\gamma_1^0, \ldots, \gamma_T^0, \beta^0, \alpha^0)$ and arbitrary (q^0, p^0) as the initial values. Note that the reparametrization $v_i = \exp(u_i)$ is convenient in solving the maximization problem since we do not need to impose restrictions on the sign of unobserved heterogeneity (i.e. $0 < v_i < \infty$ holds for any u_i.) For the distribution of v_i, we start with two points of support (q_1, q_2) and keep adding more points of support as long as all the estimates $\widehat{q}_j$ are distinct. Leaving the mean of v unrestricted, we omit the first term of baseline hazard (i.e. $\gamma_1 = 0$) so that the parameter to estimate is $\Theta = (\gamma_2, \ldots, \gamma_T, \beta', \alpha, q_1, \ldots, q_m, p_1, \ldots, p_{m-1})'$. The Simulated Annealing (e.g. Goffe, Ferrier, and Rogers, 1994) grid search is helpful to find or confirm the global maximum.

The EM algorithm is known to be robust to the choice of initial values and practically guarantees convergence to at least a local maximum. One of its disadvantages is that it dose not provide standard errors as an immediate by-product unlike the Newton-Raphson type methods. Louis (1982), Meng and Rubin (1991), Guo and Rodriguez (1992) and Oakes (1999) propose a way to find a positive-definite observed information matrix within the EM algorithm framework, from which we can obtain the asymptotic variance matrix. More precisely, under the regularity conditions, Louis (1982) shows that the observed information matrix $\mathcal{I}(\Theta)$ can be obtained as $\mathcal{I}(\Theta) = \mathbb{E}_v \left(\frac{\partial^2}{\partial\Theta\partial\Theta'} \log L(\Theta) \right) - \mathbb{V}_v \left(\frac{\partial}{\partial\Theta} \log L(\Theta) \right)$,where $L(\Theta)$ is the complete data likelihood function in (2.19). The expectation $\mathbb{E}_v(\cdot)$ and variance $\mathbb{V}_v(\cdot)$ are taken over the conditional distribution of v given the observed data $\{\delta_i, Z_i\}_{i=1}^n$. As noted in Guo and Rodriguez (1992), the first term of $\mathcal{I}(\Theta)$ can be interpreted as the conditional expectation of the observed information when v is observed, whereas the second term represents the missing information associated with the conditional distribution of v given the observed data. The explicit form of the observed information matrix $\mathcal{I}(\Theta)$ in our model is given in the Appendix B.

2.5 The Passage of the Motorcycle Helmet Use Law

2.5.1 Data

Motorcycle-helmet-use (MHU) law The history of each state's coverage of the MHU law is obtained from the Insurance Institute for Highway Safety (http://www.iihs.org/laws/helmet_history.html.) In our model $d_{i,t} = 1$ means state i chooses the universal MHU law while $d_{i,t} = -1$ denotes all choices other than the universal law, which includes not only repealing the law but also passing the partial MHU law. Note that the partial MHU law applies only to young riders under a certain age (e.g. not older than 17) and adult riders either inexperienced (e.g. instruction permit holders) or without sufficient medical insurance. So the proportion of riders covered by the partial law is rather small. Under such definition, we have 15 states with no revision and 33 states with revisions. More precisely, we have 27 states with one revision, 4 states with 2 revisions, and other two states with 3 and 4 revisions, respectively, during the study period. Among these 42 revisions, 34 cases are from -1 to 1 (i.e. adopt the universal law) whereas 8 cases are from 1 to -1 (i.e. repeal or reduce the universal law.) Even though we have the dates of law changes, we set the unit of time as a month because of the availability of covariates, so that the entire sample period is 383 months (from February 1975 to December 2006; $T = 383$) for 48 states excluding Hawaii and Alaska ($n = 48$.)

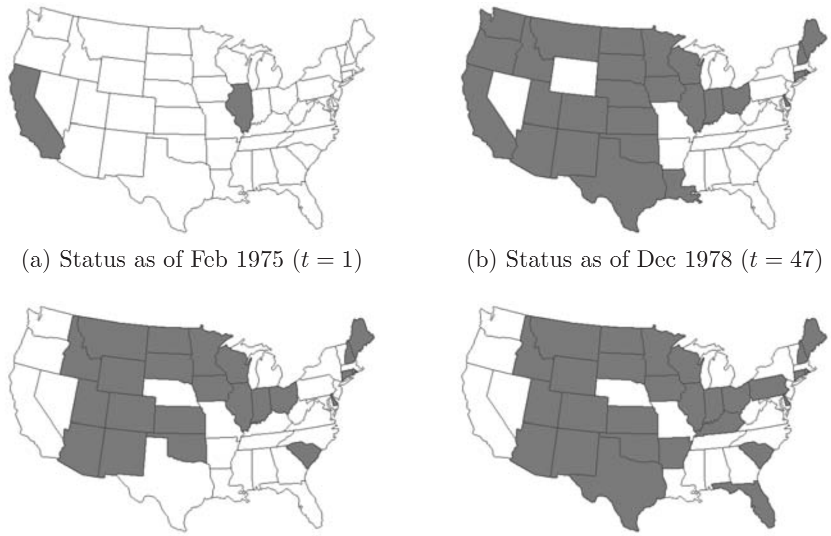
Though the study period is long, number of passages is limited and the baseline hazard γ_t does not change frequently. To estimate the piecewise constant baseline hazard (2.12) more efficiently, we group the time periods into four such that they reflect the history of legislating activities regarding to the MHU law during the sample period as follows (source: http://www.iihs.org/research/qanda/helmet_use.html):

• *From Feb. 1975 to Dec. 1978 (*$1 \leq t \leq 47$*)*: In 1966, the Highway Safety Act was introduced by the federal government, which required states to have mandatory MHU law if they wanted to receive federal funds for highway maintenance and construction. 47 states had complied by 1975; but in 1975, Congress withdrew it and half of the states had repealed the law within three years.

- *From Jan. 1979 to Dec. 1991 (*$48 \leq t \leq 203$*)*: There were no special activities to remark.
- *From Jan. 1992 to Sep. 1995 (*$204 \leq t \leq 248$*)*: In the Intermodal Surface Transportation Efficiency Act of 1991, signed by President Bush in December 1991, Congress created incentives for states to enact helmet and safety belt use laws. States with both laws were eligible for special safety grants, but states that had not enacted them by October 1993 had up to 3 percent of their federal highway allotment redirected to highway safety programs.
- *From Oct. 1995 to Dec. 2006 (*$249 \leq t \leq 383$*)*: Four years after establishing the incentives, Congress again reversed itself. In the fall of 1995, Congress lifted federal sanctions against states without MHU laws, paving the way for state legislatures to repeal the MHU laws.

Figure 2.1 shows the status of choices at four selected time periods, t = 1, 47, 248 and 383. Shaded states have no requirement or maintain the partial MHU law (i.e. $d_{i,t} = -1$.)

Note that we define two types of spells and failures: "type -1 spell" ("type 1 spell") denotes a spell in which individual i stays with $d_{i,t} = -1$ ($d_{i,t} = 1$) throughout. An individual starts "type -1 spell" if her choice changes from 1 to -1 (i.e. "(1 : -1)-failure") and "type 1 spell" if her choice changes from -1 to 1 (i.e. "(-1 : 1)-failure.") Each spell ends by either a new choice (i.e. a failure) or censoring then the other type of spell starts. Figure 2.2-(a) depicts the Kaplan-Meier estimator for survivor function on both types of failures where the vertical lines separate the time period as described above. In more details, Figure 2.2-(b) plots the two Kaplan-Meier estimators by the type of failures separately. We observe that all the failures in the first period are



(a) Status as of Feb 1975 ($t = 1$)

(b) Status as of Dec 1978 ($t = 47$)

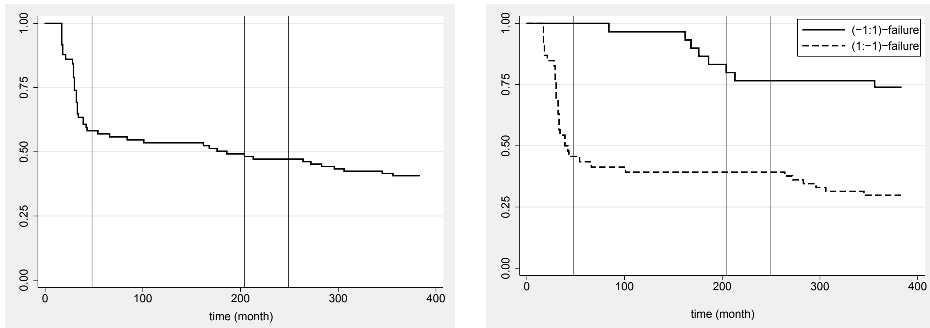
(c) Status as of Sep 1995 ($t = 248$)

(d) Status as of Dec 2006 ($t = 383$)

Figure 2.1: History of Motorcycle-Helmet-Use Law

Note: States are shaded when they maintain the partial motorcycle-helmet-use law or have no requirement.

from "(1 : -1)-failure." Big drop at the beginning of the third period accounts for a revision from a partial to universal coverage in California. (The records shows that only two states expanded the coverage during the third period: California expanded the law to universal coverage in January 1992 and Rhode Island expanded coverage to operators 20 years-old and younger in July 1992. In our definition of revision, however, only California's revision counts.) Finally most revisions in the last period are in "(1 : -1)-failure." It suggests that when the Federal government lifted the incentive to adopt the universal MHU law, states reacted quickly to repeal or reduce the law as shown in the first and the last period. States, however, have the incentive to adopt the universal law voluntarily as shown in the second period group where no incentive or requirement was given by the Federal government.



(a) For both failure types

(b) Grouped by failure types

Figure 2.2: Kaplan-Meier Estimator for Survival Functions

Social interactions To analyze the social interaction between neighboring states, we consider 48 states excluding Hawaii and Alaska, where the neighborhood is defined based on the geographical locations. More precisely, if two states are adjacent or share borders, then they are defined as neighbors: $w_{ij} = w_{ji} = 1$ if i and j are neighbors to each other, $w_{ij} = 0$ otherwise. We naturally assume that the neighborhood is time invariant in this case. We redefine w_{ij} by dividing it by the total number of borders shared (i.e. total number of neighbors) of state i, to make it as a well-defined weight as well as row-normalized (i.e. $\sum_{j=1}^{n} w_{ij} = 1$.)

Considering geographical locations as a measure of the social-dependence measure is intuitive in this particular example, even after controlling for any geographical, meteorological and cultural similarities among the states. Motocyclists frequently traveling between state boarders would like to have a homogenous regulation between them. Insurance companies may present the data comparing fatality rates across state border to emphasize that introducing the universal MHU law reduces the probability of mortality. If we regard a state's decision on the coverage law as a conclusion of the residents' consensus, therefore, the geographical proximity would be a strong factor to measure the social distance in this decision making procedure.

Table 2.1: Definitions of Variables and Summary Statistics

Variable	Description	Mean	Std.Dev	Min	Max
LRoadway	Total public road and street mileage in million miles in log scale	-2.750	0.795	-5.271	-1.184
Elevdiff	Difference between highest and lowest elevation in ft divided by 100,000	0.525	0.042	0.003	0.148
LPrecip	Mean number of days in a month with precipitation .01 inches or higher divided by 10 then logarithms taken	-0.133	0.367	-1.984	0.514
Population	Total population divided by a million	4.967	5.209	0.334	34.55
Registered	Number of registered motorcycles divided by 100,000	1.011	1.074	0.012	7.457
FatalRate	Number of fatalities in the previous month divided by Population	0.232	0.430	0	2.995
NbhdAvg	Average decisions of neighbors in the previous month	0.333	0.662	-1	1

Control variables National Highway Traffic Safety Administration (NHTSA) maintains the Fatality Analysis Reporting System (FARS,) which contains monthly fatality data from 1975 by the vehicle types. In this study, we count fatalities occurred by motorcycles only excluding mopeds, mini-bikes and motor scooters, and use the fatality rates per population. Since agents can only observe the fatality rate up to previous month at each time, we include the fatality rate lagged by one month. The number of registered motorcycles is provided by the Federal Highway Administration (FHWA.) National Climate Data Center (NCDC) by National Oceanic and Atmospheric Administration (NOAA) provides the mean number of days in a month with precipitation 0.01 inch or higher, which could control for seasonality.

Tables 2.1 and 2.2 summarize the variables used in the estimation and their descriptive statistics. *LRoadway*, *ElevDiff* and *LPrecip* represent the possibility of accident occurrences; *ElevDiff* and *LPrecip* represent the driving condition in each state. States have more incentive to adopt the universal MHU law for higher values of *LRoadway*, *ElevDiff* or *LPrecip*, and we expect the coefficients for these three covariates to be positive. *Population* and *Registered* represent the pressure of two

Table 2.2: Correlation between Variables

	LRoadway	*ElevDiff*	*LPrecip*	*Population*	*Registered*	*FatalRate*	*NbhdAvg*
LRoadway	1						
ElevDiff	0.094	1					
LPrecip	-0.188	-0.455	1				
Population	0.499	0.038	-0.084	1			
Registered	0.493	0.084	-0.147	0.844	1		
FatalRate	-0.319	0.031	0.001	-0.311	-0.262	1	
NbhdAvg	-0.103	-0.219	0.193	0.202	0.033	-0.224	1

conflicting opinions. Although they are naturally highly correlated, their effects on the choices of states are in the opposite direction. That is, as population grows, more car drivers are expected and so is higher pressure to introduce or maintain the universal MHU law. On the other hand, as the number of registered motorcycles rises, motorcyclists' opinion becomes more substantive. With regard to the state's decisions on the MHU law, the coefficient of *Population* is thus expected to be positive while that of *Registered* be negative.

The covariates *FatalRate* and *NbhdAvg* are of the particular interest. The fatality rate is expected to have a high positive impact on the hazard rate for "(-1 : 1)-failure." That is, if a state's safety concern is sufficiently substantial, it tends to have the universal MHU law when the previous fatality rate is high. Also, the effect of *NbhdAvg* is anticipated to be highly positive when the social interactions matters in this decision making. Note that the positive coefficient of *NbhdAvg* means that as more neighboring states adopt the universal MHU law, the higher hazard rate for "(-1 : 1)-failure" (i.e. revision of choice from -1 to 1) the state would encounter if it does not currently enforce the universal MHU law. At the same time, if the universal law is currently effective in the state, the hazard rate for "(1 : -1)-failure" (i.e. revision of choice from 1 to -1) decreases with the positive coefficient of *NbhdAvg*. In other words, the probability rate that the state would repeal or reduce the universal law becomes lower as more neighbors adopt the universal law.

2.5.2 Estimation results

For estimation, we assume the logistic specification for $G(\cdot)$ in (2.7). More precisely, conditional on $z_{i,t}$, we assume that $(\varepsilon_i(t,-1) - \varepsilon_i(t,1))/\sigma$ follows the independent type I extreme value distribution with variance $\sigma^2 > 0$ so that $G(z) = \exp(z/\sigma)/[1+\exp(z/\sigma)]$ for all $-\infty < z < \infty$. For the identification purpose, we assume the unit scale parameter ($\sigma = 1$.) In addition, since the decisions can be in both directions (i.e. 1 to -1 by repealing the law and -1 to 1 by adopting the law,) we specify the duration model as $\lambda_i\left(t|v_i\right) = \exp(\gamma_t)\phi_{d_{i,t-1}}(x'_{i,t}\beta + \alpha\sum_{j\neq i} w_{ij}d_{j,t-1})v_i$, where $\phi_{d_{i,t-1}}\left(z\right) = G(2z)$ if $d_{i,t-1} = -1$ and $1 - G(2z)$ if $d_{i,t-1} = 1$. Then the analysis becomes similar to a multi-spell duration analysis by letting the spell ends if any revision of the law is effective. In this case, the log-likelihood function is given by averages over all spells $k = 1, \ldots, K$ as

$$
\begin{aligned}
& \log L\left(\gamma, \theta|v\right) \\
= \quad & \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\substack{d_{i,T_i-1} \\ \in\{-1,1\}}} \delta_i(k)\log\left(1 - \exp\left(-\exp(\gamma_{\overline{T}_i(k)})\phi_{d_{i,\overline{T}_i(k)-1}}\left(x'_{i,\overline{T}_i(k)}\beta + \alpha W d_{\overline{T}_i(k)-1}\right)v_i\right)\right) \\
& -\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{s=\underline{T}_i(k)}^{\overline{T}_i(k)-1}\sum_{\substack{d_{i,s-1} \\ \in\{-1,1\}}} \exp\left(\gamma_s\right)\phi_{d_{i,s-1}}\left(x'_{i,s}\beta + \alpha W d_{s-1}\right)v_i, \qquad (2.22)
\end{aligned}
$$

where the k-th spell of individual i is from $\underline{T}_i(k)$ to $\overline{T}_i(k)$ and $\delta_i(k)$ is the censoring indicator of the spell k. In this context, we could understand the model as an alternating state model, where each spell does not rely on the history of events occurred before. Therefore, the interpretation of the sign of $(\alpha, \beta')'$ should be such that the positive values accelerate "(-1 : 1)-failure" that corresponds to the higher probability rate of switching from -1 to 1. Negative values, on the other hand, accelerate "type 1 failure" and decelerates the "type -1 failure" at the same time. It thereby represents the lower probability rate of switching from -1 to 1.

Table 2.3 summarizes the estimation results. The first two columns show the ML estimates obtained without unobserved heterogeneity. When *NbhdAvg* is omitted, all estimates except for *Elevdiff* are highly significant but the signs for *LRoadway* and *FatalRate* are in the opposite direction to what were expected. Signs of all the estimates except for *LPrecip* and *FatalRate* are as expected when the social interaction *NbhdAvg* is taken into account. Note that, however, the estimates for *LPrecip* and (especially) *FatalRate* are not significant at the 5% level, while the effects from social interactions is highly significant.

The last column of the table displays the estimates of the MPH model assuming two levels of heterogeneity (q_1, q_2) using the EM algorithm with the simulated annealing to find the global maximum. The values on the second column are used for the initial values. We set $\gamma_1 = 0$, whereas the values of q_1 and q_2, and accordingly the mean of v, are left unrestricted. Signs of all the estimates are as expected except for *LPrecip* though it is not significantly different from zero at the 5% level. For other covariates, *LRoadway* is the only significant variable that affects the decision of the MHU law. It suggests that the driving condition is not an important factor when the policy makers decide the level of the MHU law. However, as the total length of road grows, the state tends to have a stricter MHU law. Both parameters for covariates *Population* and *Registered*, which represent the pressure of two conflicting opinions, are significantly different from zero at the 5% significance level and are signed as expected.

Note that the sign of the estimates for *FatalRate* is as expected in the full model although it is not significant at the 5% level. However, the statistical insignificance does not mean that the MHU law is ineffective on the fatality rate. The estimation result in Table 2.3 only suggests that the previous fatality rate is not relevant when a state considers revision of the current MHU law. On the other hand, the estimates for *NbhdAvg* is significantly positive at the 5% level, which suggests that the policy makers in each state tend to make a parallel decision with neighboring states. In other words, though this analysis does not give an answer to a long-debated issue whether to introduce the mandatory motorcycle-helmet-use law, it explains a behavioral aspect of the legislative decision making procedure (i.e. social interactions) and empirically shows how the proximity between agents affects the decision making.

To show the magnitude of effect on hazard rate by each covariate, Table 2.4 reports the averaged elasticities of hazard rates for each covariates. For log-transformed

Table 2.3: Estimation Results

	No Unobs. Hetero. I	No Unobs. Hetero. II	With Unobs. Hetero.
LRoadway	-0.650**	0.354**	0.346**
	(0.059)	(0.073)	(0.057)
Elevdiff	3.549	3.243	1.759
	(2.170)	(2.134)	(2.099)
LPrecip	1.043**	-0.023	-0.599*
	(0.264)	(0.248)	(0.329)
Population	0.400**	0.255**	0.127**
	(0.015)	(0.067)	(0.022)
Registered	-1.579**	-1.519**	-0.705**
	(0.070)	(0.237)	(0.103)
FatalRate	-0.917**	-0.051	0.227
	(0.230)	(0.205)	(0.341)
NbhdAvg		4.124**	1.106**
		(0.327)	(0.232)
Base. Hazard: γ_2	-6.131**	-5.369**	-1.569**
	(0.016)	(0.013)	(0.358)
γ_3	-6.158**	-5.580**	-2.854**
	(0.248)	(0.488)	(0.563)
γ_4	-6.106**	-5.605**	-1.935**
	(0.010)	(0.023)	(0.352)
Unobs. Hetero.: q_1			0.036**
			(0.006)
q_2			0.000**
			(0.000)
p_1			0.688**
			(0.109)
Log-Likelihood	-767.588	-579.465	-238.370

Note: ML estimation is used for the case without unobserved heterogeneity, whereas EM method is used for the case with unobserved heterogeneity. Numbers in parentheses denote standard errors. * and ** represent significance at 10% and 5%, respectively.

covariates (*LRoadway* and *LPrecip*,) the values are obtained with respect to the percentage change of the original levels. For *NbhdAvg*, the values are obtained by the change in one of the neighbors' choices from -1 to 1 (e.g. Halvorsen and Palmquist, 1980). Note that elasticities do not depend on the baseline hazard by virtue of the proportional hazard specification. A 10% increase in total length of public roadway (*Roadway*) leads to 4.01% increase and 2.91% decrease in the hazard rate for "(-1 : 1)-failure" (i.e. likelihood to introduce the law) and "type 1 failure" (likelihood to repeal or reduce the law,) respectively. Same movements are found for *Precip*, *Population*. and *FatalRate*. The percentage change in the elevation difference (*Elevdiff*) would not induce much of percentage changes in hazard rates. On the other hand, 10% increase in the number of registered motorcycles (*Registered*) results in 9.12% decrease in hazard rate for "(-1 : 1)-failure" whereas 5.13% increase in that for "(1 : -1)-failure." Finally, when a neighbor changes its choice from -1 to 1, 169.9% increase and 37.9% decrease in hazard rate for "type -1 failure" and "(1 : -1)-failure" are obtained, respectively. It shows that, as the elasticity by a neighbor's choice turns out substantially high, each state indeed reacts to the neighbors' decisions very sensitively. However, the degree of sensitivity is not symmetric: pressure from social interactions with neighboring states is higher, on the average, toward the direction to adopting the universal law than the other way around.

Finally, the estimation result of the baseline hazard in the full model with unobserved heterogeneity (the last column of Table 2.3) shows that the piecewise baseline hazard estimate $(\exp(\widehat{\gamma}_1), \ldots, \exp(\widehat{\gamma}_4)) = (1.000, 0.208, 0.058, 0.144)$ are all highly significant. It well demonstrates the overall behavior of the states, which corresponds to the common trend of the historic changes of the federal regulations as summarized in the previous subsection. Note that the baseline hazard during the third period turns out to be substantially low, which shows that the incentives for states to enact the MHU law turned out to be not so effective.

Table 2.4: Elasticities of Hazard Rates

	Elasticity for "(-1 : 1)-Failure"	Elasticity for "(1 : -1)-Failure"
Roadway	0.401	-0.291
Elevdiff	0.103	-0.082
Precip	-0.694	0.504
Population	0.706	-0.555
Registered	-0.912	0.513
FatalRate	0.312	-1.046
NbhdAvg	1.699	-0.379

Note: Letting $\lambda_i[d] = \exp(\gamma_t)\phi_d(x'_{i,t}\beta + \alpha \sum_{j \neq i} w_{ij} d_{j,t-1})v_i$ for $d = -1, 1$, the table represents averaged estimates of the elasticities $\partial \log \widehat{\lambda}_i[d]/\partial \log x_{i,t}$ over the observations. Elasticities for *Roadway* and *Precip* are calculated with respect to the change of *Roadway* and *Precip* before taking logs. Elasticities for *NbhdAvg* is obtained as (e.g., for the case of "(-1: 1)-failure") $(\widehat{\lambda}_i^c[-1] - \widehat{\lambda}_i[-1])/\widehat{\lambda}_i[-1]$, where $\widehat{\lambda}_i^c[-1]$ is the counterfactual hazard rate when *one* of the neighbors changes from -1 to 1, whereas $\widehat{\lambda}_i[-1]$ is the original hazard rate estimate.

For the levels of heterogeneity, the higher level ($\widehat{q}_1$,) which corresponds to the states that have higher tendency to change the law, is much larger than the lower level ($\widehat{q}_2$) and the overall probability assigned to the former is about 2.2 times higher than that associated with the latter. As the estimates converge, the posterior probabilities ($\widehat{\pi}_{i1}, \widehat{\pi}_{i2}$) for each state i converges to the extreme values either 0 or 1. Figure 2.3 shows that the posterior probabilities correctly describe each state's behavior of changing the MHU law in that all states who have revised their law at least once are assigned

(a) Number of law changes

(b) Probability assigned to higher level of heterogeneity ($\widehat{q}_1$)

Figure 2.3: Law Changes and Heterogeneity

to higher level of unobserved heterogeneity and vice versa. Note that in contrast to the standard duration data, in which a high-risk group is likely to have early events, a high-risk group in the longitudinal case appears to have frequent events, which can be described the faster rate of the Poisson process with higher level of v_i.

2.6 Concluding Remarks

The mandatory MHU law is a long-debated issue. One concerning safety supports the universal MHU law (e.g. Houston and Richardson, 2007; Muller, 2004; Weiss, 1992), which is mainly backed by automobile drivers (i.e. non-motorcyclists) and insurance companies. The opposite opinion adheres to the idea that wearing helmet is a personal choice: Motorcyclists and liberalists insist that society's role is not to mandate personal safety but rather to provide the education and experience necessary to aid people in making these decisions for themselves. Moreover, even though it is a common wisdom that wearing helmet reduces the mortality in motorcycle accidents, medical evidences are still controversial (e.g. Cooter, McLean, David, and Simpson, 1988; Goldstein, 1986; Huston and Sears, 1981; Krantz, 1985; Stolzenberg and D'Alessio, 2003).

Though we aware such a long debate on the mandatory MHU law, we do not attempt to answer to this question in this chapter. The main point of this chapter is to find the evidence that the coverage of the MHU law in a state is spatially dependent on neighboring states' coverage. We develop a model analyzing states' decision on the coverage of the MHU law. Reflecting the fact that each decision making can be only realized at certain times, a hazard model naturally follows. In particular, we introduce a grouped mixed proportional hazard (MPH) model with social interactions, which serves as the econometric foundation of the states' choice model. Note that, however, this hazard model is different from the cross-sectional duration dependent models (e.g. Sirakaya, 2006) since the social interaction is based on others' discrete choice variables instead of the durations.

Based on the fact that not many revisions were made, we assume that the decisions are myopic. As a natural extension, a fully dynamic model incorporating a forward-looking behavior of agents is to be developed.

CHAPTER III

Estimation of Latent Dependent Variables in Dynamic Panel Data Models with Lagged Ranking as a Regressor: The Best States for Business

3.1 Introduction

Economic variables are sometimes reported by their ranking only, without revealing their values. Examples include rankings on companies' market shares, average prices, and sales, to name a few. It occurs more frequently on variables for marketing research including brand power ranking, top innovative companies, and so on. The approach presented in this chapter provides a methodology for latent variables with panel data in the presence of both lagged ranking and unobservable individual heterogeneity.

In cross-sectional settings, models with ranked data have used a probit or logit framework with constant or random utility specifications. They may extend to applications for panel data. Typically, they both need ranking observations on alternatives chosen by respondents. More precisely, each respondent (or consumer) is provided a set of alternatives (e.g. goods and brands) and she answers the order of her preference which forms a ranking observation for this particular respondent. With a limited number of alternatives and many respondents, a researcher can conduct an analysis in a framework with a limited dependent variable.

In this study, however, we analyze a panel data in which the rankings of alternatives are only observed over time and no respondent-level choices are available. In addition, the current latent dependent variable is affected by the revealed ranking in the previous period due to the "reputation" effect. For instance, the weekly ranking of NCAA college football teams is based on last week's ranking and the current week's game results or performances. The ranking is determined by polls but actual votes are not revealed. If one is interested in a team's (latent) "ability," she has to make an inference based on past rankings and performances, which are only observable.

Provided only information on output ranking rather than micro-level data, the difficulty arises with respect to estimation. Mainly, the number of data is not sufficient to estimate with precision by conventional probit or logit. In each time period, only one full ranking order is observed. This study gets around this problem by utilizing the exploding property of the logit model where each observation is exploded into several pseudo-observations for the purposes of estimation.

The remainder of the chapter is organized as follows. Section 3.2 introduces the annual ranking for business environment announced by Forbes magazine and addresses issues in econometrical modeling. Section 3.3 presents the econometric model for dynamic panel data with latent dependent variable in the presence of unobserved individual effects where lagged ranking among individuals enters as a regressor. Section 3.4 discusses how to estimate the model with simulation. Section 3.5 makes remarks on ties in the ranking. Section 3.6 discusses the estimation results, and section 3.7 is devoted to conclusions.

3.2 Ranking for Business Environment

3.2.1 Background

Since 2006, Forbes magazine has announced the annual ranking of the "best states for business." They declare that their ranking measures six vital categories for businesses: costs, labor supply, regulatory environment, current economic climate, growth prospects and quality of life (http://www.forbes.com/2009/09/23/best-states-for-business-beltway-best-states.html.) Also, according to them, business costs, which include labor, energy, and taxes are weighted the most heavily. We consider "business environment" as a latent output which takes an evolution mechanism relying on current economic climate and last year's ranking. The dependence on last year's ranking is twofold. On one hand, some factors cannot be easily changed over time including reserves of natural resources, arable land area, and quality of life. So previous ranking may be kept in the next period with regard to these factors. On the other hand, as Forbes remarked, the growth prospects are one of the main factors. Since current ranking reflects the prospected growth for next several years, it may well carry over the next period. For example, Forbes denoted top-ranked states in 2009 share an expanding, educated workforce. It said

> *"[t]he three states that followed [1st-ranked] Virginia in the rankings (Washington, Utah, and Colorado) also ranked in the top four along with Virginia in labor supply category, which looks at high school and college attainment, as well as net migration and projected population growth."*

(http://www.forbes.com/2009/09/23/best-states-for-business-beltway-best-states.html) So the projected expansion in population (or workforce) raises a state's ranking in current year and makes its business environment look better in next year as well.

3.2.2 Issues in econometric modeling

In this subsection, we address some issues in econometric modeling with the annual ranking of the best states for business.

First, we have a latent variable, namely business environment. What we observe is the ordinal ranking of business environment in each state along with observable business environment shifters, e.g. unemployment rate and population. In other words, business environment itself is a latent variable while we observe its relative positions so that a researcher can tell a state's business environment is better or worse than all the others for all states.

Second, provided only information on output ranking rather than micro-level data, the difficulty arises with respect to estimation. Mainly, the number of data is not sufficient to estimate with precision by conventional probit or logit. In each time period, only one full ranking order is observed. To utilize information as much as possible, we use the exploding property of the logit model where each observation is exploded into several pseudo-observations for the purposes of estimation. For example, suppose we observe a ranking ordered as $(A > B > C > D)$. Then we can extract the following three pieces of information (or pseudo-observations) from this single observation of ranking order: $(A > \{B, C, D\})$, $(B > \{C, D\})$, and $(C > D)$.

Third, as briefly introduced in the previous section, the reputation effect plays an important role in establishing business environment. We take this effect into account by using a lagged ranking as a regressor.

Fourth, each state has its own unobservable (to the researcher) individual heterogeneity. For example, residents' attitude to a certain company or industry in a state would not be observable.

Fifth, though it is not occurred in the ranking of the best states for business, ties in ranking make the estimation even more difficult. We make a remark on this issue more in detail in section 3.5.

3.3 The Model

3.3.1 Basic specification

Suppose we have a latent ("output") variable y^*_{it} that has a data generating process (DGP) as follows:

$$y^*_{i0} = c_i + \varepsilon_{i0}, \tag{3.1}$$

$$y^*_{it} = c_i + \alpha g(r_{i,t-1}) + x_{it}\beta + \varepsilon_{it}, \tag{3.2}$$

$$\varepsilon_t | c, r_{t-1}, \ldots, r_0, x \sim F(0, \Omega), \tag{3.3}$$

$$r_{it} = 1 + \sum_{j=1}^{n} I\{y^*_{it} < y^*_{jt}\}, \tag{3.4}$$

where we index individual and time as $i = 1, \ldots, n$ and $t = 1, \ldots, T$, respectively. Here c_i denotes the unobservable (to the researcher) individual heterogeneity, x_{it} is strictly exogenous regressors with length K_x, and ε_{it} is an unobserved error term. The vector of regressors throughout entire time periods is denoted as $x_i := (x_{i1}, \ldots, x_{iT})$. The definition (3.4) represents the *rank* of y^*_{it} at time t, which is observable instead of the latent variable y^*_{it}. The indicator function $I\{\cdot\}$ has value one if the expression in the parenthesis is true and zero otherwise. Notice that the "best" rank is 1 and n is the "worst." Hence $1 = r_{1t} = r_{2t} - 1 = r_{3t} - 2 = \ldots = r_{nt} - (n-1)$ if and only if $y^*_{1t} > y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt}$. If $r_{it} < r_{jt}$ ($r_{ij} > r_{jt}$, resp.), we say individual i is ranked *higher* (*lower*, resp.) than individual j or individual i is positioned *above* (*below*, resp.) individual j at time t.

In equation (3.2), the function $g(\cdot)$ accommodates the effect of the individual's realized ranking in the previous period, $r_{i,t-1}$, on her current (latent) output, y^*_{it}. This set up accounts for the "reputation" effect on the output variable. That is, individual i who performed well among competitors in the previous period is considered to have high valued output in the current period.[1] Notice that $g(r_{i,t-1})$ is exogenous to y^*_{it} since it is lagged.

To distinguish a case with no tie in ranking to others, we introduce a notion of *complete* ranking where no tie of ranking occurs. Formal definitions of complete and incomplete rankings follows.

Definition III.1. Denote $\mathcal{N}_R$ and $\mathcal{N}_I$ be the set of rankings and the set of individuals, respectively. A ranking is said to be *complete* if there exists a one-to-one function $\pi_t : \mathcal{N}_R \to \mathcal{N}_I$ for all $t = 0, \ldots, T$. Otherwise, it is *incomplete*.

Hence for any rankings, rank k goes to individual $\pi_t(k)$ at each time t. Only for a complete ranking there is a unique individual $\pi_t(k)$ at each time t. For now, we assume that the output ranking is complete and we will relax this assumption and discuss it in section 3.5.

Denote the systematic component of the DGP (3.2) as follows:

$$\mu_{it} := c_i + \alpha g(r_{i,t-1}) + x_{it}\beta \tag{3.5}$$

so that the output of individual i at time t is represented as $y^*_{it} = \mu_{it} + \varepsilon_{it}$. Consider a *ranking event*, $\{\pi_t(i) = i : i \in \mathcal{N}_I\}$ so that $y^*_{1t} > y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt}$ out of all possible $n!$ ranking events. Then we obtain a probability of a ranking event $y^*_{1t} > y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt}$ as

$$\begin{aligned} \mathbb{P}(\{\pi_t(i) = i : i \in \mathcal{N}_I\}) &= \mathbb{P}(y^*_{1t} > y^*_{2t}, y^*_{2t} > y^*_{3t}, \ldots, y^*_{n-1,t} > y^*_{nt}) \\ &= \mathbb{P}(y^*_{2t} - y^*_{1t} < 0, y^*_{3t} - y^*_{2t} < 0, \ldots, y^*_{nt} - y^*_{n-1,t} < 0). \end{aligned} \tag{3.6}$$

[1]Many ranking/rating schemes in sports take the previous ranking into account when they access the current ranking of teams or players. Examples include World Football Elo Ratings, a ranking system for men's national teams in association football. The factors taken into consideration when calculating a team's new rating are (i) the team's old rating, (ii) the considered weight of the tournament, (iii) the result of the match including the goal difference of it, and (iv) the expected result of the match.

Define a $(n-1) \times n$ transformation matrix

$$M_t = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & 0 & -1 & \ddots & 0 \\ & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \dots & 0 & -1 \end{pmatrix}$$

where the elements $M_t(k,k) = M_t(n,n) = -1$ and $M_t(k,k+1) = 1$ for $k = 1, \dots, n-1$, and all other entries are zero. Stack the individuals from 1 to n, so that the DGP (3.2) is represented for each time t in vector form as $y_t^* = \mu_t + \varepsilon_t$, where $\varepsilon_t \sim F(0, \Omega)$. Then we have

$$\begin{aligned} \mathbb{P}(\{\pi_t(i) = i : i \in \mathcal{N}_I\}) &= \mathbb{P}(M_t y_t^* < 0) \\ &= \mathbb{P}(M_t \mu_t + M_t \varepsilon_t < 0) \\ &= \mathbb{P}(M_t \varepsilon_t < -M_t \mu_t). \end{aligned}$$

In general, one can construct a transition matrix M_t such that $M_t(k, \pi_t(k)) = M_t(n, \pi_t(n)) = -1$, $M_t(k, \pi_t(k+1)) = 1$ for $k = 1, \dots, n-1$, and all the other elements are zero. Then the vector of error differences $M_t\varepsilon_t$ is distributed jointly a certain distribution with mean zero and variance-covariance matrix $M_t \Omega M_t'$.

The problem occurs when one tries to estimate the model (3.1)-(3.4) with a typical panel data set where n is large and T is small. The full ranking orders are observed only T times while the number of possible ranking events are enormous: $n!$ orderings at each time. As n gets large, it becomes in general almost impossible to estimate the parameters with precision. This study gets around this problem by utilizing the exploding property of logit model where each observation is exploded into several pseudo-observations for the purposes of estimation.

3.3.2 Mixed exploded logit

Let us assume that ε_{it} are *iid* with a type I extreme value distribution, given by $P(\varepsilon_{it} \leq x) = \exp(-\exp(x))$. The odds that individual i is ranked higher than j ($y^*_{it} > y^*_{jt} \iff r_{it} < r_{jt}$) is given by $\exp(\mu_{it} - \mu_{jt})$. For notational ease, denote $g_{it} := g(r_{it})$. We again take a ranking event $r_t = \{\pi_t(i) = i : i \in \mathcal{N}_I\}$ ($y^*_{1t} > y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt}$) into consideration. Provided a ranking event, consider a *subranking event* by excluding the highest value: $y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt}$ in this case. Then we obtain a probability of a ranking event $y^*_{1t} > y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt}$ as

$$\mathbb{P}(y^*_{1t} > y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt}) = \mathbb{P}(y^*_{1t}|\mathcal{N}_I) \cdot \mathbb{P}(y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt}), \tag{3.7}$$

where $\mathbb{P}(y^*_{it}|\mathcal{S})$ represents the probability that the output of individual i is ranked first among those of elements in a set $\mathcal{S}$ at time t. Luce and Suppes (1965) provides this decomposition theorem for constant utility model. See Chapman and Staelin (1982) for application within the random utility model. As remarked in Chapman and Staelin (1982), the probability of a ranking event, $\mathbb{P}(y^*_{1t} > y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt})$, is decomposed into the product of the probability of a *top-ranking event*, $\mathbb{P}(y^*_{1t}|\mathcal{N}_I)$, and the probability of a subranking event, $\mathbb{P}(y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt})$. By successively applying the decomposition (3.7) to the subranking events, we can rewrite the decomposition (3.7) with $n-1$ top-ranking events such that

$$\begin{aligned}&\mathbb{P}(y^*_{1t} > y^*_{2t} > y^*_{3t} > \ldots > y^*_{nt})\\&\quad = \mathbb{P}(y^*_{1t}|\mathcal{N}_I) \cdot \mathbb{P}(y^*_{2t}|\mathcal{N}_I\backslash\{1\}) \ldots \mathbb{P}(y^*_{n-1,t}|\mathcal{N}_I\backslash\{1, \ldots, n-2\}).\end{aligned} \tag{3.8}$$

Then the probability of a top-ranking event that individual 1 is ranked first from

$\mathcal{N}_I$ given regressor values is that

$$\begin{aligned}
&\mathbb{P}(y_{1t}^*|\mathcal{N}_I, c, g_{t-1}, x_t) \\
&\quad = \mathbb{P}\left(\sum_{j=1}^{n} I\{y_{1t}^* < y_{jt}^*\} = 0 \middle| c, g_{t-1}, x_t\right) \\
&\quad = \mathbb{P}\left(y_{1t}^* > y_{jt}^*, \forall j \neq 1 \middle| c, g_{t-1}, x_t\right) \\
&\quad = \frac{\exp(\mu_{1t})}{\sum_{j=1}^{n} \exp(\mu_{jt})}.
\end{aligned} \tag{3.9}$$

See Anderson, de Palma, and Thisse (1992, pp.39∼40) for proof of last equality. Similarly, given that y_{1t}^* is the greatest among all outputs, the probability that y_{2t}^* is the greatest among outputs of remaining individuals ($y_{2t}^* > y_{jt}^*$, for all $j \in \mathcal{N}_I \backslash \{1\}$) is then

$$\begin{aligned}
&\mathbb{P}(y_{2t}^*|\mathcal{N}_I \backslash \{1\}, c, g_{t-1}, x_t) \\
&\quad = \mathbb{P}\left(\sum_{j=1}^{n} I\{y_{2t}^* < y_{jt}^*\} = 1 \middle| y_{2t}^* < y_{1t}^*, c, g_{t-1}, x_t\right) \\
&\quad = \mathbb{P}\left(\sum_{j \neq 1} I\{y_{2t}^* < y_{jt}^*\} = 0 \middle| y_{2t}^* < y_{1t}^*, c, g_{t-1}, x_t\right) \\
&\quad = \mathbb{P}\left(y_{2t}^* > y_{jt}^*, \forall j \neq 1, 2 \middle| y_{2t}^* < y_{1t}^*, c, g_{t-1}, x_t\right) \\
&\quad = \frac{\exp(\mu_{2t})}{\sum_{j=1}^{n} \exp(\mu_{jt}) - \exp(\mu_{1t})}.
\end{aligned} \tag{3.10}$$

So proceeding from probability (3.9) to (3.10), we observe that the numerator remains as the exponential of current rank's systematic component while the denominator is decreased by the numerator of individual who ranked immediately above. Finally, for the last two outputs, we obtain

$$
\begin{aligned}
&\mathbb{P}(y^*_{n-1,t}|\mathcal{N}_I\backslash\{1,\dots,n-2\}, c, g_{t-1}, x_t)\\
&\quad = \mathbb{P}\left(y^*_{n-1,t} > y^*_{nt} \middle| y^*_{1t} > \dots > y^*_{n-2,t}, c, g_{t-1}, x_t \right)\\
&\quad = \frac{\exp(\mu_{n-1,t})}{\sum_{j=1}^{n}\exp(\mu_{jt}) - \sum_{k=1}^{n-2}\exp(\mu_{kt})}\\
&\quad = \frac{\exp(\mu_{n-1,t})}{\exp(\mu_{n-1,t}) + \exp(\mu_{nt})}.
\end{aligned}
\tag{3.11}
$$

Therefore substituting the probabilities of $n-1$ top-ranking events into (3.8), one can generally obtain the density of a ranking event $r_t = (r_{1t}, \dots, r_{nt})$ given (c, g_{t-1}, x_t) is

$$
f_t(r_t|c, g_{t-1}, x_t) = \prod_{i=1}^{n} \left\{ \frac{\exp(\mu_{it})}{\sum_{j=1}^{n} \delta_{ijt} \exp(\mu_{jt})} \right\}, \tag{3.12}
$$

where $\delta_{ijt} = 1$ if $r_{it} \leq r_{jt}$ or $i = j$ is ranked lowest with no tie, and 0 otherwise.

Equation (3.12) implies that a ranking event of n outputs can be represented as being the same as $n-1$ independent orderings between two individual outputs, also known as *pseudo-observations*. So a ranking event at each time period constitutes an observation and is written as if it were multiple pseudo-observations. For this explosion into multiple pseudo-observations for the purposes of estimation, a logit model on ranked observations is called an *exploded logit*. Salient applications of exploded logit models include Beggs, Cardell, and Hausman (1981), Chapman and Staelin (1982), and Hausman and Ruud (1987) to name a few.

Under the assumptions that the dynamics of the conditional distributions are correctly specified and $x_i = (x_{i1}, \dots, x_{iT})$ is strictly exogenous conditional on c_i, the joint density of $(r_1, \ldots, r_T)$ given (r_0, x, c) is given by

$$f(r_1, \ldots, r_T | r_0, x, c) = \prod_{t=1}^{T} f_t(r_t | c, g_{t-1}, x_t; \theta) \tag{3.13}$$

To construct the likelihood function, it is necessary to integrate the joint density (3.13) over the distribution of the unobserved individual heterogeneity c_i. We specify the density of c given (r_{i0}, x_i) with parameter vector η as $h(c|r_{i0}, x_i; \eta)$. Notice that

$$\begin{aligned}\log f(r_1, \ldots, r_T | r_0, x, c) &= \sum_{t=1}^{T} \sum_{i=1}^{n} \mu_{it} - \sum_{t=1}^{T} \sum_{i=1}^{n} \log \left(\sum_{j=1}^{n} \delta_{ijt} \exp(\mu_{jt}) \right) \\ &= \sum_{i=1}^{n} \sum_{t=1}^{T} \mu_{it} - \sum_{i=1}^{n} \sum_{t=1}^{T} \log \left(\sum_{j=1}^{n} \delta_{ijt} \exp(\mu_{jt}) \right). \end{aligned} \tag{3.14}$$

Given that $h(c|r_{i0}, x_i; \eta)$ is correctly specified, the log-likelihood function is given by

$$\begin{aligned} & l(\theta, \eta | r_0, \ldots, r_T, x_1, \ldots, x_n) \\ & \quad = \sum_{i=1}^{n} \int \left[\sum_{t=1}^{T} \mu_{it} \right] h(c|r_{i0}, x_i; \eta) dc \\ & \qquad - \sum_{i=1}^{n} \int \left[\sum_{t=1}^{T} \log \left(\sum_{j=1}^{n} \delta_{ijt} \exp(\mu_{jt}) \right) \right] h(c|r_{i0}, x_i; \eta) dc \end{aligned} \tag{3.15}$$

where θ denotes all the parameters except those for c. Since the integral in the likelihood function (3.15) does not have a closed form solution, we resort on simulation for estimation. We will revisit issues in estimation in section 3.4.

3.4 Estimation with Simulation

Throughout the rest of this study, we assume that the individual heterogeneity c_i is specified as

$$c_i = \eta_0 + \eta_1 r_{i0} + z_i \eta_2 + \nu_i \tag{3.16}$$

where z_i is an individual-specific regressors with length K_z, $\eta = (\eta_0, \eta_1, \eta_2)$ is a length $(K_z + 2)$ parameter vector, and $\nu_i | r_{i0}, z_i \sim iid\, \mathcal{N}(0, \sigma_\nu^2)$. Denote $\Phi(\cdot)$ and $\phi(\cdot)$ be the cdf and pdf of the standard normal distribution, respectively. Given the equivalency in their role, we denote $w_{it} := (g_{i,t-1}, x_{it})'$ and $\gamma := (\alpha, \beta)$ for further notational ease.

Assume that ε_{i0} is *iid* with standard logistic distribution, and no tie was observed in the data set. Then the log-likelihood (3.15) is rewritten as

$$\begin{aligned} & l(\theta, \eta | r_0, \ldots, r_T, x_1, \ldots, x_n) \\ & \quad = nT\eta_0 + T\left(\eta_1 \sum_{i=1}^{n} r_{i0} + \eta_2 \sum_{i=1}^{n} z_i\right) + \sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}\gamma \\ & \qquad - \sum_{i=1}^{n} \int \left[\sum_{t=1}^{T} \log\left(\sum_{j=1}^{n} \delta_{ijt} \exp(\mu_{jt})\right)\right] \frac{1}{\sigma_\nu} \phi\left(\frac{\nu}{\sigma_\nu}\right) d\nu \end{aligned} \tag{3.17}$$

where $\theta := (\gamma, \sigma_\nu^2)$. As denoted at the end of section 3.3.1, it is necessary to simulate the each value of integral for each individual in the likelihood (3.17). It is approximated through simple simulation steps for any given value of $(\theta, \eta, \sigma_\nu^2)$: At each period of simulation s, (i) draw a value of ν_0 from a standard normal distribution and label it ν_0^s, (ii) multiply ν_0 by σ_ν to obtain $\nu^s = \sigma_\nu \nu_0^s$, and (iii) calculate the probability in the square brackets with this adjusted draw ν^s. Then (iv) repeat steps i through iii S times. The average becomes the value of integral except $1/\sigma_\nu$ which can be pulled out of integral. That is,

$$
\begin{aligned}
&\int \left[\sum_{t=1}^{T} \log\left(\sum_{j=1}^{n} \delta_{ijt} \exp(\mu_{jt})\right)\right] \phi\left(\frac{\nu}{\sigma_\nu}\right) d\nu \\
&\qquad \simeq \frac{1}{S}\sum_{s=1}^{S}\left[\sum_{t=1}^{T} \log\left(\sum_{j=1}^{n} \delta_{ijt} \exp(\eta_0 + \eta_1 r_{j0} + z_j\eta_2 + w_{jt}\gamma + \nu^s)\right)\right]. \qquad (3.18)
\end{aligned}
$$

The simulated log-likelihood function is given by

$$
\begin{aligned}
&l^{\mathrm{sim}}(\theta, \eta | r_0, \ldots, r_T, x_1, \ldots, x_n) \\
&\quad = nT\eta_0 + T\left(\eta_1 \sum_{i=1}^{n} r_{i0} + \eta_2 \sum_{i=1}^{n} z_i\right) + \sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}\gamma \\
&\qquad - \frac{1}{\sigma_\nu}\sum_{i=1}^{n}\frac{1}{S}\sum_{s=1}^{S}\left[\sum_{t=1}^{T}\log\left(\sum_{j=1}^{n}\delta_{ijt}\exp(\eta_0 + \eta_1 r_{j0} + z_j\eta_2 + w_{jt}\gamma + \sigma_\nu \nu_0^s)\right)\right].
\end{aligned}
\qquad (3.19)
$$

3.5 Incomplete Rankings

In this section, we investigate the case when several individuals are observed as being tied. When $r_{it} = r_{jt}$ for some individuals i and j ($i \neq j$) at time t, it is equivalent to $y^*_{it} = y^*_{jt}$. However, with continuous distribution F it is an event of probability zero. So we consider a threshold B, to be estimated, defined as follows:

$$
r_{it} = r_{jt} \iff |y^*_{it} - y^*_{jt}| \leq b \qquad (3.20)
$$

for all i, j, t, and $i \neq j$. In other words, the observed ranking order is distinguished only when two latent outputs are sufficiently different. Therefore, a tie further reduces information on the latent output. Following the notation in section 3.3.1, we represent an allocation of ranking by a mapping $\widetilde{\pi}_t : \mathcal{N}_R \mapsto \mathcal{N}_I$. Hence rank k goes to individual $\widetilde{\pi}_t(k)$, possibly multiple values, at each time t.

Consider the standard competition ranking, also known as "1224" ranking. That is, individuals that have output value less than b apart receive the same ranking number, and then a gap is left in the ranking numbers. The number of ranking numbers that are left out in this gap is one less than the number of individuals that have equal value of output. If two (or more) individuals tie for a position in the ranking, the position of all those ranked below them is unaffected. In our notation, if $\#\widetilde{\pi}_t(k) = m > 1$, ranks $k+1, \ldots, k+m-1$ are left out and thus $\widetilde{\pi}_t(k+1) = \ldots = \widetilde{\pi}_t(k+m-1) = \varnothing$. In this scheme, ranking is newly defined as

$$\widetilde{r}_{it} = 1 + \sum_{j=1}^{n} I\{y^*_{it} < y^*_{jt} - b\}. \tag{3.21}$$

Notice that $\widetilde{r}_{it} = r_{it}$ when $b = 0$.

For propositional purpose, assume that $\widetilde{\pi}_t(i) = \{i, i+1\}$ and $\widetilde{\pi}_t(k) = k$ for $k = 1, \ldots, i-1, i+2, \ldots, n$. That is, individuals are sorted by their output values in descending order, and individuals i and $i+1$ are tied. In this ordering, observing $\widetilde{r}_{it} = \widetilde{r}_{i+1,t} = i$ is the outcome of two possibilities: (A) outcome of individual i is greater than that of $i+1$ and the difference is less than b $(0 < y^*_{it} - y^*_{i+1,t} < b)$ or (B) outcome of individual i is smaller than that of $i+1$ and the difference is less than b $(0 < y^*_{i+1,t} - y^*_{it} < b)$. Both events A and B are given that $y^*_{1t}, \ldots, y^*_{i-1,t}$ are greater than $y^*_{it} + b$ and $y^*_{i+1,t} + b$, and $y^*_{i+2,t}, \ldots, y^*_{nt}$ are smaller than $y^*_{it} - b$ and $y^*_{i+1,t} - b$. Since events A and B are mutually exclusive, we have

$$\begin{aligned}
\mathbb{P}(\widetilde{\pi}_t(i) = \{i, i+1\} | \mathcal{N}_I \backslash \{1, \ldots, i-1\}) &= \mathbb{P}(y^*_{it}, y^*_{i+1,t} | \mathcal{N}_I \backslash \{1, \ldots, i-1\}) \\
&= \mathbb{P}(A \text{ or } B) \\
&= \mathbb{P}(A) + \mathbb{P}(B).
\end{aligned} \tag{3.22}$$

By definition, the probability $\mathbb{P}(A)$ is further decomposed into two probabilities: $(A1)$

probability that $y^*_{it} - y^*_{i+1,t} > 0$ and $(A2)$ probability that $y^*_{it} - y^*_{i+1,t} < b$ given that the ordering of other individuals' outcomes are set as above and they are sufficiently apart from y^*_{it} and $y^*_{i+1,t}$.

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A1) \cdot \mathbb{P}(A2) \\
&= \mathbb{P}(y^*_{it} > y^*_{i+1,t} > y^*_{i+2,t} + b > \ldots > y^*_{nt} + b) \\
&\quad \times \mathbb{P}(y^*_{i+1,t} + b > y^*_{it} > y^*_{i+2,t} + b > \ldots > y^*_{nt} + b) \\
&= \left(\frac{\exp(\mu_{it})}{\exp(\mu_{it}) + \exp(\mu_{i+1,t}) + \sum_{j=i+2}^{n} \exp(\mu_{jt} + b)} \right) \\
&\quad \times \left(\frac{\exp(\mu_{i+1,t} + b)}{\exp(\mu_{it}) + \sum_{k=i+1}^{n} \exp(\mu_{kt} + b)} \right). \qquad (3.23)
\end{aligned}$$

We can derive $\mathbb{P}(B)$ in a similar manner. So the ith term in the tth component of the likelihood function is

$$\begin{aligned}
&\mathbb{P}(A) + \mathbb{P}(B) \\
&= \left(\frac{e^{\mu_{it}}}{e^{\mu_{it}} + e^{\mu_{i+1,t}} + \sum_{j=i+2}^{n} e^{\mu_{jt}+b}} \right) \left(\frac{e^{\mu_{i+1,t}+b}}{e^{\mu_{it}} + e^{\mu_{i+1,t}+b} + \sum_{j=i+2}^{n} e^{\mu_{jt}+b}} \right) \\
&\quad + \left(\frac{e^{\mu_{i+1,t}}}{e^{\mu_{it}} + e^{\mu_{i+1,t}} + \sum_{j=i+2}^{n} e^{\mu_{jt}+b}} \right) \left(\frac{e^{\mu_{it}+b}}{e^{\mu_{it}+b} + e^{\mu_{i+1,t}} + \sum_{j=i+2}^{n} e^{\mu_{jt}+b}} \right)
\end{aligned}$$

In general, one can derive for an observed $\widetilde{\pi}_t(k)$

$$\begin{aligned}
&\mathbb{P}(\widetilde{\pi}_t(k) | \mathcal{N}_I \backslash (\widetilde{\pi}_t(1) \cup \ldots \cup \widetilde{\pi}_t(k-1))) \\
&= \sum_{i=\widetilde{\pi}_t(k)} \Bigg\{ \left(\frac{\exp(\mu_{it})}{\sum_{m \in \widetilde{\pi}_t(k)} \exp(\mu_{mt}) + \sum_{s=\widetilde{\pi}_t(k+1) \cup \ldots \cup \widetilde{\pi}_t(n)} \exp(\mu_{st} + b)} \right) \times \\
&\prod_{j \in \widetilde{\pi}_t(k) \backslash \{i\}} \left(\frac{\exp(\mu_{jt} + b)}{\exp(\mu_{it}) + \sum_{l \in \widetilde{\pi}_t(k) \backslash \{i\}} \exp(\mu_{lt} + b) + \sum_{s=\widetilde{\pi}_t(k+1) \cup \ldots \cup \widetilde{\pi}_t(n)} \exp(\mu_{st} + b)} \right) \Bigg\}
\end{aligned} \qquad (3.24)$$

So the contribution of an output observed tied with some outputs is its probability of being ranked highest among outputs where higher ranked outputs are censored out multiplied by each probability of being close to an output tied with it. Notice that those ranked lower than an output enter with the threshold b considered. Hence the likelihood function for a single time period t is given by

$$\prod_{k=1}^{n} \sum_{i=\widetilde{\pi}_t(k)} \left\{ \left(\frac{\exp(\mu_{it})}{\sum_{m\in\widetilde{\pi}_t(k)} \exp(\mu_{mt}) + \sum_{s=\widetilde{\pi}_t(k+1)\cup\ldots\cup\widetilde{\pi}_t(n)} \exp(\mu_{st}+b)} \right) \times \right.$$
$$\left. \prod_{j\in\widetilde{\pi}_t(k)\setminus\{i\}} \left(\frac{\exp(\mu_{jt}+b)}{\exp(\mu_{it}) + \sum_{l\in\widetilde{\pi}_t(k)\setminus\{i\}} \exp(\mu_{lt}+b) + \sum_{s=\widetilde{\pi}_t(k+1)\cup\ldots\cup\widetilde{\pi}_t(n)} \exp(\mu_{st}+b)} \right) \right\} \tag{3.25}$$

When no tie is allowed, i.e. $b = 0$ and $\#\widetilde{\pi}_t(k) = 1$ for all $k = 1, \ldots, n$, this expression reduces to equation (3.12) derived in the previous section. On the other hand, when all the outputs are tied, i.e. $b \to \infty$ and $\widetilde{\pi}_t(1) = \mathcal{N}_I$, it collapses to 1 and provides no information on parameters.

The likelihood function for estimation would be the product across time periods of the likelihood for a single time period in equation (3.25). However it is very computationally intensive to estimate. Instead, one might think of an approximation for the likelihood. Similar attempts to approximate the likelihood function with tied rankings unfortunately have been shown to be inaccurate when the number of tied outputs is substantial fraction of outputs at that rank or higher. (See Farewell and Prentice (1980) and Kalbfleisch and Prentice (2002) for details.) Moreover, when we think of a mixed model where μ_{it}'s are stochastic, the inaccuracy even becomes worse.

3.6 Estimation Results and Discussion

3.6.1 Data

We use the 5 different annual rankings announced 2006 through 2010. Regressors are selected to represent each state's current environment and growth prospect. For individual heterogeneity, two covariates are introduced: The average bond rating for state governments (*BondRating*) and number of cities with population of 100,00 or more in a state (*NBigCity*.) The covariate *BondRating* is obtained by taking average of three major indices by Standard & Poor's, Moody's, and Fitch over three years (2005-2007.) It represents both current economic environment and expectations for future climate of a state. Average is taken since three indices are almost identical for a state and they do not change significantly over time. Based on the well-known equivalency across indices (e.g. AA+ of S&P is equivalent to Moody's and Fitch Aa1) they are quantified from 1 to 9. As shown in Table 3.1, all of the states' ratings are higher than equal to 7, the lowest rating for investment grades. The covariate *NBigCity* is obtained from U.S. Census Bureau and it represents the degree of urbanization. One can expect the business environment is nicer in (or near) a large city than rural areas owing to its high demand and labor supply.

The regressors x_{it} include population, household median income, energy price, gross state product (GSP), unemployment rate, and tax environment. First two covariates explains the demand side of each state. As noted above, the business climate gets more favorable in general as population increases. However, a diminishing pattern would be revealed when population reaches a certain amount. To accommodate the diminishing marginal effect of population to business climate, population is included in log scale ($LPopulation$.) Household median income (*MedInc*) denotes the purchasing power of consumers. In the cost side, each state's total energy average price (*EnPrice*) level is used. This price index includes the cost for electricity, gas, and so on. To explain each state's general economic environment, three covariates are finally included: GSP, unemployment rate, and tax environment. Each state's real GSP (*RGSP*) and unemployment rate (*UnempRate*) are indicators representing its economic performance. Note that the per capita GSP is not used to avoid the multicollinearity problem since it may well be highly correlated with household median income.

A corporate tax index provided by Tax Foundation (*CorpTaxEnv*) is a score showing how favorable a state's corporate tax system is for business. The higher the score, the more favorable the system is for business. Using the corporate tax rate directly would not be preferable in two reasons: First, several states hold multiple brackets and arbitrary choice from those rates (e.g. averaging over rates or taking the maximum rate) would not precisely pick a representative value. Second, what essentially we want to include here is the incentive/disincentive that a state government offers with regard to corporate tax, a representative measure of state's regulation. When merely the tax rate is used, any sort of incentives are excluded. Even though some incentives are represented by tax rate, they are typically in a special form of rates (e.g. exemption when a certain condition holds.) Then one still needs to create a proxy index for tax rates.

Table 3.1 displays the definitions of covariates and their summary statistics. The data sources are denoted at the rightmost column.

3.6.2 Estimation results

The function for lagged ranking is chosen to be $g(x) = -\sqrt{x}$ thereby the coefficient for ranking is expected to be positive and the gap between y_i^* and y_j^* are getting bigger as two states i and j are ranked higher. The number of simulation is set to be $S = 500$. Table 3.2 reports the estimation results.

As expected, the estimated parameter for the previous ranking, $\widehat{\alpha}$, is obtained to be positive and is highly significant. Among covariates, state's energy average price (*EnPrice*) and unemployment rate (*UnempRate*) are significant with expected direction. That is, when a state faces high cost of production, its business environment becomes harsh. Similarly, a state's business climate gets worse as unemployment rate rises. The only other significant covariate is the number of big cities in a state

Table 3.1: Definitions of Variables and Summary Statistics

Variable	Description	Mean	Std.Dev.	Min	Max	Source*
BondRating	Bond ratings for state governments in scale from 1 to 9	8.147	0.437	7	9	SMF
NBigCity	Number of cities with populations of 100,000 or more in a state	4.760	8.351	0	56	CSS
LPopulation	State's population in log scale	15.132	1.013	13.148	17.426	CSS
MedInc	State's household median income in 2009 dollars divided by 10,000	5.096	0.781	3.508	7.240	CSS
EnPrice	State's total energy average price ($ per million BTU)	18.276	3.556	11.136	36.211	EIA
RGSP	Real gross state product divided by a million	0.246	0.299	0.019	1.8	BEA
UnempRate	State's unemployment rate	0.056	0.022	0.025	0.133	BLS
CorpTaxEnv	State's corporate tax index	5.298	1.349	3.292	10	TXF

Note: *Data Source: SMF = S&P, Moody's, and Fitch, CSS = U.S. Census Bureau, EIA = Energy Information Administration, BEA = U.S. Bureau of Economic Analysis, BLS = U.S. Bureau of Labor Statistics, TXF = Tax Foundation.

Table 3.2: Estimation Results

Variable	Parameter	Estimate	Std.Err.
Constant	η_0	-23.713	21.270
Initial Ranking	η_1	32.741*	2.440
BondRating	η_{21}	-46.192	23.683
NBigCity	η_{22}	-35.705*	12.880
Previous Ranking	α	84.352*	23.622
LPopulation	β_1	15.985	12.199
MedInc	β_2	-1.461	0.910
EnPrice	β_3	-71.094*	12.199
RGSP	β_4	-0.788	2.596
UnempRate	β_5	-94.492*	20.707
CorpTaxEnv	β_6	-21.546	13.247
Std.Dev.	σ_ν	0.100*	0.001

Note: *represents significant at 5% level.

(*NBigCity.*) It's influence, however, is in the opposite direction than expected. Although it's not significant, the population in a state affects in a parallel way: the higher the population, the better the business environment is. From these two findings, a possible interpretation is what really matters to a state's business environment is not how many mid-size cities it has but whether it has a metropolitan area with highly dense population or not.

Noticeably, each state's climate with respect to the corporate tax (*CorpTaxEnv*) appears to be insignificant. It is because a favorable tax environment may represent both good and bad climate for business. On one hand, tax incentives attract more business into a state. Firms can relocate their headquarters or manufacturing facilities to a state that exempt or reduce the corporate tax. On the other hand, they may be signs that a state struggles with unfavorable business environment and attempts to boost business: when a state observes a harsh environment it may introduce a tax incentive to attract more business and to change the climate. Due to these two conflicting implications, *CorpTaxEnv* hardly influences the business environment of a state as obtained above.

Table 3.3: Predicted Percent Changes in Business Environment

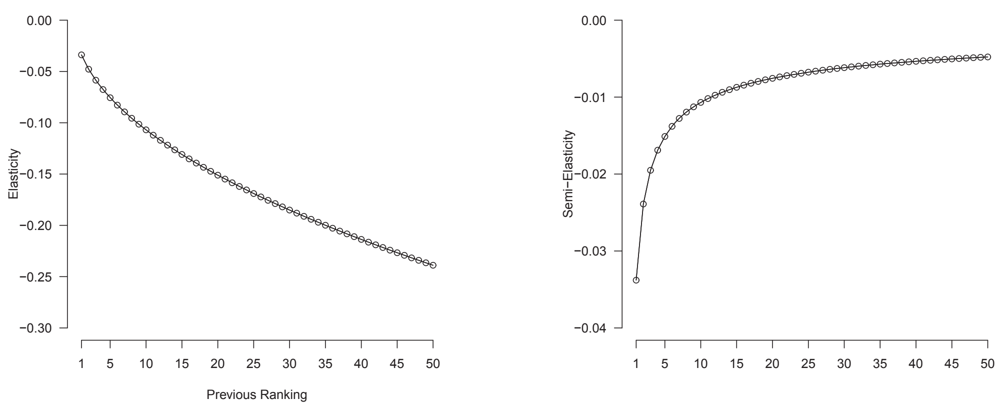
	2007	2008	2009	2010		2007	2008	2009	2010
Alabama	-37.52%	-11.48%	-4.10%	-32.17%	Montana	-40.47%	-20.21%	9.59%	-14.17%
Alaska	-126.21%	-50.26%	-17.68%	-57.99%	Nebraska	-14.40%	-15.51%	1.64%	-8.12%
Arizona	-11.73%	-9.64%	-4.30%	-19.02%	Nevada	-13.29%	-7.98%	-3.74%	-17.52%
Arkansas	-23.57%	-7.43%	-15.27%	-15.99%	New Hampshire	-16.26%	-13.39%	-10.21%	-13.39%
California	-6.36%	-4.36%	-3.47%	-6.86%	New Jersey	-15.33%	-16.41%	-11.16%	-21.64%
Colorado	-12.90%	-10.55%	1.10%	-8.85%	New Mexico	-27.54%	-10.61%	4.78%	-34.87%
Connecticut	-22.54%	-19.96%	-10.11%	-15.71%	New York	-26.95%	-12.42%	-7.41%	-14.67%
Delaware	-14.14%	-18.25%	-6.18%	-18.62%	North Carolina	-12.04%	-6.91%	-5.99%	-11.02%
Florida	-11.05%	-8.55%	-3.53%	-16.54%	North Dakota	-16.18%	-1.36%	-14.62%	-8.25%
Georgia	-15.48%	-9.33%	4.51%	-17.64%	Ohio	-25.86%	-16.77%	-4.64%	-14.88%
Hawaii	-32.62%	-17.07%	2.69%	-72.79%	Oklahoma	-18.28%	-18.54%	-0.16%	-15.46%
Idaho	-12.22%	-7.88%	-5.32%	-14.57%	Oregon	-19.26%	-10.38%	1.33%	-10.82%
Illinois	-33.11%	-9.12%	-9.00%	-13.39%	Pennsylvania	-35.89%	-17.63%	-8.18%	-18.82%
Indiana	-28.22%	-7.22%	-2.84%	-27.00%	Rhode Island	-35.39%	-32.27%	-1.80%	-26.43%
Iowa	-21.91%	-9.30%	-1.81%	-5.23%	South Carolina	-22.99%	-5.73%	-12.56%	-18.40%
Kansas	-17.10%	-7.15%	-6.68%	-10.97%	South Dakota	-14.90%	-14.59%	-2.74%	-4.99%
Kentucky	-32.56%	-23.37%	-10.98%	-22.87%	Tennessee	-17.11%	-4.32%	-8.68%	-21.09%
Louisiana	43.64%	-21.00%	-10.69%	-69.14%	Texas	-11.82%	-7.79%	-6.31%	-10.22%
Maine	-71.14%	-45.54%	-17.81%	-12.30%	Utah	-12.71%	-6.64%	-2.71%	-11.77%
Maryland	-16.48%	-15.70%	-7.77%	-12.77%	Vermont	-22.41%	-15.65%	-12.24%	-25.17%
Massachusetts	-25.29%	-24.25%	-3.93%	-16.65%	Virginia	-10.70%	-6.99%	-3.49%	-13.61%
Michigan	-37.84%	-26.99%	-8.87%	-18.43%	Washington	-10.84%	0.32%	-3.21%	-12.87%
Minnesota	-15.60%	-5.44%	-6.46%	-17.78%	West Virginia	-9.64%	-49.69%	-17.45%	-54.13%
Mississippi	-87.39%	-19.13%	-3.75%	-49.73%	Wisconsin	-33.16%	-22.07%	-9.53%	-21.04%
Missouri	-16.62%	-5.01%	-15.87%	-11.74%	Wyoming	-23.85%	-16.70%	-3.96%	-12.00%
National Average	-23.87%	-14.88%	-6.03%	-20.40%					

Note: Each value represents the predicted percent change in business environment of each state from the previous year.

Table 3.4: Elasticities of Business Environment

Variable	Parameter	Elasticity
Initial Ranking	η_1	1.5620*
BondRating	η_{21}	-0.4902
NBigCity	η_{22}	-0.1364*
Previous Ranking	α	0.3237*
LPopulation	β_1	0.3185
MedInc	β_2	-0.0092
EnPrice	β_3	-1.4089*
RGSP	β_4	-0.0002
UnempRate	β_5	-0.0071*
CorpTaxEnv	β_6	-0.1493

Note: *represents that the parameter estimate is significant at 5% level.



(a) Elasticities (b) Semi-elasticities

Figure 3.1: Previous Ranking Elasticities of Business Environment

Note: Elasticities are evaluated at the mean value of predicted business environment.

Table 3.3 shows the predicted percent change in business environment from the previous year for each state using the estimates obtained in Table 3.2. Overall, we find falls of predicted business environment in most states, particulary due to the financial hardship in US during the sample period. Of course, even though a state's business environment is worsen, it is still able to achieve higher ranking in case the environment of its competitors (i.e. states closely ranked) deteriorates more severely.

Table 3.4 is the elasticities of business environment with respect to each covariate. A 10% increase in the previous ranking (e.g. from the 20th to the 22nd) results in 3.2% decrease in business environment. Also, the percentage increase in energy price (*EnPrice*) induces the highest percent decrease in business environment among covariates, x's. Although it is statistically significant, the influence of unemployment rate (*UnempRate*) is marginal.

Figure 3.1 plots the previous ranking elasticities of business environment, evaluated at the mean value of predicted business environment for different values of the previous ranking. As shown in Figure 3.1 (a), states ranked higher in the previous year show higher values of elasticity, meaning less elastic to the percent change in the previous ranking. To understand the influence of a unit change in the previous ranking (i.e. from the first to the second, from the second to the third, and so on,) Figure 3.1 (b) draws the semi-elasticities. From this figure, it is clear that the higher a state was ranked in the previous year, the more the percent decrease occurs by a unit-drop in the previous ranking. For example, a state that ranked first in the previous year would have 3.4% decrease in business environment on average if it was ranked second. However, a state ranked 25th and one ranked 49th would obtain 0.68% and 0.48% lower business environment on average if it was ranked 26th and 50th, respectively. Notice that it is convex according to how the g function is defined.

Finally, Tables 3.5 and 3.6 are the percent changes in business environment in each state and each year from the previous year induced by changes in ranking.

3.7 Concluding Remarks

In this chapter, we have set up a dynamic panel model with latent output variable and its observable ranking to analyze the ranking for best states for business

Table 3.5: Percent Changes in Business Environment: Part I

	2006	2007		2008		2009		2010	
	Ranking	Ranking	%ΔEnv	Ranking	%ΔEnv	Ranking	%ΔEnv	Ranking	%ΔEnv
Alabama	40	35	5.58%	28	7.49%	28	-	25	2.61%
Alaska	47	47	-	48	-1.10%	42	5.56%	42	-
Arizona	15	18	-1.94%	18	-	36	-9.30%	23	3.99%
Arkansas	24	21	2.71%	32	-9.88%	26	3.79%	32	-3.62%
California	36	34	0.51%	40	-1.51%	38	0.45%	39	-0.22%
Colorado	5	8	-3.41%	6	1.62%	4	1.90%	4	-
Connecticut	28	31	-1.74%	33	-0.92%	35	-0.81%	36	-0.34%
Delaware	8	11	-3.08%	12	-0.74%	21	-6.01%	20	0.43%
Florida	9	7	1.42%	8	-0.74%	18	-6.71%	26	-3.07%
Georgia	10	15	-4.37%	5	6.53%	6	-1.18%	8	-1.84%
Hawaii	42	37	3.10%	27	5.64%	39	-8.14%	46	-2.29%
Idaho	6	6	-	7	-1.22%	11	-4.30%	12	-0.75%
Illinois	44	40	3.85%	35	4.63%	24	9.99%	37	-12.57%
Indiana	32	27	4.76%	25	1.93%	30	-4.88%	29	0.70%
Iowa	25	24	0.88%	22	1.65%	14	6.78%	13	1.01%
Kansas	21	20	0.78%	21	-0.75%	15	4.12%	10	3.66%
Kentucky	33	41	-8.55%	44	-2.33%	43	0.68%	31	6.68%
Louisiana	50	49	2.38%	49	-	44	8.97%	44	-
Maine	46	48	-3.12%	46	2.10%	41	4.54%	50	-7.71%
Maryland	11	12	-0.86%	14	-1.42%	12	1.22%	14	-1.17%
Massachusetts	37	36	0.63%	36	-	34	0.99%	16	7.83%
Michigan	45	46	-1.00%	47	-0.78%	49	-1.41%	47	1.17%
Minnesota	14	10	3.63%	11	-1.02%	17	-5.48%	15	1.25%
Mississippi	48	43	7.09%	42	1.26%	40	2.45%	48	-6.71%
Missouri	22	16	4.42%	30	-11.52%	29	0.52%	18	5.19%

Note: Each column denoted as %ΔEnv represents the percent change in business environment of each state from the previous year induced by change in ranking.

Table 3.6: Percent Changes in Business Environment: Part II

	2006	2007		2008		2009		2010	
	Ranking	Ranking	%ΔEnv	Ranking	%ΔEnv	Ranking	%ΔEnv	Ranking	%ΔEnv
Montana	38	42	-4.31%	24	15.33%	13	13.71%	24	-16.32%
Nebraska	7	17	-11.46%	10	4.46%	9	0.84%	9	-
Nevada	26	22	2.38%	19	1.79%	31	-7.45%	28	1.24%
New Hampshire	18	14	2.87%	20	-4.31%	19	0.54%	19	-
New Jersey	16	19	-2.35%	34	-9.26%	45	-4.57%	40	1.48%
New Mexico	29	26	2.36%	15	8.26%	27	-12.46%	35	-4.59%
New York	35	33	1.27%	38	-2.92%	32	3.04%	21	5.29%
North Carolina	3	3	-	4	-1.32%	5	-1.08%	3	1.73%
North Dakota	13	9	4.84%	13	-5.74%	7	6.25%	11	-5.25%
Ohio	34	38	-3.09%	39	-0.63%	37	1.18%	38	-0.53%
Oklahoma	19	30	-9.24%	26	2.26%	20	3.63%	33	-7.77%
Oregon	31	28	2.48%	16	9.46%	10	6.34%	6	4.83%
Pennsylvania	41	39	1.90%	41	-1.66%	33	5.98%	30	2.10%
Rhode Island	43	45	-1.59%	45	-	50	-2.89%	49	0.43%
South Carolina	27	23	3.37%	29	-5.18%	25	2.73%	34	-5.59%
South Dakota	17	25	-6.26%	23	1.13%	16	4.00%	17	-0.65%
Tennessee	20	13	5.41%	17	-3.68%	23	-4.44%	22	0.52%
Texas	2	4	-2.70%	9	-4.43%	8	0.56%	7	0.53%
Utah	4	2	2.87%	2	-	3	-1.85%	1	2.71%
Vermont	30	32	-1.35%	36	-2.26%	47	-5.23%	45	0.66%
Virginia	1	1	-	1	-	1	-	2	-1.94%
Washington	12	5	6.07%	3	2.69%	2	1.69%	5	-5.49%
West Virginia	49	50	-3.05%	50	-	46	6.87%	41	5.81%
Wisconsin	39	44	-4.94%	43	0.76%	48	-3.52%	43	2.75%
Wyoming	23	29	-5.38%	31	-1.37%	22	5.73%	27	-3.37%

Note: *Each column denoted as %ΔEnv represents the percent change in business environment of each state from the previous year induced by change in ranking.

announced annually by Forbes magazine. We have seen that why a typical panel model fails the estimation with precision when n is large and T is small. This study overcomes this challenge by utilizing the mixed exploded logit which uses the information from all pairwise ranking orders. The estimation results for five time periods show that the revealed output ranking explains its latent variable well with selected observable covariates.

APPENDICES

APPENDIX A

Head Size as a Quality Measure in 2005-2009

In this appendix, we discuss why almost all observable product characteristics are taste characteristics: consumers perceive many product characteristics in a Hotelling sense, as modeled in Section 1.3. Also we see why the size of a driver's head can serve as a quality measure in the time frame of dataset analyzed in this study: 2005-2009.

As introduced in Section 1.1 a specific feature of the golf driver is almost all observable product characteristics depend on consumers' taste while having a quality difference over products at least over time. This feature poses both a challenge and an opportunity to the researcher. A challenge to an econometrician is requesting to find an appropriate observable and discernable quality measure in a vertical sense. On the other hand, an opportunity of examining consumers' behavioral characteristics opens up once the quality is controlled for. A carefully chosen quality measure successfully explains true quality of products and lets consumers' behavioral characteristics identified if there is any.

The characteristics of a driver that depend on consumers' taste are decomposed into two parts: (1) bona fide taste characteristics and (2) characteristics having trade-offs between them. This decomposition is worth to compare to the distinction between search and experience characteristics as described in Stigler (1961). Before they purchase consumers can observe and verify search characteristics, e.g. the shape and loft of head. Contrastingly, experience characteristics are not typically known to consumers before testing the product, e.g. feel of grip or ball flight. Consumers however may have information on experience characteristics via magazine reviews, announced test results, or golf buddies' opinion. Characteristics in our first category is a comparable mix of search and experience characteristics whereas the second category mainly consists of experience characteristics.

First, the bona fide taste characteristics include all characteristics subject to each consumer's own taste purely. Hence no consensus can be obtained due to the nature of them. Some consumers like the hitting sound of a driver, which is an important factor that golfers care when they test a new driver, but some think its sound is detestable. Letting alone the unobservable taste characteristics, many of observable characteristics of a driver fall into this category including the shape and loft of head, length and stiffness of shaft, and feel of grip among others. The bona fide taste characteristics are completely free from consideration in a vertical sense.

The second category of characteristics needs more careful attention. Many characteristics of a driver, observable at least indirectly through a good many available test results, are conceptualized in a vertical sense individually, e.g. hitting distance, forgiveness, accuracy, and controllability (or playability[1]) In other words, consumers unanimously agree that a certain behavior of a driver is superior to the opposite behavior on a specific characteristic *ceteris paribus*, e.g. consumers prefers longer distance and higher accuracy holding others equal. Due to the technological trade-off among the characteristics in this category, however, a combination of this type of features is not well-ordered. A driver loses forgiveness in exchange of improved distance,

[1]The term "playability" is defined as the degree of how easy or difficult the clubs are to play for golfers of different skill levels. For example, a cavity back iron with significant perimeter weighting is clearly easier for most golfers to handle than a muscleback blade. We say the former shows higher playability. To golfers with advanced skills who care accuracy or distance a lot, however, higher playability of a golf club does not necessarily mean better quality since there is a trade-off between accuracy or distance and in ease of control.

one of the well-known trade-offs. Consumers' choices regarding these two characteristics are made in a horizontal sense depending on their taste or more precisely own order of importance between the characteristics. Moreover, even on a specific characteristic each consumer assesses differently: one feels driver A hits longer than B but another may think in the opposite direction.

Both Table A.1 and Table A.2 show the professional test results of selected golf drivers: notice both categories of taste characteristics are tested. To evaluate the characteristics of drivers, 60 golfers of varying skill levels hit a group of 5 to 7 driver models each year. They are told not to compare one driver to another but only to rate how they performed with each driver. The ratings are highly concentrated within a range far less than 1 in all criteria on a scale of 1 to 10 in each year as a result of averaging over testers with varying skill levels therefore wide range of taste.

A question worth to investigate is then whether the quality is enhanced over time or not in a vertical sense. As the ratings in Tables A.1 and A.2 are normalized over time, the test results suggest overall quality has been improved from 2006 to 2009: average of overall rating has risen from 7.9 to 8.43. Even though the difference is not huge the newer model of each brand receives generally higher ratings throughout all test categories (see Cleveland Launcher Ti460 (2006) versus Cleveland Launcher (2009) and Mizuno MX500 460 (2006) versus Mizuno MX 700 (2009) for instance.) It arouses the need of quality measure that captures evolution of products, as firms advertise more often than not that their new model is longer, faster, and stronger than its predecessors. Enhanced quality is achieved in two ways: (1) a Pareto improvement, e.g. showing longer hitting distance with better forgiveness or at least without weakening forgiveness, and (2) not a Pareto improvement but the gain in a characteristic dominates the loss in another, e.g. achieving much longer distance with a slight loss in forgiveness.[2]

[2]Think of putting two balls in different size, red and white, into a box. You want the total volume of balls as big as possible in your own preference. If you prefer red balls much to white balls, you

Table A.1: Driver Test Results on a Scale of 1 to 10 (2006)

Driver	Distance	Control	Accuracy	Forgiveness	Ball Flight	Sound	Feel	Overall
Cleveland Launcher Ti460	8.3	7.9	7.8	7.8	8.3	8.1	8.5	8.2
Mizuno MX500 460	8.4	8.2	8.2	7.9	8	8.1	8.4	8.2
Yonex Cyberstar Nanospeed 460	8.4	8	8.1	8	8.2	8	8.4	8.2
Nickent Genex 3DX T-spec 460	8.2	8.1	8	8	8.3	8.1	8.4	8.1
Nike SasQuatch 460	8.4	8.1	8.2	7.9	8.3	7.9	7.9	8.1
Callaway Fusion FT 3 460	8.4	8.1	8	7.6	8.2	7.7	8.1	8
MacGregor MACTEC NVG 445	8	8	8.1	7.8	8	7.5	8.1	8
Ping G5 460	7.9	7.8	7.9	7.7	7.9	7.9	8.3	7.9
Feel TI-Carbon 460	7.9	8	7.8	7.8	8	8	8.1	7.9
TechPower Speed Offset 460	7.8	8	8	7.8	7.7	7.8	7.9	7.9
Adams Redline RPM 460D	8.1	8.1	8	7.7	8.2	7.7	7.9	7.8
KZG Gemini 460	8	7.8	7.8	7.7	7.9	7.7	8	7.8
Bridgestone J33R 460	8.1	7.7	7.7	7.6	8	7.5	7.9	7.8
Infiniti ProPlusion Ti Comp 460	8	7.7	7.8	7.7	7.8	7.6	7.9	7.8
Natural Golf ST Hammer 420	7.9	7.8	7.7	7.9	7.8	8	7.8	7.8
Black Shark S.E.T. 455	7.8	7.7	7.8	7.9	7.7	7.8	7.6	7.8

Note: Source: http://www.golftestusa.com/drivers.html. Refer to the webpage for complete listing of test results. The test scores are normalized across years.

Table A.2: Driver Test Results on a Scale of 1 to 10 (2009)

Driver	Distance	Control	Accuracy	Forgiveness	Ball Flight	Sound	Feel	Overall
Titleist 909 D2	8.7	8.6	8.5	8.5	8.6	8.5	8.7	8.7
Cleveland Launcher	8.6	8.5	8.5	8.6	8.5	8.5	8.5	8.6
Titleist 909 D3	8.5	8.5	8.5	8.4	8.5	8.4	8.6	8.6
Ping G10	8.6	8.6	8.5	8.6	8.6	8.5	8.5	8.6
Mizuno MX 700	8.5	8.4	8.5	8.6	8.5	8.4	8.5	8.5
Pat Simmons Liberator 460	8.6	8.4	8.5	8.4	8.5	8.3	8.4	8.5
Cleveland HiBore Monster XLS	8.4	8.4	8.5	8.5	8.4	8.2	8.5	8.5
Bobby Jones Workshop Edition	8.4	8.3	8.5	8.7	8.4	8.5	8.5	8.4
Nike SQ DYMO	8.5	8.4	8.5	8.5	8.5	8.5	8.5	8.4
Infiniti xMOI Propulsion	8.3	8.3	8.4	8.3	8.4	8.2	8.2	8.3
Natural Golf Hammer	8.3	8.3	8.4	8.4	8.2	8.2	8.2	8.3
Nickent 4DX EVOLVER	8.3	8.4	8.3	8.3	8.2	8.3	8.3	8.3
Pat Simmons Liberator 420	8.3	8.2	8.3	8.2	8.2	8.3	8.3	8.3
Pinemeadow Doublewall	8.3	8.3	8.2	8.4	8.3	8.2	8.3	8.3
Srixon Z-RW	8.2	8.2	8.3	8.2	8.3	8.3	8.3	8.3
Wilson Staff Spine	8.3	8.2	8.4	8.3	8.3	8.3	8.2	8.3

Note: Source: http://www.golftestusa.com/drivers.html. Refer to the webpage for complete listing of test results. The test scores are normalized across years.

It is not appropriate to take the ratings in Tables A.1 and A.2 as a measure of quality since they are on the taste characteristics as described above and rating values are tightly close to each other subsequently.[3] In short of remained observable characteristics, a candidate for a quality indicator must explain the technological difference within the time frame of the study by representing the main stream innovation issue successfully in a specific period with sufficient variability.

The time period of 2005-2009 was the era of maturing the head size. Figure A.1 shows the time paths of average head size in the market: simple and weighted (by sales) averages from December 2004 to December 2009. Both simple and weighted averages show a gradual increase in average head size. They hit the maximum volume level of 460cc at the start of 2008 then stay close to it.

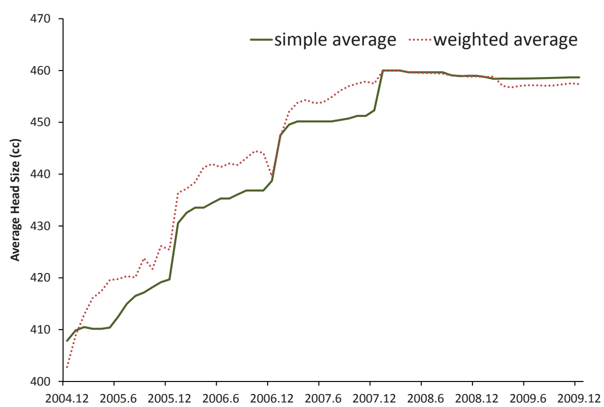


Figure A.1: Monthly Average Head Size of Drivers

Note: Head size is in cubic centimeters. The US Golf Association institutes a size rule which states that no clubhead can measure greater than 460cc. The weighted average is weighted by the volume of sales.

put a big red ball into the box by sacrificing the size of white ball, and vice versa. What really matters is then the size of the box, not each ball's relative size. Each person has different preference on relative size of balls but can always obtain bigger sum of ball sizes as she would like when the box got larger. Quality of a driver corresponds to the box size in this example.

[3]Another challenge in practice is not all models in the data are tested and rated due to various reasons. Some major brands decline to participate in the test and many minor models are not included lacking in mass interest.

The heads of almost all drivers are made of titanium and the Coefficient of Restitution (COR) of all drivers has reached its permitted limit.[4] It is generally accepted that a driver with a bigger head outperforms the significantly smaller-headed ones. First and foremost, it definitely yields better forgiveness with a larger face. Even though a golfer mis-hits a ball more or less, it may carry the ball in the desired direction. As a result of the rapidly-increasing size of driver heads in the late 1990s and the advantage of a bigger head size, the US Golf Association (USGA) curbed the volumetric growth of drivers by instituting a size rule which states that no clubhead can measure greater than 460 cubic centimeters. Manufacturers maintain a light enough weight of a big-headed driver with the use of lighter, stronger, and more expensive material, titanium. Firm's cost of developing/manufacturing a bigger head would be higher than smaller ones. It calls for the use of expensive material and requires more subtle technology to balance the center of gravity and so on. In sum, a bigger head size typically represents a better performance and a higher manufacturing cost.

[4]The material of head has been meliorated from wood to stainless steel in late 1980's and to titanium in mid-1990's. The COR has been grown. The COR is a measure of the energy loss or retention when two objects collide. The higher the value of COR a driver has, the harder a ball is bounced resulting a longer hitting distance. A time of competition to improve the COR has passed since all drivers reached the limit permitted by the US Golf Association (USGA) as of early 2000s. Once a technological improvement has been matured in an aspect then R&D efforts and accordingly advertising is focused on new innovation.

APPENDIX B

Derivation of the Observed Information Matrix

Louis (1982) shows that the observed information matrix $\mathcal{I}$ can be obtained as

$$\mathcal{I}(\Theta) = \mathbb{E}_v \left(\frac{\partial^2}{\partial\Theta\partial\Theta'} \log L(\Theta) \right) - \mathbb{V}_v \left(\frac{\partial}{\partial\Theta} \log L(\Theta) \right) \equiv \mathcal{I}_1(\Theta) - \mathcal{I}_2(\Theta).$$

We investigate these two terms separately. The multi-spell case in Section 5 can be handled similarly using the generalized log-likelihood (2.22).

Complete data information matrix $\mathcal{I}_1(\Theta)$ Under the regularity conditions, it can be obtained as $\mathcal{I}_1(\Theta) = -\partial^2 Q(\Theta)/\partial\Theta\partial\Theta'$ from the E-step. More precisely, from the formula (2.21) and for individual i,

$$Q_i(\Theta) = \sum_{j=1}^{m} \widehat{\pi}_{ij} \log p_j + \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i \log\left(1 - \exp\left(-q_j \psi_{i,T_i}\right)\right) - \sum_{s=1}^{T_i - 1} q_j \psi_{i,s} \right\},$$

where $\psi_{i,t} = \exp(\gamma_t)\phi(z'_{i,t}\theta)$. Since the maximization procedures of p and (γ, θ, q) can be separated, the Hessian matrix of $Q_i(\Theta)$ is block diagonal. In particular, for $t = 1, \cdots, T$ and $k = 1, \cdots, m$,

$$\begin{aligned}
\frac{\partial}{\partial \gamma_t} Q_i(\Theta) &= \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i \frac{q_j \psi_{i,T_i} \exp(-q_j \psi_{i,T_i})}{1 - \exp(-q_j \psi_{i,T_i})} 1\{t = T_i\} - \sum_{s=1}^{T_i - 1} q_j \psi_{i,s} 1\{t = s\} \right\}, \\
\frac{\partial}{\partial \theta} Q_i(\Theta) &= \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i \frac{q_j \psi_{i,T_i} \exp(-q_j \psi_{i,T_i})}{1 - \exp(-q_j \psi_{i,T_i})} \Psi_{i,T_i}^{(1)} - \sum_{s=1}^{T_i - 1} q_j \Psi_{i,s}^{(1)} \right\}, \\
\frac{\partial}{\partial q_k} Q_i(\Theta) &= \widehat{\pi}_{ik} \left\{ \delta_i \frac{\psi_{i,T_i} \exp(-q_k \psi_{i,T_i})}{1 - \exp(-q_k \psi_{i,T_i})} - \sum_{s=1}^{T_i - 1} \psi_{i,s} \right\}, \\
\frac{\partial}{\partial p_k} Q_i(\Theta) &= \frac{\widehat{\pi}_{ik}}{p_k} - \frac{\widehat{\pi}_{im}}{p_m} \quad (\text{for } k = 1, \cdots, m-1)
\end{aligned}$$

with $p_m = 1 - \sum_{j=1}^{m-1} p_j$, where $\Psi_{i,t}^{(1)} = \partial \psi_{i,t} / \partial \theta = \exp(\gamma_t) \phi^{(1)}(z_{i,t}' \theta) z_{i,t}$. We let

$$A_{ij} = \frac{\psi_{i,T_i} \exp(-q_j \psi_{i,T_i})}{1 - \exp(-q_j \psi_{i,T_i})}. \tag{B.1}$$

Then, for $t, r = 1, \cdots, T$ and $k, \ell = 1, \cdots, m$, we have

$$\begin{aligned}
\frac{\partial^2 Q_i(\Theta)}{\partial \gamma_t \partial \gamma_r} &= \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i q_j A_{ij} (1 - q_j \psi_{i,T_i} - q_j A_{ij}) 1\{t = r = T_i\} - \sum_{s=1}^{T_i - 1} q_j \psi_{i,s} 1\{t = r = s\} \right\}, \\
\frac{\partial^2 Q_i(\Theta)}{\partial \theta \partial \gamma_t} &= \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i q_j A_{ij} \left(1 - q_j \psi_{i,T_i} - q_j \frac{A_{ij}}{\psi_{i,T_i}}\right) \Psi_{i,T_i}^{(1)} 1\{t = T_i\} - \sum_{s=1}^{T_i - 1} q_j \Psi_{i,s}^{(1)} 1\{t = s\} \right\}, \\
\frac{\partial^2 Q_i(\Theta)}{\partial \theta \partial \theta'} &= \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i q_j A_{ij} \left[\left(1 - q_j \psi_{i,T_i} - q_j \frac{A_{ij}}{\psi_{i,T_i}}\right) \Psi_{i,T_i}^{(1)} \Psi_{i,T_i}^{(1)\prime} + \Psi_{i,T_i}^{(2)} \right] - \sum_{s=1}^{T_i - 1} q_j \Psi_{i,s}^{(2)} \right\}, \\
\frac{\partial^2 Q_i(\Theta)}{\partial \gamma_t \partial q_k} &= \widehat{\pi}_{ik} \left\{ \delta_i A_{ij} (1 - q_j \psi_{i,T_i} - q_j A_{ij}) 1\{t = T_i\} - \sum_{s=1}^{T_i - 1} \psi_{i,s} 1\{t = s\} \right\}, \\
\frac{\partial^2 Q_i(\Theta)}{\partial \theta \partial q_k} &= \widehat{\pi}_{ik} \left\{ \delta_i A_{ij} \left(1 - q_j \psi_{i,T_i} - q_j \frac{A_{ij}}{\psi_{i,T_i}}\right) \Psi_{i,T_i}^{(1)} - \sum_{s=1}^{T_i - 1} \Psi_{i,s}^{(1)} \right\}, \\
\frac{\partial^2 Q_i(\Theta)}{\partial q_k \partial q_\ell} &= -\widehat{\pi}_{ik} \delta_i \left(\psi_{i,T_i} A_{ik} + A_{ik}^2 \right) 1\{k = \ell\}, \\
\frac{\partial^2 Q_i(\Theta)}{\partial p_k \partial p_\ell} &= -\frac{\widehat{\pi}_{ik}}{p_k^2} 1\{k = \ell\} + \frac{\widehat{\pi}_{im}}{p_m^2} \quad (\text{for } k, \ell = 1, \cdots, m-1),
\end{aligned}$$

where $\Psi_{i,t}^{(2)} = \partial^2 \psi_{i,t} / \partial\theta\partial\theta' = \exp(\gamma_t)\phi^{(2)}(z_{i,t}'\theta) z_{i,t} z_{i,t}'$. The complete data information matrix is then given by $\mathcal{I}_1(\Theta) = -\sum_{i=1}^n \partial^2 Q_i(\Theta) / \partial\Theta\partial\Theta'$, where $\partial^2 Q_i(\Theta) / \partial\Theta\partial\Theta'$ consists of the second derivatives obtained above conformably with the array of parameters

$$\Theta = (\gamma_1, \cdots, \gamma_T, \theta', q_1, \cdots, q_m, p_1, \cdots, p_{m-1})'. \tag{B.2}$$

All other terms are zero since $\partial^2 Q_i(\Theta) / \partial\Theta\partial\Theta'$ is block diagonal.

Missing information matrix $\mathcal{I}_2(\Theta)$ From (2.19), we have

$$\log L_i(\Theta) = \sum_{j=1}^m \eta_{ij} \log p_j + \sum_{j=1}^m \eta_{ij} \left\{ \delta_i \log(1 - \exp(-\psi_{i,T_i} q_j)) - \sum_{s=1}^{T_i - 1} \psi_{i,s} q_j \right\}$$

for individual i, which gives (for $t = 1, \cdots, T$ and $k = 1, \cdots, m$)

$$\begin{aligned}
\frac{\partial}{\partial\gamma_t} \log L_i(\Theta) &= \sum_{j=1}^m \eta_{ij} \left\{ \delta_i q_j A_{ij} 1\{t = T_i\} - \sum_{s=1}^{T_i-1} q_j \psi_{i,s} 1\{t = s\} \right\}, \\
\frac{\partial}{\partial\theta} \log L_i(\Theta) &= \sum_{j=1}^m \eta_{ij} \left\{ \delta_i q_j A_{ij} \Psi_{i,T_i}^{(1)} - \sum_{s=1}^{T_i-1} q_j \Psi_{i,s}^{(1)} \right\}, \\
\frac{\partial}{\partial q_k} \log L_i(\Theta) &= \eta_{ik} \left\{ \delta_i A_{ik} - \sum_{s=1}^{T_i-1} \psi_{i,s} \right\}, \\
\frac{\partial}{\partial p_k} \log L_i(\Theta) &= \frac{\eta_{ik}}{p_k} - \frac{\eta_{im}}{p_m} \quad \text{(for } k = 1, \cdots, m-1\text{)},
\end{aligned}$$

where A_{ij} is defined as (B.1). We let

$$\begin{aligned}
B_{ij}^{\gamma_t} &= \delta_i q_j A_{ij} 1\{t = T_i\} - \sum_{s=1}^{T_i-1} q_j \psi_{i,s} 1\{t = s\}, \\
B_{ij}^{\theta} &= \delta_i q_j A_{ij} \Psi_{i,T_i}^{(1)} - \sum_{s=1}^{T_i-1} q_j \Psi_{i,s}^{(1)}, \\
B_{ij}^{q} &= \delta_i A_{ij} - \sum_{s=1}^{T_i-1} \psi_{i,s}.
\end{aligned}$$

Then, since $Cov_v(\eta_{ik}, \eta_{i\ell}) = \pi_{ik}(1-\pi_{ik}) 1\{k = \ell\}$ for $k, \ell = 1, \cdots, m$ from the definition of η_{ij}, where the covariance $Cov_v(\cdot)$ and the variance $\mathbb{V}_v(\cdot)$ are taken over the conditional distribution of v given the observed data $\{\delta_i, Z_i\}_{i=1}^n$, we derive that

$$
\begin{aligned}
\mathbb{V}_v\left(\frac{\partial \log L_i(\Theta)}{\partial \theta}\right) &= \sum_{j=1}^m \pi_{ij}(1-\pi_{ij}) B_{ij}^{\theta} B_{ij}^{\theta\prime}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \gamma_t}, \frac{\partial \log L_i(\Theta)}{\partial \gamma_s}\right) &= \sum_{j=1}^m \pi_{ij}(1-\pi_{ij}) B_{ij}^{\gamma_t} B_{ij}^{\gamma_s} 1\{t = s\}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial q_k}, \frac{\partial \log L_i(\Theta)}{\partial q_\ell}\right) &= \pi_{ik}(1-\pi_{ik})(B_{ik}^q)^2 1\{k = \ell\}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial p_k}, \frac{\partial \log L_i(\Theta)}{\partial p_\ell}\right) &= \frac{\pi_{ik}(1-\pi_{ik})}{p_k^2} 1\{k = \ell\} + \frac{\pi_{im}(1-\pi_{im})}{p_m^2}
\end{aligned}
$$

for $t, r = 1, \cdots, T$ and $k, \ell = 1, \cdots, m$. Similarly,

$$
\begin{aligned}
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \theta}, \frac{\partial \log L_i(\Theta)}{\partial \gamma_t}\right) &= \sum_{j=1}^m \pi_{ij}(1-\pi_{ij}) B_{ij}^{\theta} B_{ij}^{\gamma_t}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \theta}, \frac{\partial \log L_i(\Theta)}{\partial q_k}\right) &= \pi_{ik}(1-\pi_{ik}) B_{ik}^{\theta} B_{ik}^{q}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \theta}, \frac{\partial \log L_i(\Theta)}{\partial p_\ell}\right) &= \frac{\pi_{i\ell}(1-\pi_{i\ell})}{p_\ell} B_{i\ell}^{\theta}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \gamma_t}, \frac{\partial \log L_i(\Theta)}{\partial q_k}\right) &= \pi_{ik}(1-\pi_{ik}) B_{ik}^{\gamma_t} B_{ik}^{q}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \gamma_t}, \frac{\partial \log L_i(\Theta)}{\partial p_\ell}\right) &= \frac{\pi_{i\ell}(1-\pi_{i\ell})}{p_\ell} B_{i\ell}^{\gamma_t}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial q_k}, \frac{\partial \log L_i(\Theta)}{\partial p_\ell}\right) &= \frac{\pi_{ik}(1-\pi_{ik})}{p_k} B_{ik}^{q} 1\{k = \ell\}
\end{aligned}
$$

for $t = 1, \cdots, T$, $k = 1, \cdots, m$ and $\ell = 1, \cdots, m-1$. The missing information matrix is then given by $\mathcal{I}_2(\Theta) = \sum_{i=1}^n \mathbb{V}_v(\partial \log L_i(\Theta)/\partial \Theta)$, where $\mathbb{V}_v(\partial \log L_i(\Theta)/\partial \Theta)$ consists of covariance matrices obtained above conformably with the array of parameters Θ in (B.2). $\square$

BIBLIOGRAPHY

BIBLIOGRAPHY

Aguirregabiria, V., and P. Mira (2010): "Dynamic Discrete Choice Structural Models: A survey," *Journal of Econometrics*, 156(1), 38–67.

Ali, G. (2008): "Expected Utility in Econometric Random Utility Models," *Sankhyā: The Indian Journal of Statistics*, 70-B(1), 63–68.

Anderson, S. P., A. de Palma, and J.-F. Thisse (1992): *Discrete Choice Theory of Product Differentiation*. The MIT Press, Cambridge, Massachusetts.

Anselin, L., and A. K. Bera (1998): "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," in *Handbook of Applied Economics Statistics*, ed. by A. U. Giles, and D. E. A. Marcel Dekker, New York.

Becker, G. S., and K. M. Murphy (1993): "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics*, 108(4), 941–964.

Beggs, S., S. Cardell, and J. A. Hausman (1981): "Assessing the Potential Demand for Electric Cars," *Journal of Econometrics*, 16, 1–19.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Berry, S., and A. Pakes (2000): "Estimation from the Optimality Conditions for Dynamic Controls," *Manuscript*, Yale University and Harvard University.

Brock, W. A., and S. N. Durlauf (2001a): "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68(2), 235–260.

——— (2001b): "Interactions-Based Models," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5, pp. 3297–3380. North-Holland, Amsterdam.

Carranza, J. E. (2010): "Product Innovation and Adoption in Market Equilibrium: The Case of Digital Cameras," *International Journal of Industrial Organization*, 28(6), 604–618.

Carruthers, B., T. Guinnane, and Y. Lee (2012): "Bringing "Honest Capital" to Poor Borrowers: The Passage of Uniform Small Loan Law, 1907-1930," *Journal of Interdisciplinary History*, 42(3), 393–418.

Casella, G., and R. L. Berger (2001): *Statistical Inference*. Duxbury Press, New York, 2nd edn.

Chapman, R. G., and R. Staelin (1982): "Exploiting Rank Ordered Choice Set Data within the Stochastic Utility Model," *Journal of Marketing Research*, 19, 288–301.

Conlon, C. T. (2010): "A Dynamic Model of Costs and Margins in the LCD TV Industry," *Manuscript*, Columbia University.

Cooper, R., and A. John (1988): "Coordinating Coordination Failures in Keynesian Models," *Quarterly Journal of Economics*, 103(3), 441–463.

Cooter, R. D., A. J. McLean, D. J. David, and D. A. Simpson (1988): "Helmet-Induced Skull Base Fracture in a Motorcyclist," *Lancet*, 1, 84–85.

Cox, D. (1972): "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, B(20), 187–220.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data Via Em Algorithm," *Journal of the Royal Statistical Society Series B-Methodological*, 39(1), 1–38.

Eckstein, Z., and K. I. Wolpin (1989): "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey," *Journal of Human Resources*, 24(4), 562–598.

Elbers, C., and G. Ridder (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 49(3), 403–09.

Everitt, B., and D. J. Hand (1981): *Finite Mixture Distributions*, Monographs on Applied Probability and Statistics. Chapman and Hall, London ; New York.

Farewell, V. T., and R. L. Prentice (1980): "The Approximation of Partial Likelihood with Emphasis on Case-Control Studies," *Biometrics*, 67, 273–278.

Goffe, W. L., G. D. Ferrier, and J. Rogers (1994): "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60(1-2), 65–99.

Goldstein, J. P. (1986): "The Effect of Motorcycle Helmet Use on the Probability of Fatality and the Severity of Head and Neck Injuries," *Evaluation Review*, 10, 355–375.

Gowrisankaran, G., and M. Rysman (2011): "Dynamics of Consumer Demand for New Durable Goods," *Manuscript*, University of Arizona and Boston University.

Guo, G. A., and G. Rodriguez (1992): "Estimating a Multivariate Proportional Hazards Model for Clustered Data Using the Em Algorithm, with an Application to Child Survival in Guatemala," *Journal of the American Statistical Association*, 87(420), 969–976.

Halvorsen, R., and R. Palmquist (1980): "The Interpretation of Dummy Variables in Semilogarithmic Equations," *American Economic Review*, 70(3), 474–475.

Hausman, J. A., and P. A. Ruud (1987): "Specifying and Testing Econometric Models for Rank-Ordered Data," *Journal of Econometrics*, 34, 83–104.

Heckman, J. J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52(2), 271–320.

Hendel, I., and A. Nevo (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74(6), 1637–1673.

Hitsch, G. J. (2006): "An Empirical Model of Optimal Dynamic Product Launch and Exit Under Demand Uncertainty," *Marketing Science*, 25(1), 25–50.

Houston, D. J., and L. E. Richardson (2007): "Motorcycle Safety and the Repeal of Universal Helmet Laws," *American Journal of Public Health*, 97(11), 2063–2069.

Hui, K.-L. (2004): "Product Variety under Brand Influence: An Empirical Investigation of Personal Computer Demand," *Management Science*, 50(5), 686–700.

Huston, R., and J. Sears (1981): "Effect of Protective Helmet Mass on Head/Neck Dynamics," *Journal of Biomedical Engineering*, 103, 18–23.

Jiang, R., P. Manchandab, and P. E. Rossi (2009): "Bayesian Analysis of Random Coefficient Logit Models Using Aggregate Data," *Journal of Econometrics*, 149(1), 136–148.

Kalbfleisch, J., and R. Prentice (1973): "Marginal Likelihoods Based on Cox's Regression and Life Model," *Biometrika*, 60(2), 267–278.

Kalbfleisch, J. D., and R. L. Prentice (2002): *The Statistical Analysis of Failure Time Data, 2nd ed.* John Wiley & Sons, Inc., Hoboken, New Jersey.

Kelejian, H. H., and I. R. Prucha (1999): "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40(2), 509–533.

Kiefer, J., and J. Wolfowitz (1956): "Consistency of the Maximum Likelihood Estimator in the Presencse of Infinitely Many Incidental Parameters," *Annuls of Mathematical Statistics*, 27(4), 887–906.

Krantz, K. P. G. (1985): "Head and Neck Injuries to Motorcycle and Moped Riders with Special Regard to the Effect of Protective Helmets," *Injury*, 16, 253–258.

Lancaster, T. (1979): "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47(4), 939–956.

Lancaster, T. (1990): *The Econometric Analysis of Transition Data*, Econometric Society Monographs No. 17. Cambridge University Press, Cambridge ; New York.

Lee, L.-F. (2000): "A Numerically Stable Quadrature Procedure for the One-Factor Random-Component Discrete Choice Model," *Journal of Econometrics*, 95(1), 117–129.

——— (2004): "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72(6), 1899–1925.

Lee, Y. (2007): "Grouped Mixed Proportional Hazard Models with Social Interactions," Discussion paper, University of Michigan.

Louis, T. A. (1982): "Finding the Observed Information Matrix When Using the Em Algorithm," *Journal of the Royal Statistical Society Series B-Methodological*, 44(2), 226–233.

Luce, R. D., and P. Suppes (1965): "Preference, Utility, and Subjective Probability," in *Handbook of Mathematical Psychology–Volume 3,* R. Duncan Luce, Robert R. Bush, and Eugene Galanter, eds. New York: John Wiley & Sons, Inc., 249-410.

Manski, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60(3), 531–42.

Melnikov, O. (2001): "Demand for Differentiated Durable Products: The Case of the US Printer Market," *Manuscript*, Cornell University.

Meng, X. L., and D. B. Rubin (1991): "Using Em to Obtain Asymptotic Variance - Covariance Matrices - the Sem Algorithm," *Journal of the American Statistical Association*, 86(416), 899–909.

Meyer, B. D. (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58(4), 757–82.

——— (1995): "Semiparametric Estimation of Hazard Models," Discussion paper, Northwestern University.

Muller, A. (2004): "Florida's Motorcycle Helmet Law Repeal and Fatality Rates," *American Journal of Public Health*, 94(4), 556–558.

Nair, H. (2007): "Intertemporal Price Discrimination with Forward-Looking Consumers: Application to the US Market for Console Video-Games," *Quantitative Marketing and Economics*, 5(3), 239–292.

Oakes, D. (1999): "Direct Calculation of the Information Matrix Via the Em Algorithm," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 61, 479–482.

Prentice, R., and L. Gloeckler (1978): "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data," *Biometrics*, 34, 57–67.

Rust, J. (1994): "Structural Estimation of Markov Decision Process," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden. Elsevier Science, B. V.

Sirakaya, S. (2006): "Recidivism and Social Interactions," *Journal of the American Statistical Association*, 101(475), 863–877.

Song, I., and P. K. Chintagunta (2003): "A Micromodel of New Product Adoption with Heterogeneous and Forward-Looking Consumers: Application to the Digital Camera Category," *Quantitative Marketing and Economics*, 1(4), 371–407.

Stigler, G. J. (1961): "The Economics of Information," *Journal of Political Economy*, 71(3), 213–225.

Stigler, G. L., and G. S. Becker (1977): "De Gustibus Non Est Disputandum," *American Economic Review*, 67(2), 76–90.

Stolzenberg, L., and S. J. D'Alessio (2003): ""Born to Be Wild" the Effect of the Repeal of Florida's Mandatory Motorcycle Helmet-Use Law on Serious Injury and Fatality Rates," *Evaluation Review*, 27(2), 131–150.

Sueyoshi, G. T. (1995): "A Class of Binary Response Models for Grouped Duration Data," *Journal of Applied Econometrics*, 10(4), 411–431.

Trussell, J., and T. Richards (1985): "Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure," *Sociological Methodology*, pp. 242–276.

Van den Berg, G. J. (2001): "Duration Models: Specification, Identification, and Multiple Durations," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5. North-Holland, Amsterdam.

Wallis, K. (1980): "Econometric Implications of the Rational Expectations Hypothesis," *Econometrica*, 48(1), 49–73.

Weiss, A. A. (1992): "The Effect of Helmet Use on the Severity of Head Injuries in Motorcycle Accidents," *Journal of the American Statistical Association*, 87, 48–56.

Zhao, Y. (2008): "Why are Prices Falling Fast? An Empirical Study of the US Digital Camera Market," *Manuscript*, Queens College, City University of New York.