Discrete Choice Demand Estimation Leveraging Overlapping
Groups of Consumers: Estimating the Informative and Prestige
Effects of Celebrity Endorsements

Cheng Chou   Tim Derdenger*

May 2021

Abstract

Multiple groups of consumers usually reside in the same market. These groups often “overlap”—for a consumer from a high-income group, there is a consumer from a low-income group that has the same unobserved heterogeneity, e.g. price sensitivity, due to within group variation. Leveraging such cross-group matching of consumers, we combine the information from each groups’ product market shares to directly estimate the conditional choice probabilities (CCP) as a function of unobserved consumer heterogeneity. Armed with our novel CCP estimator, we develop a new approach using group level market share data to model, identify and estimate a dynamic discrete demand model for durable goods with continuous unobserved consumer heterogeneity and unobserved product characteristics. Applying our new method and using a unique source of identification of Tiger Woods’ infidelity scandal, we separately identify the prestige and informational effects associated with a celebrity endorsement.

Keywords: Dynamic discrete choice, dynamic selection, celebrity endorsement, market shares, privacy

1 Introduction

Dynamic discrete choice models play a key role in modeling consumer demand due to their ability to incorporate the dynamics of the state of the market and the intertemporal preferences of consumers. The incorporation of these dynamic aspects comes at the cost of complexity of estimation

*Cheng Chou is an Assistant Professor of Economics at School of Business, University of Leicester. Email: cchou@le.ac.uk. Tim Derdenger is an Associate Professor of Marketing and Strategy at the Tepper School of Business, Carnegie Mellon University. Email: derdenge@andrew.cmu.edu
and obscurity of identification. Specifically, defining a tractable state space while accounting for all the products in the market is often a difficult task, leading some to adopt ad hoc approximation methods. The task becomes even more challenging when the researcher wants to include multidimensional unobserved state variables, consumer and product specific, while having access to only aggregate sales data. Besides the estimation difficulties, it is also uncertain whether (or which of) the structural parameters are identified when there are both continuous unobserved consumer heterogeneity and product characteristics. In the market of durable products where consumers typically leave after purchasing, we have the additional problem that the distribution of unobserved consumer heterogeneity (e.g. random price coefficients), for those consumers who remain in the market, is likely to change over time. It is necessary to understand the consequences of such non-random attrition of consumers (also known as dynamic selection), which usually causes estimation bias in panel data analysis if ignored.

Our main methodological contribution is to develop a novel approach using market level data to model, identify, and estimate a dynamic discrete choice demand model for durable goods with dynamic selection, continuous unobserved consumers heterogeneity and continuous unobserved product characteristics, in addition to the commonly included individual-product idiosyncratic errors. The unobserved product characteristics are specified as serially correlated and correlated with the observed product characteristics, particularly price. The continuous unobserved consumer heterogeneity (e.g. random price coefficient) can be multidimensional, and its distribution varies over time due the non-random attrition of consumers. We provide a new method to estimate all model primitives, including the consumer's discount factor, without the need to reduce the dimension of the state space or by other approximation techniques such as discretizing state variables. We also provide new identification results that show the model is identified while being agnostic about how consumers form their beliefs regarding the state transition distribution.\footnote{Recently, \cite{An, Hu and Xiao 2020} use individual level panel data to identify agents preference and their subjective beliefs, which do not need to be rational expectation or myopic. Our results are based on market level data, the discount factor in this paper will be identified without belief restriction (the discount factor is assumed to be known in their paper), and our state variables, observed and unobserved, are all continuous (the state variables, excepting for the conventional utility shocks, are discrete in their paper). By no means, we are claiming that our results are more general. We limit our research scope to the market of durable goods, where purchasing can be viewed a terminal action hence simplifying the task, but their paper focuses on general dynamic discrete choice models.} The implementation of our new estimator only involves nonlinear least squares (NLS) and 2 stage least squares (2SLS). Particularly, one does not need to solve or simulate the dynamic programming discrete
choice model. The estimation simplicity allows researchers to estimate multiple model specifications at little computational cost. With the absence of the curse of dimensionality, it also makes the dynamic demand model more applicable to markets with many differentiated products. Indeed, including more products improves the efficiency of our estimation rather than causing the curse of dimensionality. We apply our new estimator to study the celebrity endorsements effect in the golf equipment market, in which there are 118 products, and each has 17 continuous observed characteristics in our sample. We obtain plausible price elasticities and discount factor from our estimated model.

The proposed new approach relies on our new idea of estimating the conditional choice probabilities (CCP) functions, which is our second contribution. In its original form (Hotz and Miller, 1993), the CCP function is a function of observed state variables. Applying the original CCP estimator to the market of durable goods has two major difficulties. The first is the large dimension of product space and/or product characteristics space. The second is the continuous multidimensional unobserved state variables (unobserved consumer preference heterogeneity) whose unknown distribution could also vary over the course of time due to non-random attrition of consumers. We provide a new perspective by exploring market level data about multiple demographic groups of consumers in the same market. Instead of viewing the CCP as a function of all observed state variables as in individual level panel data, our objective is to estimate the CCP as a function of unobserved consumer heterogeneity for each group and market. Recovering the CCP for each group and market directly along with the value of unobserved consumer heterogeneity is the central pillar of our estimator and essential for addressing the dynamic selection problem due to non-random attrition of consumers after purchasing. We discover that when we observe the market shares for a product in multiple groups of consumers in the same market, we can easily estimate the CCP function that includes unobserved consumer heterogeneity by NLS. We show that in practice market share data from only two groups will suffice. It is worth noting that this new CCP estimator can be applied to other demand models. Particularly, our new CCP estimator can simplify the estimation of the popular BLP model (Berry, Levinsohn and Pakes, 1995), which can be viewed as a myopic/static version of our dynamic demand model.

To see the intuition, note that for a demographic group \( g \), the known market share in this group is the integrated unknown CCP of group \( g \) with respect to the unknown distribution of unobserved

\footnote{Arcidiacono and Miller (2011) made important progress so that the CCP function can depend on an agent's unobserved discrete type.}
consumer heterogeneity in group $g$. This can be viewed as one moment condition. If the number of unknown CCP functions grows with the number of groups, we can never recover these unknown CCPs. The key insight is that when there are multiple groups of consumers residing in the same market, these groups usually “overlap” statistically—for a consumer from one group there could be a consumer from another group where they both have the same unobserved heterogeneity, e.g. price sensitivity, due to within group variation. Because these two consumers also face the same state of the market, their CCPs are the same. By leveraging such cross-group matching of consumers, we can combine the information in the market shares for a product from the overlapping groups to directly estimate one single CCP as a function of unobserved consumer heterogeneity.

The presence of multiple groups of consumers in the same market creates within-market variation, which also plays a key role in simplifying our CCP estimation. Exploiting the variation of group market shares within the same market, we can avoid the estimation issues due to the unobserved product characteristics and the possibly high dimension of product characteristics (since they are fixed given one particular market). We explicitly show how the effect of the state of the market on demand is aggregated into the parameters of our CCP as a function of unobserved consumer heterogeneity, which are then further mapped to the parameters of consumer flow utility functions.

As our last but not least contribution, the empirical part of this paper separately identifies the prestige and informational effects of a celebrity endorsement using data from the golf equipment market. Recent work shows that celebrity endorsements can and do increase firm value via sales [Chung, Derdenger and Srinivasan, 2013] or stock prices [Agrawal and Kamakura, 1995; Elberse and Verleug, 2012]. However, what still remains unanswered empirically is the question of how an endorsement affects a consumer’s decision process. Does it provide information to consumers (e.g. about quality), generate prestige for consumers to associate themselves with the celebrity, or some combination? For instance, at the beginning of 2000, Tiger Woods was the number one player in the world and was playing Titleist equipment to win many golf tournaments. Does the fact he was the number one player and playing Titleist equipment provide information to consumers about

---

3The use of within-market variation is not new in the literature of demand identification and estimation. Recently, Berry and Haile [2020] develop new results of nonparametric identification of demand function using the variation in the choice probabilities of different individual consumers within the same market. Within a market, the state of the market (including demand shocks within the market) does not change, but the observed consumer heterogeneity can still shift demand quantity—observably different consumers have different choice probabilities. Hence, the within-market variation of observed consumer heterogeneity is natural instruments for demand quantities.
Titleist, perhaps that it was high quality. Or, does his endorsement lead consumers to purchase Titleist products because they want to associate themselves with the number one player in the world (social prestige)? In our data, we observe the market shares of golf clubs in two consumer groups: those who belong to a private golf club and those who do not. Our key insight and source of identification comes from Tiger Woods’ infidelity scandal in November 2009, which presumably affects the prestige effect of his endorsement and not the informational effect due to his remarkable streak of being ranked number one in the sport of professional golf for 281 consecutive weeks.

Our model determines the benefit from Tiger Woods’ endorsement largely consists of a transfer of information to consumers about the quality of the product and less so by a consumer’s drive to associate him/herself with Tiger Woods by playing the same brand of equipment. We find the social prestige effect consists of roughly 9 percent of Tiger Woods overall endorsement effect. That said, this result is not inconsequential as it is the first paper (to our knowledge) to empirically illustrate that consumers do adopt products based upon (social) prestige. In a related paper regarding the informative effects of advertising, Ackerberg (2003), using data from the yogurt market, determines that consumer behavior is driven by a large informative effect for advertising and a statistically insignificant prestige effect.

With having discussed the paper’s innovations, we believe it is important to also discuss the data requirements for implementing our new methodology and its relevance for the future. Up to now, researchers who employ market level sales data have been in search of a methodology that is able to accommodate unobserved state variables as well as continuous forms of unobserved consumer heterogeneity in preference parameters, but without the cost of reducing the state space via approximation. With our methodology along with a simple data request for a panel of product sales or market share data for two or more consumer group, researchers can now account for both needs at no cost. We believe in a day and age where consumers are more aware of privacy and companies are unwilling to share their most prized digital asset (individual data), the need to find such a method becomes even more important.

In the rest of the introduction, we discuss the literature. Our identification results are novel relative to the literature on identifying dynamic discrete choice (DDC) models. Our model for durable goods can be understood as a general DDC model in which a subset of unobserved state variables (unobserved product characteristics herein) are continuous, serially correlated and correlated with other observed state variables. The existing identification results (Magnac and Thesmar, 2002; Norets, 2009; Kasahara and Shimotsu, 2009; Arcidiacono and Miller, 2011, 2018; Hu and Shum, 2022)
in the literature of DDC models cannot be applied here.

Most of the research focusing on individual-level data do not include persistent unobservable state variables (e.g. Bajari et al., 2016; Daljord, Nekipelov and Park, 2018). The following exceptions involving persistent unobservables are worth noting. Hu and Shum (2012) study dynamic binary choice models with continuous unobserved state variables, but their identification result is limited to the conditional choice probabilities and state transition distribution functions, not to model primitives like flow utility functions and the discount factor. Norets (2009) does include a serially correlated unobservable idiosyncratic error, which is individual-specific rather than an aggregate product shock like in our case. Arcidiacono and Miller (2011) model persistent unobservables, but limit them to a discrete set of values.

Our estimation approach is also new relative to the literature on estimating DDC models. First, our estimation approach is not an approximation method, and thus does not rely on the validity of specific approximations like interpolation or other value function approximations, or behavioral assumptions that consumers only consider some function of the state space and not the entire state (Melnikov, 2013; Gowrisankaran and Rysman, 2012). Second, our estimator does not exhibit a curse of dimensionality, because it does not require the estimation or approximation of the ex-ante expected value function, as is almost always the case with prior papers (e.g. Rust, 1994; Bajari et al., 2016). Third, we estimate more model primitives than the current literature since our method recovers not just the preference parameters but also the discount factor.

Our work builds on several foundational papers in the demand estimation literature. First is the result that the difference between choice-specific payoff is a function of individual choice probabilities (Hotz and Miller, 1993) in static and dynamic settings. The work of Berry (1994) and the BLP model (Berry, 1994; Berry, Levinsohn and Pakes, 1995; Berry and Haile, 2014) on demand estimation with market level data including unobservable product characteristics have been extensively used. This is similar to our setting, but focused on a static environment.

Extending the BLP models to a dynamic setting with forward-looking agents is challenging. Some researchers either do not model persistent unobserved shocks (Song and Chintagunta, 2003), or make them time-invariant (Goettler and Gordon, 2011). Others have focused on improving the computational speed of fixed point estimators with a variety of approaches. Melnikov (2013) and

---

4 We note that Daljord, Nekipelov and Park (2018) presents an innovative way to identify the discount factor in DDC models with individual data. The primary differences is that our setting involves persistent unobservable state variables, whereas those are not present in the aforementioned paper.
Gowrisankaran and Rysman (2012) develop an approximation based on inclusive value sufficiency that allows the researcher to collapse the multi-dimensional state into one dimension, making the problem much more computationally tractable. Moreover, the formal identification in the paper is not specified. Derdenger and Kumar (2019) have studied the approximation properties of this approach, and have shown that in general it is a biased and an inconsistent estimator. Dubé, Fox and Su (2012) propose a constrained optimization approach (Su and Judd 2012) to estimate static and dynamic structural models base on aggregate data.

The rest of the paper is structured as follows. In section 2, we describe our data for studying the celebrity endorsement in the golf equipment market. In section 3, we present the basic modeling approach. In section 4, we show the identification and estimation assuming that the CCP functions have been estimated. The CCP estimation is discussed in section 5. We report and discuss the estimation results in section 6. Section 7 concludes this paper. The appendix contains technical proofs/details, simulation studies and some extensions of the main theory.

2 Celebrity Endorsement Effect: Prestige vs. Information

Given the widespread use of celebrity endorsements, industry clearly views them as a method to increase firm value. But how prevalent are the above views with regard to the underling mechanism within corporations? Do CEOs or CMOs believe consumers purchase products because an endorsement sends information to consumers about a products quality? A quote from the Titleist President in 2000, Ed Abrain, in response to whether he saw an increase in sales when Tiger Woods endorsed his Titleist brand, highlights how industry executives and marketing professionals view celebrity endorsements and the role they play in attracting consumers.

How much of that you can attribute to Tiger [Woods], I don’t know. We didn’t sign him to specifically put a number on how much he could increase our sales. We signed him because he represented an affirmation that we were a credible company with outstanding products.^[5]

The role celebrity endorsements play in revealing information to consumers is an understudied area. Our research looks to understand the underlying mechanisms at play. Specifically, we

---

look to empirically separate and analyze the role celebrity endorsements play in providing quality information to consumers from the impact of (social) prestige and do so with the focus on Tiger Woods.

2.1 Data

We focus our attention on the setting of the US golf equipment market, specifically that of drivers during the period June 2008–June 2011 (36 months). Using this data we employ our new dynamic discrete choice methodology to obtain estimates of consumer preferences to disentangle Tiger Wood’s endorsement effect into prestige and information. The data consist of aggregate monthly sales data and prices for all drivers sold in the US market during June 2008 through June 2011 for two consumer groups: those who belong to a private golf club and those who do not. Our primary sales data originates from Golf Datatech a third party collector of retail sales data within the golf industry. We also source advertising expenditure data from TNS Adspender.

In Table 1 we present summary statistics regarding the data. Most notable is the discrepancy of sales across groups with players of public golf course purchasing roughly on average 4 times more drivers than golfer who play at private golf clubs each month. Additionally, on average in each month there are roughly 55 unique drivers on the market with 118 unique products available during our data period.

In order to begin our analysis, we define the respective market size for each consumer segment to generate market shares, given that we only have aggregate monthly sales of each product and group. To do so, we use data from Golf Datatech’s Summer 2009 Golf Product Attitude and Usage (A&U) Survey, which surveys 1000 golfers in the summer of 2009.

To begin, the total number of US golfers in 2009 was 26.1 million. We then leverage the A&U survey to determine the number of golfers in the market for a driver in a given period of time. The A&U survey asks whether a golfer is planning on purchasing a driver in the next twelve months. Using the tabulation from this question, we assume that 35 percent of the golfing population is actively in the market (includes those who answered yes and maybe). Next, we define how many of those golfers are a part of the two defined segments: those who belong to a private golf club and those who do not. Again, we leverage the A&U survey. Unfortunately, the survey does not ask such a specific question. Rather it poses two related questions that enable us to confidently

---

6 Note the data is collected at the retail channel level (on-course and off-course), but below we present a strong argument for such a classification using additional survey data.
### Table 1: Data Summary

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales: Private</td>
<td>505.87</td>
<td>(257.48)</td>
</tr>
<tr>
<td>Sales: Public</td>
<td>1929.50</td>
<td>(675.57)</td>
</tr>
<tr>
<td>Advertising by Brand (in Millions)</td>
<td>0.889</td>
<td>(0.682)</td>
</tr>
<tr>
<td>Price</td>
<td>202.38</td>
<td>(12.04)</td>
</tr>
<tr>
<td>Number of Products (Golf Drivers)</td>
<td>55.27</td>
<td>(3.61)</td>
</tr>
<tr>
<td>Total Number of Products (Golf Drivers)</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>Total Number of Brands</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Total Number of Observation</td>
<td>3,980</td>
<td></td>
</tr>
</tbody>
</table>

conclude the correct percentage of consumers belonging to a private golf club and those who do not. The first question asks where golfers most frequently play their golf (at a public or private course). The answer to this question reveals that 30 percent of golfers play private golf courses most frequently. A second question asks where golfers purchase their equipment. The results to this latter question illustrate that roughly 30 percent of golfers purchase at on-course golf shops. It is these two survey questions and there corresponding results that allow us to confidently assume 30 percent of golfers belong to a private club and purchase their golf equipment at on-course golf shops whereas 70 percent do not belong to a private golf club and purchase at off-course facilities. Given this information, we determine the initial month’s market size for each segment is 2.74 million golfers for private golf club members and 6.4 million golfers for those who do not belong to a private club and purchase equipment at off-course stores.

### 2.1.1 Tiger Woods Endorsement

In 1996, Tiger Woods left behind his golf scholarship at Stanford University to turn professional. Upon officially declaring himself a professional golfer he entered into endorsement contracts with Nike Golf and Titleist. The Titleist contract was for five years, 20 million dollars, and required him to endorse Titleist equipment, whereas the Nike contract was for 40 million dollars, for five years.\footnote{We certainly recognize that this definition is not perfect but it does provide a reasonable estimate. Additionally, the implicit assumption regarding our definition of market size is that there is no mixture occurring (e.g. those who do not belong to a private club, purchase equipment at a green-grass golf shop)}
years, and required him to endorse Nike Golf apparel. The reasoning for Tiger Woods’ endorsing only Nike apparel at the time was that Nike had yet to enter the golf equipment market. Nike’s entry into equipment began in February 1999, with the introduction of Nike golf balls, and in 2002, with clubs. In June of 2000, Tiger Woods officially began his switch to Nike equipment, with a five year endorsement contract (apparel, shoes, and equipment). He was the first professional player to switch golf balls from Titleist to Nike and received 100 million dollars from Nike Golf for doing so. After the five year contract expired, Tiger Woods and Nike entered into another five year, 100 million dollar endorsement deal. Finally in 2009, Tiger Woods’ found himself at the center of an infidelity scandal, which was extensively followed by the press starting shortly after Thanksgiving 2009 and continued its coverage through 2010. As a result, Tiger Woods’ 2011 contract was cut in half to 10 million dollars per year.\footnote{Dashiell Bennett, “Is Tiger Woods Running Out Of Money?”, Business Insider, Jul 15, 2011, \url{https://www.businessinsider.com/tiger-woods-broke-nike-2011-7}.}

2.1.2 Capturing Tiger Woods’ (and Others) Endorsement Effect

We capture a professional golfer’s endorsement effect by employing a measure of long term performance. We follow \cite{Chung2013}, which takes a complementary view, yet we deviate from their approach in capturing the overall endorsement effect. Rather than using the inverse of the golfers world golf ranking, our approach instead measures player quality through the golfer’s world ranking points. Both measures are based on a rolling two-year performance. But, by employing a points measure, we are able to capture more variation relative to rankings since Tiger Woods was ranked 1 in the world for 281 consecutive weeks. The use of world golf points also captures the true dominance of Tiger Woods’ whereas the measure of World Golf Ranking is unable to capture the separation between rankings (e.g. ranking of 1 with 19.234 vs a ranking of 2 and 9.742 points) by definition. As we will present below, this measure enters a consumer’s utility function as a consumer may prefer an endorsement from a professional golfer with high long-term points. Why? Because a product endorsed by a professional golfer may create an image that is recognizable by others (social prestige) leading consumers to associate with the golfer through the purchasing of the same brand or product the professional golfer uses. But it may also convey information about the quality of a product, if an endorser’s points is significantly high. For instance, if Tiger Woods is ranked number one and has a sizeable amount of points, this could signal to consumers that the brand or product Tiger Woods is playing with is of high quality. Again, this is
Table 2: Professional Golfer–Brand Endorsement Pair

<table>
<thead>
<tr>
<th>Prof. Golfer</th>
<th>Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam Scott</td>
<td>Titleist</td>
</tr>
<tr>
<td>Bubba Watson</td>
<td>Ping</td>
</tr>
<tr>
<td>Dustin Johnson</td>
<td>TaylorMade</td>
</tr>
<tr>
<td>Jason Day</td>
<td>TaylorMade</td>
</tr>
<tr>
<td>Justin Rose</td>
<td>TaylorMade</td>
</tr>
<tr>
<td>KJ Choi</td>
<td>Nike</td>
</tr>
<tr>
<td>Lee Westwood</td>
<td>Ping</td>
</tr>
<tr>
<td>Luke Donald</td>
<td>TaylorMade</td>
</tr>
<tr>
<td>Phil Mickelson</td>
<td>Callaway</td>
</tr>
<tr>
<td>Rory McIlroy</td>
<td>Nike</td>
</tr>
<tr>
<td>Sergio Garcia</td>
<td>TaylorMade</td>
</tr>
<tr>
<td>Tiger Woods</td>
<td>Nike</td>
</tr>
</tbody>
</table>

because the number one player in the world is playing it to earn prize money.

The functional form we employ to capture the endorsement effect of a given player on the sales of the endorsed brand $b$, by golfer $r$ in period $t$ is

$$Endow_{r,b,t} = \begin{cases} \ln \left( \text{World Golf Points}_{r,t} \right) & \text{if } D_{r,b,t} = 1, \\ 0 & \text{otherwise} \end{cases}$$

where $D_{r,b,t}$ is an indicator equaling one if player $r$ endorses brand $b$ in period $t$ and zero otherwise.

Let $Endow_{b,t}$ be the vector of endorsement effect of all players on brand $b$. By taking into account the variability of skill level over time, we assume that if there exists an endorsement effect, it will be larger when a player’s world golf points is higher. In practice, we include any golfer who was listed in the top 20 in the World in June 2008 and June 2011. These golfers are Tiger Woods, Adam Scott, Bubba Watson, Dustin Johnson, Jason Day, Jim Furyk, Justin Rose, KJ Choi, Lee Westwood, Luke Donald, Phil Mickelson, Rory McIlroy, and Sergio Garcia. Table 2 presents the respective brand that each golfer endorses. Figure 1 illustrates the variation in the selected players’ world golf points.
3 The Consumer Model

Our dynamic model follows the previous literature on dynamic discrete choice models of demand, particularly those that employ market level data. Although the model is general, it is especially appropriate for durable products (e.g. golf drivers in our empirical study), since consumers in such markets are typically forward looking and weigh the trade-off of making a purchase now versus the option value of waiting.

There will be two dynamic aspects in the model. The first aspect rests on consumer’s belief about the transition of product and market characteristics over the time. For example, if the product prices are expected to decline over time (see Figure 2) then the consumer is incentivized to wait.

This dynamic aspect will be reflected by the definition of lifetime payoffs of different alternatives. One interesting feature of our method is that we can identity and estimate the flow utility functions and the discount factor without restricting consumer’s belief about the law of state transition. For example, using our method, researchers do not have to assume that consumers have rational expectation regarding state transition distribution.

The second aspect concerns about the non-random attrition of consumers. For the market of durable goods, it is reasonable to assume that consumers will exit the market after purchasing. Depending on the market size, such non-random attrition of consumers could change significantly the distribution of unobserved consumers heterogeneity over the course of time. The consequence of such attrition has not been fully explored; we show that the attrition of consumers dramatically
changes the way of identifying the discount factor.

Though our estimation method belongs to the genre of two step CCP estimation [Hotz and Miller (1993)], both our CCP and post-CCP estimation are new.

3.1 The Model

For a market, we observe $T$ periods indexed by $t = 1, \ldots, T$. The timeline of our model is the following. In each period $t$, consumer $i$ observes the state of market $\Omega_{it}$ and considers whether or not to purchase a durable product from the available goods $1, \ldots, J$. The associated expected lifetime payoffs are $v_{i1t}, \ldots, v_{iJt}$. If she decides to purchase, she then chooses which to buy by comparing payoffs $v_{i1t}, \ldots, v_{iJt}$. Once a consumer has purchased a product, she exits the market completely, hence purchasing is a terminal action in our model. If she decides not to purchase now, she chooses the outside good $0$ and remains in the market for the next period. In other words, the outside good $0$ is “wait-and-see”. Let $v_{i0t}$ denote her discounted expected future value. Later, we will assume that there are some preference shocks that follow the extreme value distribution.

---

For the simplicity of exposition, we let the product space being fixed. The arguments do not change if we consider time varying choice set.
(EVD), with the implied CCP of choosing product \( j = 0, 1, \ldots, J \) taking the familiar logit form,

\[
\sigma_{ijt} = \frac{\exp(v_{ijt})}{\exp(v_{i0t}) + \sum_{k=1}^{J} \exp(v_{ikt})}.
\]

The rest of this section is to formalize what the payoff \( v_{ijt} \) is in our dynamic model.

The lifetime payoff is a “sum” of per period or flow utilities. We first state the flow utilities. If consumer \( i \) does not purchase in period \( t \) (choosing the outside option), she receives the flow utility \( \varepsilon_{i0t} \) in period \( t \) and stays in the market. When consumer \( i \) purchases product \( j \) at time \( t \), her flow indirect utility during the purchase period \( t \) is

\[
\delta_j + \gamma' X_{jt} + \xi_{jt} - \alpha_i P_{jt} + \varepsilon_{ijt}.
\]

She then receives the identical flow utility \( \delta_j + \gamma' X_{jt} + \xi_{jt} \) in each period following her purchase. Here \( X_{jt} \) is a vector of observable product attributes other than price, \( P_{jt} \) is the price, \( \delta_j \) is product fixed effect, and \( \xi_{jt} \) is unobserved product characteristics. Assume \( E(\xi_{jt}) = 0 \). Let \( X_t \equiv (X'_{1t}, \ldots, X'_{Jt})' \), and \( P_t \) and \( \xi_t \) are defined similarly. Lastly, \( \varepsilon_{it} \equiv (\varepsilon_{i0t}, \ldots, \varepsilon_{iJt})' \) is a vector of utility shocks. The unobserved product characteristics \( \xi_{jt} \) can be serially correlated and correlated with prices and observable product characteristics.\(^{10}\) Let \( \Omega_{it} \equiv (X'_t, P'_t, \xi'_t, \varepsilon'_it)' \).

For our empirical application \( X_{jt} \) includes product age, the endorsement effects \( Endow_{b,t} \), the corresponding interaction terms for Tiger Wood’s infidelity scandal, and the log transformation of planned media spending in dollars \( Ad_{b,t} \) by brand \( b \).

Having stated the flow utilities, we now clarify how we model consumer heterogeneity (beyond the individual utility shocks). Consumers are heterogeneous in their price coefficient \( \alpha_i \), which depends on observed discrete groups \( g = 1, \ldots, G \) and unobserved continuous heterogeneity \( U_i \).\(^{11}\)

In our golf application, we observe two consumer groups: those who belong to a private golf club \((g = \text{Private})\) and those who do not \((g = \text{Public})\). In general, these groups can be defined by income bracket, age group, sex etc. Suppose that there are \( G \) groups in the market facing the same price \( P_t \) and products characteristics \( (X_t, \xi_t) \). We use a vector of dummy variables \( D_i \equiv (D^{(1)}_i, \ldots, D^{(G)}_i)' \) to indicate the membership—\( D^{(g)}_i = 1 \) if consumer \( i \) belongs to group \( g \), and \( D^{(g)}_i = 0 \) otherwise. Assume \( \sum_{g=1}^{G} D^{(g)}_i = 1 \). Consider

\[
\alpha_i = \alpha^{(1)} + \tau^{(2)} D^{(2)}_i + \cdots + \tau^{(G)} D^{(G)}_i + \omega U_i,
\]

\(^{10}\)Note the explicit assumption that unobserved product characteristics \( \xi_{jt} \) does not vary across consumer in the same market \( t \). This is commonly assumed in the vast literature of the BLP models (e.g. Berry and Haile [2020]).

\(^{11}\)We limit our attention to the heterogeneity in price coefficient \( \alpha_i \). It is easy to generalize our approach to some or all components of \( \gamma \) (see Appendix E).
where $\tau^{(g)}$ captures the between group variation, and $\omega U_i$ is idiosyncratic unobserved price preference, which captures the within group variation of the price coefficient. Below, we normalize the variance of $U_i$ to be 1, hence $\omega \geq 0$ controls the size of within group variation. When we plot the probability density functions (PDF) of consumers’ price coefficients, the PDFs of two similar groups of consumers will “overlap” (see Panel I of Figure 4). We will use the information of the overlapping groups to identify and estimate the model. It will be convenient to define $\tau^{(1)} = 0$. It is tempting to conclude that $\alpha^{(1)} + \tau^{(g)}$ refers to the mean price coefficient, which will determine price elasticities, for group $g$. This is in general not correct because of the attrition of consumers. The attrition will makes the distribution of $U_i$ skewed over the course of time, hence $\alpha^{(1)} + \tau^{(g)}$ is no longer the mean of price coefficient $\alpha_i$. We will return to this point later.

Below, we say that a consumer $i$ is of type-$(g,U)$ if she is from group $g$ and $U_i = U$, hence her price coefficient $\alpha_i = \alpha^{(1)} + \tau^{(g)} + \omega U$. For a consumer $i$ of type-$(g,U)$, we just write $v^{(g)}_{ijt}(U)$ to denote her expected lifetime payoffs $v_{ijt}$ from product $j$.

**Assumption 1** (Type-I EVD). Assume that utility shocks $\varepsilon_{ijt}$ are serially independent, follow type 1 EVD, and they are independent of $(D_i, U_i, X_t, P_t, \xi_t)$.

**Assumption 2** (Markov Process). $\Pr(\Omega_{i,t+1} \mid \Omega_{it}, \Omega_{i,t-1}, \ldots) = \Pr(\Omega_{i,t+1} \mid \Omega_{it})$.

**Assumption 3** (Conditional Independence). For all $t$, we have (i) $\Omega_{i,t+1} \perp \varepsilon_{it} \mid (X_t, P_t, \xi_t)$; (ii) $\varepsilon_{i,t+1} \perp \Omega_{it} \mid (X_{t+1}, P_{t+1}, \xi_{t+1})$; (iii) if two consumers have the same price sensitivity, they have the same belief about the conditional distribution of $(X_{t+1}, P_{t+1}, \xi_{t+1})$ given $(X_t, P_t, \xi_t)$.

The above assumptions restricting consumers’ belief about the state transitions are standard in the literature. It is worth mentioning that we only assume that consumers’ beliefs satisfy the Markov property and certain conditional independence. What we do not assume is that the state transition distribution according to consumers’ beliefs is identical to the observed state transition distribution in the data, which is implicitly assumed in the literature as rational expectation. The last part of Assumption 3 says that the heterogeneity in price sensitivity also determines the belief about the transition of the market state variables $(X_t, P_t, \xi_t)$. This is still weaker than the common rational expectation assumption that will assume that the transition distribution of the market

---

$^{12}$To take account of heteroskedasticity (group varying within group variation), it is straightforward to consider the more flexible specification $\alpha_i = \alpha^{(1)} + \tau^{(2)} D^{(2)}_i + \cdots + \tau^{(G)} D^{(G)}_i(U) + (\alpha^{(1)} \omega^{(1)} D^{(1)}_i + \cdots + \omega^{(G)} D^{(G)}_i U_i)$, where $\omega^{(g)}$ controls the variation of $\alpha_i$ within group $g$. 

15
state variables \((X_t, P_t, \xi_t)\) is the same as contained in data—which is apparently the same for all consumers in the market.

### 3.2 Consumer’s Decision

With flow utilities specified, we can now combine them to determine the lifetime payoffs. Consider a consumer \(i\) of type \((g, U)\)—that is this consumer is from group \(g\) and her unobserved price sensitivity is \(U\). We can write the Bellman equation in terms of the value function \(V_t^{(g)}(\Omega_{it}, U)\) as follows\(^{13}\)

\[
V_t^{(g)}(\Omega_{it}, U) = \max \left( \varepsilon_{i0t} + \beta E \left[ V_{t+1}^{(g)}(\Omega_{it+1}, U) \mid \Omega_{it} \right], \max_{j \in \{1, \ldots, J\}} v_{jt}^{(g)}(U) + \varepsilon_{ijt} \right).
\]

The first term within brackets is the present discounted utility associated with the decision to not purchase any product, i.e. choosing the outside option \(j = 0\), in period \(t\). The discount factor is \(\beta \in [0, 1)\). The choice of not purchasing in period \(t\) provides flow utility \(\varepsilon_{i0t}\), and a term that captures expected future utility conditional on the current state being \(\Omega_{it}\). This last term is the option value of waiting to purchase. The second term within brackets indicates the value associated with the purchase of a product. Given the fact that consumers exit the market after the purchase of any product, a consumer’s choice specific value function can be written as the sum of the current period \(t\) utility and the stream of utilities in periods following purchase:

\[
v_{jt}^{(g)}(U) = \frac{\delta_j + \gamma^t X_{jt} + \xi_j t}{1 - \beta} - (\alpha^{(1)} + \tau^{(g)} + \omega U) P_{jt}, \quad j = 1, \ldots, J.
\]

Recall \(\alpha^{(1)} + \tau^{(g)} + \omega U\) is the price coefficient of a consumer of type \((g, U)\). Let

\[
v_{0t}^{(g)}(U) = \beta E \left[ V_{t+1}^{(g)}(\Omega_{it+1}, U) \mid \Omega_{it} \right] = \beta E \left[ V_{t+1}^{(g)}(X_{t+1}, P_{t+1}, \xi_{t+1}, U) \mid X_t, P_t, \xi_t \right],
\]

where \(V_{t+1}^{(g)}(X_{t+1}, P_{t+1}, \xi_{t+1}, U) \equiv E \left( V_{t+1}^{(g)}(X_{t+1}, P_{t+1}, \xi_{t+1}, \varepsilon_{i, t+1}, U) \mid X_{t+1}, P_{t+1}, \xi_{t+1} \right)\), and the expectation is taken over \(\varepsilon_{i, t+1}\). The second identity follows from applying Assumption 3.

Up to now, we have formalized what the payoffs \(v_{i0t}, v_{1t}, \ldots, v_{i,Jt}\) are in the definition of the CCP. They are \(v_{ijt} = v_{jt}^{(g)}(U)\) for a consumer of type \((g, U)\). Correspondingly, we use \(\sigma_{jt}^{(g)}(U)\) to denote the probability of buying product \(j\) in period \(t\) provided that consumer \(i\) is of type-\((g, U)\):

\[
\sigma_{jt}^{(g)}(U) = \frac{\exp(v_{jt}^{(g)}(U))}{\exp(v_{0t}^{(g)}(U)) + \sum_{k=1}^{J} \exp(v_{kt}^{(g)}(U))}.
\]

Particularly, in our golf market study, we have \(\sigma_{jt}^{(Public)}(U)\) and \(\sigma_{jt}^{(Private)}(U)\).

---

\(^{13}\)We often mute the argument the state variables \(\Omega_{it}\) in writing alternative specific value function for exposition simplicity.
Remark 1 (Myopic Model as One Special Case). A special case of the above model is when $\beta = 0$, that is assuming that consumers are myopic. In the myopic case, $v^{(g)}_{it}(U) = 0$, $v^{(g)}_{jt}(U) = \delta_{jt} + \gamma'X_{jt} - (\alpha^{(1)} + \tau^{(g)} + \omega U)P_{jt} + \xi_{jt}$ for product $j = 1, \ldots, J$. This setting then becomes a BLP discrete choice demand model.

3.3 Dynamic Selection Problem

In period $t$, the observed market share of product $j$ among the group $g$ of consumers is obtained by integrating the individual CCP, $\sigma^{(g)}_{jt}(U)$, over unobserved price sensitivity $U$. Let $F^{(g)}_{t}(u)$ be the cumulative distribution function (CDF) of the unobserved price sensitivity $U_i$ of consumers within group $g$ in period $t$. The market share of product $j$ in group $g$ in period $t$ is

$$S^{(g)}_{jt} = \int \sigma^{(g)}_{jt}(u) \, dF^{(g)}_{t}(u).$$

In our golf market data, we can observe two group market share $S^{(Public)}_{jt}$ and $S^{(Private)}_{jt}$.

Some comments about the composition of consumers in different periods, i.e. $F^{(g)}_{t}(U)$, are due here. The composition of consumers depends on whether or not consumers would remain in the market after purchasing. For the case of non-durable goods, like ready-to-eat oatmeal, consumers remain in the market after purchasing, hence $F^{(g)}_{t}(u)$ does not vary across time. For the case of durable goods, it is reasonable to assume that consumers will exit the market after purchasing. It is expected that such non-random attrition (or dynamic selection) of consumers could significantly change the distribution of unobserved price sensitivity $F^{(g)}_{t}(u)$, depending on the rate of attrition. This is a “selection problem” in dynamic discrete choice. In the rest, we will show that in order to fix the dynamic selection problem, it is essential to obtain the CCP $\sigma^{(g)}_{jt}(U)$ as a function of unobserved heterogeneity $U$.

The attrition has the following implications in theory, and our simulation studies show that ignoring attrition could cause substantial bias in practice. First, it changes the distribution of price sensitivity $U_i$ over the course of time even after controlling the demographic groups. It is intuitive that attrition “pushes” the distribution of $U_i$ to concentrate more and more on the price sensitive area over the time. Second, attrition also changes the composition of groups. Attrition pushes the distribution of groups to concentrate more on price sensitive groups—over the time, we see bigger and bigger weights on price sensitive groups. Lastly, the rate of attrition is different for different groups. Consumers in the group with lower average price elasticity would leave the market faster.

Below, we first assume that the unobserved price sensitivity follows a normal distribution at
Depending on whether or not there is attrition, Proposition 1 provides a formula of the distribution of price sensitivity for the subsequent periods in terms of the CCP function $\sigma^{(g)}_{jt}(U)$.

**Assumption 4** (Initial distribution of unobserved price sensitivity). For each of the G groups, assume that in the first period the unobserved consumer price sensitivity follows the standard normal distribution, that is $F^{(g)}_1(u) = \Phi(u)$. Here $\Phi(u)$ denotes the CDF of the standard normal distribution, and let $\phi(u)$ denote the respective PDF.

**Proposition 1** (Distribution of unobserved heterogeneity due to attrition). Suppose Assumptions 1 to 4 holds. Let $f^{(g)}_t(U)$ be the PDF of the unobserved price sensitivity $U$ in period $t$ and group $g$. We have that

$$f^{(g)}_t(u) = \phi(u) \times \Gamma^{(g)}_t(u),$$

where $\Gamma^{(g)}_t(u)$ satisfies the following.

(i) (Case I: No attrition) If consumers remain in the market after purchasing, $\Gamma^{(g)}_t(u) = 1$ for all $(u,t,g)$;

(ii) (Case II: Attrition) If consumers left the market after purchasing,

$$\Gamma^{(g)}_1(u) = 1, \quad \Gamma^{(g)}_t(u) = \prod_{s=1}^{t-1} \frac{\sigma^{(g)}_0(u)}{S^{(g)}_s}, \quad t \geq 2.$$

Note that the definition of $\Gamma^{(g)}_t(u)$ implies the following recursive formula:

$$\Gamma^{(g)}_1(u) = 1, \quad \Gamma^{(g)}_{t+1}(u) = \Gamma^{(g)}_t(u) \times \frac{\sigma^{(g)}_t(u)}{S^{(g)}_t}.$$

### 4 Identification and Estimation When the CCP Is Known

Assuming that we have estimated CCP $\sigma^{(g)}_{jt}(U)$ as a function of $U$ and known the distribution of $U$ for each period and group, $F^{(g)}_t(U)$, by the formulas in Proposition 1 we show how to identify and estimate the remaining structural parameters leaving the CCP estimation to the next section. The conclusion we will arrive at is that in order to estimate the structural parameters in consumer preferences, including the discount factor, one simply needs to run the two linear regressions, eq. (Linear-Reg-1) and eq. (Linear-Reg-2), below. We provide two remarks at the end explaining the consequence of dynamic selection on the model estimation, and the intuition why we can estimate the model while being agnostic about consumers’ belief about the state transition.

---

14Proposition A.1 in the appendix describes the variation of group composition when there is consumer attrition.
Model Parameters except for Discount Factor and Product Fixed Effect

Identification and estimation of model parameters outside of the discount factor and product fixed effects start from the following moment conditions taken from Chou, Derdenger and Kumar (2019). They are based on the observation that conditional on purchasing in period $t$, a consumer’s choice about which one to buy does not depend on the unknown continuation value $v_{jt}(U)$. We choose product 1 as the reference product, which results in

$$\ln \left[ \frac{\sigma_{jt}(U)}{\sigma_{1t}(U)} \right] = v_{jt}(U) - v_{1t}(U).$$

By the definition of payoff functions $v_{jt}(U)$ in eq. (2), and integrating the above display over $U$ with respect its distribution function in period $t$, we have the first condition

$$\int \ln \left[ \frac{\sigma_{jt}(U)}{\sigma_{1t}(U)} \right] dF_{jt}(U) + \omega(P_{jt} - P_{1t}) \int U dF_{jt}(U) = \frac{\delta_j - \delta_1}{1 - \beta} + (X_{jt} - X_{1t})' \frac{\gamma}{1 - \beta} - (\alpha^{(1)} + \tau^{(g)})(P_{jt} - P_{1t}) + \xi_{jt} - \xi_{1t}. \quad (4)$$

Provided that the CCP function $\sigma_{jt}(U)$ and the distribution function $F_{jt}(U)$ are known, the above two terms involving integration are also known. We then identify $(\delta_2 - \delta_1)/(1 - \beta), \ldots, (\delta_J - \delta_1)/(1 - \beta), \gamma/(1 - \beta), \alpha^{(1)}, \tau \equiv (\tau^{(2)}, \ldots, \tau^{(G)})'$ and $\omega$ using 2SLS. 2SLS is used because price $P_{jt}$ is usually correlated with $\xi_{jt}$.

In the actual estimation, we can simplify the procedure. It turns out that we can directly estimate $\tau$ and $\omega$ when we estimate the CCP functions $\sigma_{jt}(U)$. Focus on the first group, and let

$$Y_{jt} \equiv \int \ln \left[ \frac{\sigma_{jt}^{(1)}(U)}{\sigma_{1t}^{(1)}(U)} \right] dF_{jt}^{(1)}(U) + \omega(P_{jt} - P_{1t}) \int U dF_{jt}^{(1)}(U).$$

The above eq. (4) becomes

$$Y_{jt} = \frac{\delta_j - \delta_1}{1 - \beta} + (X_{jt} - X_{1t})' \frac{\gamma}{1 - \beta} - \alpha^{(1)}(P_{jt} - P_{1t}) + \xi_{jt} - \xi_{1t}, \quad \text{(Linear-Reg-1)}$$

for $j = 2, \ldots, J$ and is simple to estimate with $Y_{jt}$ known. Recall $\tau^{(1)} = 0$. We define $Y_{jt}$ because we will estimate this dependent variable as a whole directly during CCP estimation (Proposition 2).

---

15 Throughout the paper, we assume there are underlying instrumental variables (IV) available—for example, the prices of the product in other markets (Nevo 2001) or the characteristics of other products (Berry, Levinsohn and Pakes 1995). In our study of golf market, we will instrument both price and advertising expenditure.
Model Parameters: Discount Factor and Product Fixed Effect

Identification and estimation of $\beta$ and $\delta_1$ originates from a second condition that comes from
\[
\ln(\sigma^{(g)}_{it}(U)/\sigma^{(g)}_{0t}(U)) = v^{(g)}_{it}(U) - v^{(g)}_{0t}(U)
\]
and describes the trade-off from buying now and waiting. By the definition of the payoffs, it becomes
\[
\ln \left[ \frac{\sigma^{(g)}_{it}(U)}{\sigma^{(g)}_{0t}(U)} \right] = \frac{\delta_1}{1-\beta} + X'_{it} \gamma - (\alpha^{(1)} + \tau^{(g)} + \omega U)P_{it} + \frac{\xi_{it}}{1-\beta} - \beta E[\tilde{V}^{(g)}_{t+1}(U) \mid X_t, P_t, \xi_t].
\] (5)

We will identify the discount factor $\beta$ and product fixed effect $\delta_1$ using the second condition. Before that, we will first show that for any fixed unobservable price sensitivity $\tilde{U}$ (e.g. $\tilde{U} = 0$),
\[
E[W^{(g)}_t(\tilde{U})] = \delta_1 + \beta E[W^{(g)}_{t+1}(\tilde{U}) + \ln \sigma^{(g)}_{0,t+1}(\tilde{U})],
\] (Linear-Reg-2)
where $W^{(g)}_t(\tilde{U})$ is estimable and defined below. This equation will give rise to an estimable formula of $(\beta, \delta_1)'$.

We obtain eq. (Linear-Reg-2) from eq. (5) with four steps. Note that after running 2SLS of eq. (Linear-Reg-1), we already know many parameters including $\sigma^{(g)}_{jt}(U)$, $\tau$, $\omega$, $\gamma/(1-\beta)$, and $\alpha^{(1)}$. Step 1 is to define $W^{(g)}_t(U)$ by combining the terms that are already known in eq. (5). Let
\[
W^{(g)}_t(U) \equiv \ln \left[ \frac{\sigma^{(g)}_{it}(U)}{\sigma^{(g)}_{0t}(U)} \right] - \left[ X'_{it} \frac{\gamma}{1-\beta} - (\alpha^{(1)} + \tau^{(g)} + \omega U)P_{it} + \frac{\delta_1}{1-\beta} + \frac{\xi_{it}}{1-\beta} \right].
\]
Note that $W^{(g)}_t(U)$ is known for any $U$ after the CCP estimation and the above 2SLS linear regression. Step 2 is to convert the unknown integrated value function $\tilde{V}^{(g)}_{t+1}(U)$ into something we have already known using the well known expectation maximization formula of logit model:
\[
\tilde{V}^{(g)}_{t+1}(U) = v^{(g)}_{1,t+1}(U) - \ln \sigma^{(g)}_{1,t+1}(U).
\]
We have
\[
\tilde{V}^{(g)}_{t+1}(U) = -[W^{(g)}_{t+1}(U) + \ln \sigma^{(g)}_{0,t+1}(U)] + \frac{\delta_1}{1-\beta} + \frac{\xi_{1,t+1}}{1-\beta}.
\]
In step 3, we rewrite eq. (5) in terms of $W^{(g)}_t(U)$ and conclude
\[
W^{(g)}_t(U) = \delta_1 + \frac{\xi_{it}}{1-\beta} + \beta E \left( W^{(g)}_{t+1}(U) + \ln \sigma^{(g)}_{0,t+1}(U) - \frac{\xi_{1,t+1}}{1-\beta} \right) \bigg| X_t, P_t, \xi_t , \] (6)
Lastly, in step 4, for a fixed unobserved price sensitivity $\tilde{U}$, we take unconditional expectation with respect to $(X_t, P_t, \xi_t)$, and use the condition $E(\xi_{1t}) = E(\xi_{1,t+1}) = 0$ and the law of iterated expectation to reach eq. (Linear-Reg-2).

We now show how to identify the discount factor $\beta$ and product fixed effect $\delta_1$ using eq. (Linear-Reg-2). The expectations $E[W^{(g)}_t(\tilde{U})]$ and $E[\ln \sigma^{(g)}_{0t}(\tilde{U})]$ are taken over $(X_t, P_t, \xi_t)$ only with $\tilde{U}$ being
fixed for each group $g$ and each $t$. This expectation can be estimated by $T^{-1} \sum_{t=1}^{T} W_t^{(g)}(\tilde{U})$ when $(X_t, P_t, \xi_t)$ satisfies certain stationarity conditions.\footnote{We need the time series $(X_t, P_t, \xi_t)$ is ergodic, and for the chosen fixed $\tilde{U}$, $W_t^{(g)}(\tilde{U})$, as a function of $(X_t, P_t, \xi_t)$, satisfies certain continuity conditions.} With at least two groups (say 1 and 2), we have

\begin{align*}
E[W_t^{(1)}(\tilde{U})] &= \delta_1 + \beta E[W_{t+1}^{(1)}(\tilde{U}) + \ln \sigma_{0,t+1}^{(1)}(\tilde{U})] \\
E[W_t^{(2)}(\tilde{U})] &= \delta_1 + \beta E[W_{t+1}^{(2)}(\tilde{U}) + \ln \sigma_{0,t+1}^{(2)}(\tilde{U})].
\end{align*}

We can solve the discount factor $\beta$ from the above linear system of equations, and obtain

\[ \beta = \frac{E[W_t^{(1)}(\tilde{U})] - E[W_t^{(2)}(\tilde{U})]}{E[W_{t+1}^{(1)}(\tilde{U})] - E[W_{t+1}^{(2)}(\tilde{U})] + E[\ln \sigma_{0,t+1}^{(1)}(\tilde{U})] - E[\ln \sigma_{0,t+1}^{(2)}(\tilde{U})]}. \]

The discount factor can be estimated by the sample analog of above formula. Knowing the discount factor $\beta$, we have the fixed effect $\delta_1$. The other product fixed effects $\delta_2, \ldots, \delta_J$ are automatically determined since we have known $(\delta_j - \delta_1)/(1 - \beta)$.

**Remark 2** (How does the attrition affect the estimation?) Non-random attrition affects the estimation in two ways. The first is apparent. Ignoring the attrition is to let $F_t^{(1)}(u)$ be the distribution function of $U_i$ in the first period, i.e. $F_t^{(1)}(u) = \Phi(u)$. Reading the definition of the dependent variable $Y_{jt}$ of eq. (Linear-Reg-1), it is apparent that ignoring the attrition will misspecify the $F_t^{(1)}(U)$, causing bias in estimating preference parameters. The second is more subtle. Attrition will create a nonstationarity problem, which can be clearly seen from the estimation of discount factor.

Reading the equation of identifying the discount factor, eq. (Linear-Reg-2), one may wonder why not integrate out the unobserved price sensitivity $U$ since we also know its distribution function $F_t^{(g)}(U)$? It turns out this will lead us to a biased estimator of the discount factor in the presence of attrition. To understand why, consider $\tilde{W}_t^{(g)} \equiv \int W_t^{(g)}(U) \, dF_t^{(g)}(U)$. For a given period $t$, we have

\[ \tilde{W}_t^{(g)} = \int \ln \left[ \frac{\sigma_{0,t}^{(g)}(U)}{\sigma_{0,t}^{(g)}(U)} \right] \, dF_t^{(g)}(U) - X_t^1 \gamma + \beta (\alpha^{(1)} + \tau^{(g)}) P_t + \omega P_t \int U \, dF_t^{(g)}(U), \]

by the definition of $W_t^{(g)}(U)$. The integrated term $\tilde{W}_t^{(g)}$ is still estimable using our approach for each period $t$. It is also easy to verify that eq. (Linear-Reg-2) becomes

\[ \mathbb{E}(\tilde{W}_t^{(g)}) = \delta_1 + \beta \mathbb{E}\left[ \tilde{W}_{t+1}^{(g)} + \int \ln \sigma_{0,t+1}^{(g)}(U) \, dF_t^{(g)}(U) \right]. \]
We then have an alternative formula of the discount factor,

$$\beta = \frac{E(\bar{W}(1)_t) - E(\bar{W}(2)_t)}{E(\bar{W}(1)_{t+1}) - E(\bar{W}(2)_{t+1}) + E[\int \ln \sigma_{0,t+1}(U) \ d F_t^{(1)}(U)] - E[\int \ln \sigma_{0,t+1}(U) \ d F_t^{(2)}(U)],}$$

with at least two groups 1 and 2.

The problem is how to estimate $E(\bar{W}(g)_t)$ and $E[\int \ln \sigma_{0,t+1}(U) \ d F_t^{(g)}(U)]$? Taking $E(\bar{W}(g)_t)$ for example, it is tempting to use $T^{-1} \sum_{t=1}^{T} \bar{W}(g)_t$ as the estimator, however it is an inconsistent estimator when there is non-random attrition of consumers. The underlying reason is that even though $(X_t, P_t, \xi_t)$ satisfies certain stationarity conditions, $\bar{W}(g)_t$ is still non-stationary due to the attrition of consumer. In particular, both

$$\int \ln \left[ \frac{\sigma_{1t}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right] \ d F_t^{(g)}(U) \quad \text{and} \quad \int U \ d F_t^{(g)}(U)$$

in the definition of $\bar{W}(g)_t$ are nonstationary. Intuitively, consumers who are less price sensitive purchase and leave the market earlier making the average $\int U \ d F_t^{(g)}(U)$ drift upward over time. Due to this non-stationary property (caused by attrition), the temporal average will not converge in probability to $E(\bar{W}(g)_t)$, which is indeed only well defined for a fixed period. In order to estimate $E(\bar{W}(g)_t)$, one needs access to a large number of cross sectional markets for each period. Such data access is usually unavailable in empirical studies. The same comments apply to the integral term of CCP functions.

Our approach works here because we can recover the CCP function at such a precise level that the CCP for any given unobserved price sensitivity, i.e. $\sigma^{(g)}_{jt}(\tilde{U})$ herein, can be obtained. We then can avoid the problem of attrition by focusing on one type of consumer (in terms of fixing $U$). After fixing $U$, all variables, like $W_t^{(g)}(U)$, involve only stationary process $(X_t, P_t, \xi_t)$. We then can use the temporal average to estimate them and the discount factor. This again highlights the importance of recovering the CCP $\sigma^{(g)}_{jt}(U)$ in order to fix the dynamic selection problem.

Remark 3 (Why can we identify the model without a specification of belief?). Note that our estimation of flow utility functions and the discount factor does not rely on the specification of the law of state transition that is embedded in the conditional expectation $E[g(X_{t+1}, P_{t+1}, \xi_{t+1}) | X_t, P_t, \xi_t]$ (here $g(X_{t+1}, P_{t+1}, \xi_{t+1})$ denotes a generic function of $(X_{t+1}, P_{t+1}, \xi_{t+1})$). For example, a consumer’s belief about the state transition may not be that of rational expectation.

For the estimation of flow utility functions (2SLS of regression eq. [Linear-Reg-1]), the intuition is that because purchasing in our model is a terminal choice, the comparison between two products
once a consumer has decided to buy one of them does not involve the future valuation, hence it does not involve the belief about the state transition distribution.

The intuition for why we can estimate the discount factor without knowing a consumer’s belief regarding the law of state transition is less transparent. The key step of identifying the discount factor is to take unconditional expectation for eq. (6). The trick of unconditional expectation can be understood by the following story. Suppose consumer A has rational expectation about the quality of a car without further information (unconditional expectation). It is fine that A has an irrational belief about the quality of the car given the year it was manufactured (conditional expectation with year manufactured being the conditional variable). Taking unconditional expectation is to disregard the information of the year, hence the irrational belief due to year does not matter. In our model, taking unconditional expectation of eq. (6) is to get rid of the information about the current market state \((X_t, P_t, \xi_t)\), so that consumers belief about \((X_{t+1}, P_{t+1}, \xi_{t+1})\) given \((X_t, P_t, \xi_t)\) is irrelevant.

5 CCP Estimation by Using Overlapping Groups of Consumers

We have seen the essential role of the CCP \(\sigma_{jt}^{(g)}(U)\) in model estimation, particularly in the presence of dynamic selection. We propose a novel method of estimating \(\sigma_{jt}^{(g)}(U)\) using group market share data \(S_{jt}^{(g)}\) for \(g = 1, \ldots, G\) and certain constraints implied by the underlying structural model. We have \(G\) equations from the definition of group market shares \(S_{jt}^{(1)}, \ldots, S_{jt}^{(G)}\),

\[
S_{jt}^{(1)} = \int \sigma_{jt}^{(1)}(u) \, dF_t^{(1)}(u),
\]

\[
\vdots
\]

\[
S_{jt}^{(G)} = \int \sigma_{jt}^{(G)}(u) \, dF_t^{(G)}(u).
\]

It seems that there are \(G\) unknowns \(\sigma_{jt}^{(1)}(U), \ldots, \sigma_{jt}^{(G)}(U)\) in the above \(G\) equations—let alone that the CDF of unobserved price sensitivity \(F_t^{(g)}(U)\) also depends on the CCP of choosing the outside option in all previous periods, \(\sigma_{0t}^{(g)}(U), \ldots, \sigma_{0t-1}^{(g)}(U)\) according to Proposition 1. It seems hopeless to solve \(\sigma_{jt}^{(g)}(U)\). Below, we will argue that this is not the case.

5.1 Key Observation: Shifting CCP Across Similar Groups of Consumers

Our argument rests on the following observation: these \(G\) unknown functions \(\sigma_{jt}^{(1)}(U), \ldots, \sigma_{jt}^{(G)}(U)\) indeed can be viewed as one unknown CCP function after some transformation. To see this, suppose there are two income brackets, 1 (high-income) and 2 (low-income), in sample. Note that
the CCP is determined by comparing expected payoffs of different alternatives. The expected payoff of product \( j = 1, \ldots, J \) is simply

\[
v_{ijt} = \frac{\delta_j + \gamma'X_{jt} + \xi_{jt}}{1 - \beta} - \left( \alpha^{(1)} + \tau^{(2)} D_i^{(2)} + \omega U_i \right) P_{jt}.
\]

The expected payoff of the outside option depends on the expected future payoffs for all products \( j = 1, \ldots, J \). The unknown consumer’s type affects the choice probabilities by altering the expected payoffs, which can be only through the term \( \alpha_i = \alpha^{(1)} + \tau^{(2)} D_i^{(2)} + \omega U_i \). Now consider consumer \( h \) from high-income group 1, and consumer \( \ell \) from low-income group 2. Let \( U_h \) and \( U_\ell \) be the idiosyncratic price sensitivity relative to their respective group for the two consumers \( h \) and \( \ell \), respectively. Note that if \( U_h \) and \( U_\ell \) satisfy the condition that \( U_h = U_\ell + \tau^{(2)}/\omega \), we have the conclusion that these two consumers have the same price coefficient: \( \alpha_h = \alpha^{(1)} + \omega U_h = \alpha^{(1)} + \tau^{(2)} + \omega U_\ell = \alpha_\ell \). Intuitively, this says that though consumer \( h \) is from higher income group, she still has the same price sensitivity as a lower income consumer \( \ell \) because of \( h \)'s idiosyncratic relatively high price sensitivity. The same price coefficient further implies that the two CCPs \( \sigma_{jt}^{(2)}(U_{\ell}) = \sigma_{jt}^{(1)}(U_h) \) when \( U_h = U_\ell + \tau^{(2)}/\omega \). In summary, we conclude that two CCPs \( \sigma_{jt}^{(1)}(U) \) and \( \sigma_{jt}^{(2)}(U) \), viewed as a function \( U \), are essentially identical—we can obtain one by shifting the other along the axis of \( U \), i.e.

\[
\sigma_{jt}^{(2)}(U) = \sigma_{jt}^{(1)}(U + \tau^{(2)}/\omega).
\]

Figure 3 illustrates this observation using the CCP of two groups from our simulation studies. The essential observation is that the underlying structural model implies certain restrictions that can be used to transform the CCP functions of one focal group to get the CCP of the other groups. In addition, the underlying structural models also impose restrictions on CCP function \( \sigma_{jt}^{(1)}(U) \). The application of these restrictions can be better seen after expressing \( \sigma_{jt}^{(1)}(U) \) using a series multinomial logit.

We can now rewrite the \( G \) group market shares equations about \( G \) unknown CCP functions at the beginning, eq. (7), as the following \( G \) equations about one unknown \( \sigma_{jt}^{(1)}(U) \) by applying our shifting observation \( \sigma_{jt}^{(g)}(U) = \sigma_{jt}^{(1)}(U + \tau^{(g)}/\omega) \). Recall that we defined \( \tau^{(1)} = 0 \) for convenience.

---

\(^{17}\)This is most transparent by considering a two period dynamic model. In period 2 (terminal period), \( v_{i,j,t=0,t=2} = 0 \). The expected optimal payoff in period 2 is \( \ln(1 + \sum_j \exp(v_{i,j,2})) \). Then the payoff of the outside option in period 1 is \( \beta \mathbb{E}[\ln(1 + \sum_j \exp(v_{i,j})) | X_1, P_1, \xi_1] \). The statement can be generalized to infinite horizon dynamic programming problem easily.
We have

\[ S_{jt}^{(1)} = \int \sigma_{jt}^{(1)}(u + \frac{\tau^{(1)}}{\omega}) \, dF_{t}^{(1)}(u) \]

\[ \vdots \]

\[ S_{jt}^{(G)} = \int \sigma_{jt}^{(1)}(u + \frac{\tau^{(G)}}{\omega}) \, dF_{t}^{(G)}(u) \]

For each period, we now have \( J \times G \) equations, but only \( J + G \) unknowns, where \( J \) refers to the unknown CCP \( \sigma_{jt}^{(1)}(U), \ldots, \sigma_{jt}^{(1)}(U) \) for group 1, and \( G \) comes from the unknown \( \tau^{(2)}, \ldots, \tau^{(G)} \) and \( \omega \), which are common for all markets and products. It is hardly a surprise that we can recover \( \sigma_{jt}^{(1)}(U) \) (hence the other \( \sigma_{jt}^{(2)}(U), \ldots, \sigma_{jt}^{(G)}(U) \) by shifting) from the above equations by parameterizing \( \sigma_{jt}^{(1)}(U) \) (so it is known up to finite number of parameters).

5.2 Series Multinomial Logit Approximation of CCP

We now flesh out the details of solving CCP \( \sigma_{jt}^{(1)}(U) \) from eq. (8). The unknown \( \sigma_{jt}^{(1)}(U) \) is a continuous function of scalar \( U \), whose range is between 0 and 1, and \( \sum_{j=0}^{J} \sigma_{jt}^{(1)}(U) = 1 \) for any \( U \).

We approximate the CCP \( \sigma_{jt}^{(1)}(U) \) (as a function of \( U \)) by interpolation using “series multinomial logit”, which is a simple extension of the series logit in Hirano, Imbens and Ridder (2003),

\[ \sigma_{jt}^{(1)}(U; \rho_t) \equiv L_j(R_K(U; \rho_{1t}), \ldots, R_K(U; \rho_{Jt})), \]
where $L_j$ is a multinomial logit model,
\[
L_j(c_1, \ldots, c_J) \equiv \frac{\exp(c_j)}{1 + \sum_{k=1}^{J} \exp(c_k)},
\]
and $R_K(U; \rho_{jt})$ is a polynomial function,
\[
R_K(U; \rho_{jt}) \equiv \rho_{jt1} + \rho_{jt2}U + \rho_{jt3}U^2 + \cdots + \rho_{jtK}U^{K-1}.
\]
Let $\rho_{jt} = (\rho_{jt1}, \ldots, \rho_{jtK})'$, and let $\rho_t$ be the collection of $\rho_{1t}, \ldots, \rho_{Jt}$. Lastly,
\[
\sigma_{0t}^{(1)}(U; \rho_t) \equiv 1 - \sum_{j=1}^{J} \sigma_{jt}^{(1)}(U; \rho_t).
\]
The idea of series (multinomial) logit is to use the power series $R_K(U; \rho_{jt})$ to approximate the log odds ratio $\ln[\sigma_{jt}^{(1)}(U)/\sigma_{0t}^{(1)}(U)]$. In practice, we found the polynomial of degree 2, i.e. $K = 3$, is sufficient for approximation. For exposition simplicity, we let $K = 3$ hereafter, and let $\rho$ be the $(KJT) \times 1$ vector from stacking $\rho_t$ over all $T$ periods. The coefficients $\rho$ in this series expansion indeed will have economic interpretation—when $\beta = 0$, $\rho_{jt1}$ is the mean value of product $j$ among the consumers from group 1 as defined in Berry (1994) (see Remark C.1 in the appendix), otherwise the term includes both the mean value and proportions of the continuation value.

The workflow of our CCP estimation is the following. We first estimate $(\tau, \omega, \rho)$ using an NLS procedure below. Knowing $\rho$, we know group 1 CCP $\sigma_{jt}^{(1)}(U; \rho_t)$ for each alternative $j$ and each period $t$. Knowing $\tau$ and $\omega$, we know the CCP of the other groups by shifting: $\sigma_{jt}^{(g)}(U) = \sigma_{jt}^{(1)}(U + \tau^{(g)}/\omega)$.

For the rest of this section, we will explain the estimation of $(\tau, \omega, \rho)$ leaving some remarks about the limitation of our approach and asymptotic distribution at the end. The estimation is based on
\[
S_{jt}^{(g)} = \int \sigma_{jt}^{(1)}\left( u + \frac{\tau^{(g)}}{\omega}; \rho_t \right) f_t^{(g)}(u) \, du,
\]
where
\[
\sigma_{jt}^{(1)}\left( U + \frac{\tau^{(g)}}{\omega}; \rho_t \right) = \frac{\exp\left[ \rho_{jt1} + \rho_{jt2}\left( U + \frac{\tau^{(g)}}{\omega} \right) + \rho_{jt3}\left( U + \frac{\tau^{(g)}}{\omega} \right)^2 \right]}{1 + \sum_{k=1}^{J} \exp\left[ \rho_{kt1} + \rho_{kt2}\left( U + \frac{\tau^{(g)}}{\omega} \right) + \rho_{kt3}\left( U + \frac{\tau^{(g)}}{\omega} \right)^2 \right]},
\]
Proposition 1 says $f_t^{(g)}(U) = \Gamma_t^{(g)}(U)\phi(U)$, where $\phi(U)$ is the PDF of the standard normal distribution. The above equation can be rewritten as follows,
\[
S_{jt}^{(g)} = \mathbb{E}\left[ \sigma_{jt}^{(1)}\left( U^* + \frac{\tau^{(g)}}{\omega}; \rho_t \right) \Gamma_t^{(g)}(U^*) \right], \quad U^* \sim \mathcal{N}(0, 1).
\]
for all $j = 1, \ldots, J$, $g = 1, \ldots, G$, and $t = 1, \ldots, T$. In practice, it is straightforward to compute the above expectation by Gauss–Hermite quadrature:

$$GH_{jt}^{(g)}(\tau, \omega, \rho) \equiv \sum_{i=1}^{n} \eta_i \times \left[ \sigma_{jt}^{(1)} \left( u_i + \frac{\tau(g)}{\omega} \right) \right] I_{jt}^{(g)}(u_i).$$

Here $u_1, \ldots, u_n$ are $n$ nodes, and $\eta_1, \ldots, \eta_n$ are the respective weights. Both nodes and weights are predetermined known constants. By Gauss–Hermite approximation, we have

$$S_{jt}^{(g)} = GH_{jt}^{(g)}(\tau, \omega, \rho). \quad (10)$$

We then can estimate the unknown parameters $(\tau, \omega, \rho)$ by NLS:

$$\left( \hat{\tau}, \hat{\omega}, \hat{\rho} \right) \equiv \arg \min_{\tau, \omega, \rho} \sum_{j=1, g=1, t=1}^{J, G, T} \left[ S_{jt}^{(g)} - GH_{jt}^{(g)}(\tau, \omega, \rho) \right]^2 \quad (11)$$

subject to

$$\rho_{jt2} - \rho_{1t2} = -\omega(P_{jt} - P_{1t}) \quad \text{and} \quad \rho_{jt3} - \rho_{1t3} = 0, \quad j = 2, \ldots, J. \quad (12)$$

In Appendix C we provide the derivation of the constraints eq. (12) implied by our structural demand model. From eq. (12), we have seen the interpretation of the parameters of series expansion, $\rho_{jt2}$ and $\rho_{jt3}$, in terms of the structural parameters. The next proposition, whose proof is also in Appendix C provides the interpretation of $\rho_{jt1}$ using the structural parameters.

**Proposition 2** (Interpretation of series logit parameters $\rho_{jt1}$). Recall that

$$Y_{jt} \equiv \int \ln \left[ \frac{\sigma_{jt}^{(1)}(U)}{\sigma_{1t}^{(1)}(U)} \right] dF_t^{(1)}(U) + \omega(P_{jt} - P_{1t}) \int U dF_t^{(1)}(U)$$

is the dependent variable of our first identification linear regression eq. (Linear-Reg-1). We have

$$Y_{jt} = \rho_{jt1} - \rho_{1t1}. \quad \text{Remark 4 (Overlapping Groups).} \quad \text{One key step is the transformation between } \sigma_{jt}^{(1)}(U) \text{ and } \sigma_{1t}^{(2)}(U). \quad \text{Our observation is that for a consumer } \ell \text{ from observed group } 2 \text{ with idiosyncratic } U_\ell, \text{ we can find a consumer } h \text{ from group } 1, \text{ whose idiosyncratic } U_h = U_\ell + \tau^{(2)}/\omega, \text{ then the two consumers have the same price coefficient. See Panel I of Figure 4.}$$

---

18To be clear, let $u_1^*, \ldots, u_n^*$ be the $n$ nodes of Gauss-Hermite quadrature, and let $\eta_1^*, \ldots, \eta_n^*$ be the respective weights. In our simulation, we used $n = 15$ nodes. The nodes and associated weights are determined by the Hermite polynomial, and they do not depend on the function to be approximated, which is $\sigma_{jt}^{(1)}(U + \tau^{(2)}/\omega, \rho_t)$ herein. For $i = 1, \ldots, n$, define $u_i = \sqrt{2}u_i^*$ and $\eta_i = \eta_i^*/\sqrt{\pi}.$
$h \in \text{Group 1}, \alpha_h = \alpha^{(1)} + \omega U_h$, matches

$\ell \in \text{Group 2}, \alpha_\ell = \alpha^{(1)} + \tau^{(2)} + \omega U_\ell$ when

$U_h = U_\ell + \tau^{(2)}/\omega$.

Figure 4: PDF of Price Coefficients from Different Groups
However, when the two groups 1 and 3 are too distinct (that is \( \tau(3) \) is large) and/or the within group variation of idiosyncratic \( U \) is extremely small (that is \( \omega \) is tiny), the chance of cross-group matching decreases. See Panel II of Figure 4. Note that \( \tau(3)/\omega \) becomes large, when either \( \tau(3) \) is large or \( \omega \) is small. In this case, though the identity \( \sigma_{jt}^{(3)}(U) = \sigma_{jt}^{(1)}(U + \tau(3)/\omega) \) is still valid, the equation about the market share in group 3,

\[
S_{jt}^{(3)} = \int \sigma_{j}^{(1)}(u + \tau(3)/\omega) \, dF_{t}^{(3)}(u)
\]

has less information for identifying \( \sigma_{jt}^{(1)}(U) \). Empirically, this also implies that the observed \( S_{jt}^{(3)} \) is also close to zero. Intuitively, when the two groups are very different (\( \tau(3) \) is large) or there is little variation (\( \omega \) is small) within a group, it is expected that in equilibrium, the two groups of consumers will choose different products. Our simulation studies (Appendix D) shows that our estimator works very well even when the within group variation is very small (which has the same effect as the inter group variation is large).

**Remark 5** (Asymptotic distribution). The asymptotic distribution of our estimators can be analyzed using the standard (two step) M-estimator framework. The details are provided in Appendix B. Let \( \theta \equiv (\theta_1', \theta_2')' \), where \( \theta_1 \equiv (\tau', \omega, \rho', \delta_2 - \delta_1)/(1 - \beta), \ldots, (\delta_J - \delta_1)/(1 - \beta), \gamma/(1 - \beta), \alpha^{(1)} \)' and \( \theta_2 \equiv (\beta, \delta_1)' \). We can estimate \( \theta_1 \) jointly by combining eq. (11) and eq. (Linear-Reg-1). This step can be organized as a standard M-estimation problem, hence the asymptotic distribution is obtained easily. The estimation of \( \theta_2 \equiv (\beta, \delta_1)' \) uses eq. (Linear-Reg-2) and depends on the first-stage estimates of \( \theta_1 \). Equation (Linear-Reg-2) can be organized as an ordinary least square problem. Thus, we have closed form for the estimators of \((\beta, \delta_1)\). With closed form, the asymptotic distribution of the estimators of \( \theta_2 \) is obtained easily.

6 Model Estimation with Golf Data

We follow the above estimation routine to recover the the model parameters, specifically the unobserved price heterogeneity parameter \( \omega \) and to account for endogeneity concerns.

---

**Endogeneity**

In our estimation routine, we implicitly assumed that the variance of the utility shocks \( \varepsilon_{ijt} \) does not vary across groups. Compared with consumers from the public group, golfers who belong to a private golf course and purchase their golf equipment at on-course golf shops have the opportunity of trying the golf clubs before purchase on the driving range. It is reasonable to expect that such a “try-out” opportunity could affect a consumer’s utility primarily by reducing the variance of utility shocks \( \varepsilon_{ijt} \). The objective here is to understand how such a “try-out” effect will affect the estimation procedure. Briefly speaking, the effect is negligible. Let \( \varepsilon_{jt}^{(Private)}(U) + \kappa^{-1}\varepsilon_{jt}^{(Private)} \) denote the
may be an issue given that firms may set prices and advertising spend based upon the value of the unobservable product characteristic. To account for such, we instrument for price and brand advertising spend with the number of products in the marketplace in period $t$, the number of products produced by a given brand in period $t$, the number of products produced by rival brands in period $t$, and several indicator variables that capture whether time period $t$ occurs before or after Tiger Woods’ infidelity scandal (November 2009) to account for any impact the scandal may have on prices or advertising. Figure 5 illustrates the timeline of Tiger Woods’ infidelity in our sampled months. We include an indicator variable that captures the period after the scandal but before his returns to golf for the 2010 Masters in April (“Scandal Period 1” in Figure 5) and a second indicator that captures the effect of the scandal after he has returned to playing in professional golf tournaments for the time period after March 2010 (“Scandal Period 2” in Figure 5).

With data occurring before and after November 2009, we also include two additional variable that captures the impact of Tiger Woods’ infidelity scandal on the overall endorsement effect (one for “Scandal Period 1” and the other for “Scandal Period 2” in Figure 5). These variable are identical to the above instruments but are interacted with Tiger Woods’ endorsement measure.

Identifying the prestige from the informative effects requires further assumptions and leveraging the short period of time after the 2009 infidelity scandal. It is important to note that during the short period of time after the scandal occurred Tiger Woods did not play in any of the 15 scheduled PGA Tour events (4 months). During this period, our identifying assumption is that any impact from Tiger Woods’ infidelity scandal on Nike sales must be due to the elimination of the

\[
\sigma_{jt}(Private)(U, \kappa) = \exp(\kappa v_{jt}(Private)(U))/(\sum_{k=0}^{J} \exp(\kappa v_{kt}(Private)(U))).
\]

The question is can we still claim that $\sigma_{jt}(Private)(U, \kappa) \approx 1$ when $\kappa \neq 1$, we will argue below that $\sigma_{jt}(Private)(U, \kappa) - \sigma_{jt}(Private)(U, \kappa = 1)$ is negligible in our empirical study, so that the equation still holds approximately. Below we omit the dependence of $\sigma_{jt}(Private)(U, \kappa)$ on $U$ to save space. Consider the first order Taylor expansion of $\sigma_{jt}(Private)(\kappa)$ at $\kappa = 1$, we have that $\sigma_{jt}(Private)(\kappa) - \sigma_{jt}(Private)(1) \approx (\kappa - 1) \sum_{k \neq j} \sigma_{jt}(Private)(1)\sigma_{kt}(Private)(1)(v_{jt}(Private) - v_{kt}(Private))$. Through this first order expansion, we illustrate that we can in principle modify our method to estimate $\kappa$ directly. However, it is unnecessary for our purpose of estimating the celebrity effect. Note that $\sigma_{jt}(Private)(1)$ is approximately the private group’s observed market share of product $j$, which is very small in our real data, and $v_{jt}(Private) - v_{kt}(Private)$ is approximately the log shares ratio between $j$ and $k$. With the mean and maximum of the observed market share in our data being 7.04e−4 and 0.0149, respectively, the individual term $\sigma_{jt}(Private)(1)\sigma_{kt}(Private)(1)$ is negligible. Hence, the difference $\sigma_{jt}(Private)(\kappa) - \sigma_{jt}(Private)(1)$ is also negligible.
Note: Tiger Woods did not play in any of the PGA Tour events in “Scandal Period 1” (November 2009–March 2010), but he was ranked 1 by large margin.

Figure 5: Timeline of Tiger Woods’ Infidelity Scandal in Our Data

(social) prestige component of the endorsement effect and not the informative component that is communicated to consumers by him still being the number one player, having a very large world golf point total, and playing a Nike Driver.

Additionally, we assume the informative component did not decay over time during the short period Tiger Woods did not play competitive golf. Support for such an assumption is straightforward, the Thanksgiving weekend of 2009 was Tiger’s 232 consecutive week at world number one of his incredible 281 week streak (June 12, 2005 to Oct. 30, 2010). No other player before this period or after has come remotely close to amassing so many consecutive weeks at number 1.

Finally, we assume strict exogeneity of the world golf points. First, the endorsement variable measures a player’s quality through the golfer’s world golf points, which again is based on a rolling two-year performance—a difficult measure to control given points are awarded each tournament based on the strength of the field and a player’s relative finish. Secondly, the possibility that a player’s world golf points is correlated with an unobservable and time varying structural error, we believe is unlikely given the inclusion of product fixed effects and from the fact that no listed player switched brands during our specified data period.

6.1 Empirical Results

We now move to discuss our empirical results, which are presented in Table 3. This table presents results using our above method with and without heterogeneity in order to ascertain the importance of incorporating unobserved consumer price heterogeneity. In our discussion below, we discuss the non- endorsement related variables first and then move to presenting the estimated endorsements effects for Tiger Woods. We continue with addressing the overall makeup of his endorsement effect
by discussing the proportion driven by (social) prestige and information.

6.1.1 Price Estimates

The three most important parameter estimates connected to our new estimation methodology are those link to price, which recovers the price sensitivity (α\(^{(Public)}\), τ\(^{(Private)}\)) for each consumer type as well as \(\omega\) which again measures the unobserved consumer price heterogeneity within each group. Our estimates indicate the price sensitivity for those who play golf at public courses is 0.0103. For those who belong to a private golf club, their sensitivity is found by adding \(\alpha^{(Public)}\) and \(\tau^{(Private)}\). This illustrates that those who play private courses are less price sensitive than those who play public courses (0.0013 = 0.0103 – 0.0090). Additionally, there is a statistically significant estimate of within group unobserved heterogeneity (\(\omega = 0.0028\)).

We illustrate the price estimates in an economically meaningful way by providing own-price elasticity measures for each group for a select number of products. To highlight the impact of unobserved heterogeneity plays in formulating price elasticities, we also report results where \(\omega = 0\).

Our elasticity measures consider a temporary price change, such that the valuation of outside option, \(v^{(g)}(U)\), is not affected. Below we present the own and cross price elasticity formulas. The market share of group \(g = \{\text{Public, Private}\}\) in period \(t\) is

\[
S^{(g)}_{jt} = \int \sigma^{(g)}_{jt}(u) f^{(g)}_t(u) \, du = \int \sigma^{(g)}_{jt}(u) \Gamma^{(g)}_t(u) \, d\Phi(u),
\]

Recall that \(f^{(g)}_t(u)\) is the PDF of the unobserved price sensitivity in period \(t\) and group \(g\), and \(f^{(g)}_t(u) = \phi(u) \Gamma^{(g)}_t(u)\) by Proposition 1. The individual CCP is

\[
\sigma^{(g)}_{jt}(U) = \frac{\exp(v^{(g)}_{jt}(U))}{\exp(v^{(g)}_{0t}(U)) + \sum_{k=1}^{J} \exp(v^{(g)}_{kt}(U))},
\]

and

\[
\Gamma^{(g)}_1(u) = 1, \quad \Gamma^{(g)}_t(u) = \prod_{s=1}^{t-1} \frac{\sigma^{(g)}_{js}(u)}{S^{(g)}_{0s}}, \quad t \geq 2.
\]

Because \(\Gamma^{(g)}_t(u)\) does not depend on price \(P_{kt}\) in period \(t\),

\[
\frac{\partial[\sigma^{(g)}_{jt}(u) \Gamma^{(g)}_t(u)]}{\partial P_{kt}} = \Gamma^{(g)}_t(u) \frac{\partial \sigma^{(g)}_{jt}(u)}{\partial P_{kt}}.
\]

Given this, price elasticities, including unobserved consumer price heterogeneity, are

\[
\epsilon^{(g)}_{jk,t} = \frac{\partial S^{(g)}_{jt}/\partial P_{kt}}{S^{(g)}_{jt}/P_{kt}} = \begin{cases} \frac{-P_{jt}}{S^{(g)}_{jt}} \int (\alpha^{(1)} + \tau^{(g)} + \omega u) \sigma^{(g)}_{jt}(u)(1 - \sigma^{(g)}_{jt}(u)) \Gamma^{(g)}_t(u) \, d\Phi(u) & \text{if } j = k \\ \frac{P_{kt}}{S^{(g)}_{jt}} \int (\alpha^{(1)} + \tau^{(g)} + \omega u) \sigma^{(g)}_{jt}(u) \sigma^{(g)}_{kt}(u) \Gamma^{(g)}_t(u) \, d\Phi(u) & \text{if } j \neq k. \end{cases}
\]
Table 3: Estimation Results

<table>
<thead>
<tr>
<th>Variables</th>
<th>With Hetero.</th>
<th>W/O Hetero.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Adam Scott/(1 − β)</td>
<td>−0.7755 (0.7587)</td>
<td>−0.7909 (0.7900)</td>
</tr>
<tr>
<td>Bubba Watson/(1 − β)</td>
<td>−1.8704*** (0.3357)</td>
<td>−1.9230*** (0.3495)</td>
</tr>
<tr>
<td>Dustin Johnson/(1 − β)</td>
<td>0.7249*** (0.2660)</td>
<td>0.9225*** (0.2769)</td>
</tr>
<tr>
<td>Jason Day/(1 − β)</td>
<td>2.8122*** (0.5981)</td>
<td>2.8366*** (0.6227)</td>
</tr>
<tr>
<td>Justin Rose/(1 − β)</td>
<td>−0.1967 (0.4692)</td>
<td>−0.2001 (0.4885)</td>
</tr>
<tr>
<td>KJ Choi/(1 − β)</td>
<td>2.9750*** (0.8447)</td>
<td>3.2072*** (0.8794)</td>
</tr>
<tr>
<td>Lee Westwood/(1 − β)</td>
<td>0.9880** (0.5158)</td>
<td>0.8848 (0.5371)</td>
</tr>
<tr>
<td>Luke Donald/(1 − β)</td>
<td>−1.3784 (1.1658)</td>
<td>−1.3650 (1.2138)</td>
</tr>
<tr>
<td>Phil Mickelson/(1 − β)</td>
<td>1.8163 (1.8881)</td>
<td>2.0649 (1.9658)</td>
</tr>
<tr>
<td>Rory McIlroy/(1 − β)</td>
<td>6.0277*** (0.7630)</td>
<td>6.4976*** (0.7944)</td>
</tr>
<tr>
<td>Sergio Garcia/(1 − β)</td>
<td>3.3361*** (0.6423)</td>
<td>3.5164*** (0.6687)</td>
</tr>
<tr>
<td>Tiger Woods/(1 − β)</td>
<td>8.4475*** (0.7320)</td>
<td>8.5409*** (0.7621)</td>
</tr>
<tr>
<td>Tiger × Scandal (11/09–3/10)/(1 − β)</td>
<td>−0.7344*** (0.2009)</td>
<td>−0.7500*** (0.2092)</td>
</tr>
<tr>
<td>Tiger × Scandal (Post 3/10)/(1 − β)</td>
<td>−0.8499*** (0.2430)</td>
<td>−1.0041*** (0.2530)</td>
</tr>
<tr>
<td>Log of Ad Spend($)/(1 − β)</td>
<td>0.2557*** (0.0252)</td>
<td>0.2592*** (0.0262)</td>
</tr>
<tr>
<td>Product Age (Month)/(1 − β)</td>
<td>−0.0476*** (0.0220)</td>
<td>−0.0568*** (0.0229)</td>
</tr>
<tr>
<td>Price (α(Public))</td>
<td>0.0103*** (0.0028)</td>
<td>0.0094*** (0.0029)</td>
</tr>
<tr>
<td>Price (τ(Private))</td>
<td>−0.0090*** (0.0046)</td>
<td>−0.0092 (0.0060)</td>
</tr>
<tr>
<td>Within Group Deviation (ω)</td>
<td>0.0028** (0.0017)</td>
<td>— —</td>
</tr>
<tr>
<td>Discount Factor (β)</td>
<td>0.9002*** (0.4126)</td>
<td>0.9223 (1.3128)</td>
</tr>
</tbody>
</table>

Note: Product fixed effects not reported due to large number of products;
*** 95 percent significance; ** 90 percent significance
With having estimated \( \sigma_{jt}^{(g)}(u) \) and \( \Gamma_{t}^{(g)}(u) \) and using Gauss–Hermite quadrature, we have the following formula for own-price elasticity,

\[
e_{jt}^{(g)} = -\frac{P_{jt}}{s_{jt}^{(g)}} \sum_{i=1}^{n} (\alpha^{(1)} + \tau^{(g)} + \omega u_i) \sigma_{jt}^{(1)} \left( u_i + \frac{\tau^{(g)}}{\omega}; \rho_t \right) \left[ 1 - \sigma_{jt}^{(1)} \left( u_i + \frac{\tau^{(g)}}{\omega}; \rho_t \right) \right] \Gamma_{t}^{(g)}(u_i) \times \eta_i.
\]

Recall that \( u_1, \ldots, u_n \) are \( n \) known nodes, and \( \eta_1, \ldots, \eta_n \) are the respective known weights. Similarly, we have the cross-price elasticities,

\[
e_{jk,t}^{(g)} = \frac{P_{kt}}{s_{jt}^{(g)}} \sum_{i=1}^{n} (\alpha^{(1)} + \tau^{(g)} + \omega u_i) \sigma_{jt}^{(1)} \left( u_i + \frac{\tau^{(g)}}{\omega}; \rho_t \right) \sigma_{kt}^{(1)} \left( u_i + \frac{\tau^{(g)}}{\omega}; \rho_t \right) \Gamma_{t}^{(g)}(u_i) \times \eta_i.
\]

Before presenting the estimates, we want to point out the difference between the price elasticities in a static and dynamic model. In the absence of attrition, i.e. \( \Gamma_{t}^{(g)}(u) = 1 \), the above price elasticities \( e_{jt}^{(g)} \) are identical to the elasticities in a static BLP model \cite{Nevo2011}. This is because we consider a temporary price change. With attrition of consumers, however, the prices elasticities in our dynamic model are different even with temporary price change, due to the previous change of the pool of remaining consumers captured by \( \Gamma_{t}^{(g)} \). Consumers attrition is closely related to the product life cycle. So our results are useful for designing the price promotion, which is usually temporary, of a product at different stages of its life.

Table 4 presents estimates of the own-price elasticities for the most recent product release for each “major” brand across consumer groups for each of the above estimation results. In parentheses next to the product name we present the month and year the product was launched. The estimates indicate that the consumers who play their golf at private golf clubs are less price sensitive than those consumers who play at public courses when purchasing a new driver. Additionally, when incorporating unobserved consumer heterogeneity in estimation, own-price elasticities increase (more noticeably for golfers in private clubs), indicating consumers are more sensitive and the importance of capturing unobserved consumer heterogeneity.

\footnote{So far we have only computed the contemporary price elasticities—the effect of price change on the current market share. By attrition of consumers, even temporary price change could in principle change the distribution of consumers in the future. Consider \( \frac{\partial s_{jt+1}^{(g)}}{\partial P_{kt}} \) (the effect of temporary price change of product \( k \) in period \( t \) on the market share of product \( j \) in period \( t + 1 \)) for example. This elasticity mainly depends on \( \frac{\partial s_{jt}^{(g)}}{\partial P_{kt}} \) and \( \frac{\partial s_{jt+1}^{(g)}}{\partial P_{kt}} \), which are close to zero in practice because small temporary price change hardly change the market share of the outside option. For this reason, we believe the elasticities like \( \frac{\partial s_{jt+1}^{(g)}}{\partial P_{kt}} \) are close to zero, and only focus on the contemporary price elasticities.}
Table 4: Mean Own-Price Elasticity (%) Estimates

<table>
<thead>
<tr>
<th>Product</th>
<th>With Hetero.</th>
<th>W/O Hetero. ($\omega = 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public</td>
<td>Private</td>
</tr>
<tr>
<td>Nike VR Pro (March 2011)</td>
<td>-3.7926</td>
<td>-0.4058</td>
</tr>
<tr>
<td>Callaway Razr Hawk (March 2011)</td>
<td>-3.7273</td>
<td>-0.4165</td>
</tr>
<tr>
<td>Taylormade R11 (March 2011)</td>
<td>-3.6949</td>
<td>-0.5579</td>
</tr>
<tr>
<td>Ping K15 (August 2010)</td>
<td>-2.6392</td>
<td>-0.4783</td>
</tr>
<tr>
<td>Cobra S3 (March 2011)</td>
<td>-3.1998</td>
<td>-0.2627</td>
</tr>
</tbody>
</table>

6.1.2 Tiger Woods Endorsement Estimates

We now turn our attention to Tiger Woods’ endorsement effect and the proportion of such due to (social) prestige and information. Our estimates determine that Tiger Woods’ endorsement had a large impact on changing consumer behavior relative to other listed players. The marginal (product) lifetime utility associated with his overall endorsement effect ($\gamma_1 - \beta$) is 8.4475. The parameter of interest that captures the lifetime prestige effect is the world golf ranking points interacted with an indicator variable for the period between November 2009 and March 2010. Estimation indicates the marginal lifetime (social) prestige effect to be 0.7334. Given the marginal utility estimate of the combined lifetime prestige and informational effect as 8.4475, we determine that the social prestige consists of roughly 9 percent of this effect. Given such a finding, we find that the former CEO of Titleist was correct in his reason to have Tiger Woods endorse Titleist products—the benefit from Tiger Woods’ endorsement largely consists of a transfer of information to consumers about the quality of the product and less so by a consumer’s drive to associate him/herself with Tiger Woods by playing the same brand of driver. That said, this latter result is not inconsequential. This is the first paper to empirically illustrate that consumers do adopt products based upon (social) prestige. In a related paper regarding the informative effects of advertising, Ackerberg (2003), using data from the yogurt market, determines that consumer behavior is driven by a large informative effect for advertising (and a statistically insignificant prestige effect).
6.1.3 Other Estimates

We wrap-up the empirical discussion with highlighting the role product age, advertising, and the discount factor play in a consumer’s decision. Importantly, age and advertising are both found to be statistically significant and hold the proper sign. Age is found to negatively affect a product’s utility whereas advertisement increases the likelihood a consumer purchases a given brand. Finally, it is a must that we discuss our discount factor estimate. The first note of interest is that we are able to precisely estimate it (student t = 2.1818). Additionally, the magnitude is quite different from the standard assumption of either $\beta = 0.99$ or $\beta = 0.975$. For instance, a monthly discount factor set at $\beta = 0.975$ assumes consumers value future utility roughly 21 years in the future ($\beta^{12\times21} = 0.0017$). However, our estimate of $\beta = 0.90$ equates to 5 years—by the 60th month after the purchase date the associated flow utility is valued at near zero ($\beta^{60} = 0.0018$). How reasonable is this estimate? It is quite sensible according to the leading industry magazine, *Golf Digest*, which recommends golfers not own any equipment older than 5 years to capture the benefit that newer technology can bring in the form of better accuracy and more distance.\footnote{Jerry Tarde, “New Versus Old Stuff”, Golf Digest, February 08, 2018, \url{https://www.golfdigest.com/story/editors-letter-new-versus-old-stuff}}

7 Conclusion

In estimating dynamic discrete choice demand models for durable goods, it is essential to account for unobserved consumer heterogeneity and unobserved product characteristics in order to obtain unbiased estimates of important parameters like price elasticities. However, in the implementation of such models, it is inevitable to address the curse of dimensionality caused by the large number of products on the market and the high dimension of product characteristics. This paper provides a new estimation approach using group market share data that includes continuous unobserved consumer heterogeneity and unobserved product characteristics, but avoids the curse of dimensionality. As a result, our methodology can be used in the markets with many differentiated products.

Particularly, using our method and the exogenous shock from Tiger Woods’ infidelity scandal in November 2009, we separately identify the informational and prestige effect of celebrity endorsement where we find a sizable informative effect for Tiger Wood’s endorsement.

The implementation of our method is simple and requires only NLS and 2SLS. It allows researchers to consider various model specifications at little computational and programming cost.
In addition to these practical benefits, our method also has a few theoretical appealing properties. We find that the identification of dynamic demand model requires the same conditions as the identification of static demand model, which requires instrumental variables to address the endogeneity of variables like price. In establishing identification, we are agnostic about how consumers form their beliefs regarding the state transition distribution, and we explicitly take account of dynamic selection. This is useful because there is evidence of discrepancies between an individual’s subjective belief and the assumption of rational expectations (see the cited reference in An, Hu and Xiao 2020). The fact that the model is identified without imposing an assumption about consumer beliefs is novel and interesting, and is particularly true for the discount factor where existing results require not only rational expectation but also excluded variables (that affect the state transition but not the flow utility). Our paper requires neither.

Our data requirement is group market shares (or sales). This is weaker than requiring consumer-level panel data, from which we can construct the group market shares. To collect customer level panel data, companies need to track customers over time. For durable goods, this could be very costly and impractical due to their longer life span (last at least 3 years by the definition of the US Census Bureau). Additionally, in the digital age where privacy is concern, customers are unwilling to be tracked over time, and companies are unwilling to share their customer-level data with researchers. As a result, these prospects make our data requirement (and methodology) more desirable in the future.

---

References


APPENDIX

A Proofs

Proof of Proposition 1. First, without attrition, the pool of consumers does not change with time. So that $f_t^{(g)}(u) = f_1^{(g)}(u) = \phi(u)$ for any period $t$. In the rest, we focus on the case with attrition.

When $t = 1$, $f_1^{(g)}(u) = \phi(u)$ by Assumption 4. We will prove

$$f_t^{(g)}(u) = \phi(u) \times \prod_{s=1}^{t-1} \frac{\sigma_{u(s)}^{(g)}(u)}{S_{0(s)}^{(g)}}, \quad t \geq 2. \tag{A.1}$$

by the induction. Define a few notation for exposition. Let $A_{it}$ denote the discrete purchasing choice made by consumer $i$ in period $t$. Particularly, $A_{it} = 0$ means not purchase in period $t$. Let $Z_t \equiv (X_t', P_t', \xi_t')'$ denote the vector of product characteristics in period $t$. Also recall that $D_i^{(g)} = 1$ denotes that $i$ is from group $g$.

When we randomly draw a consumer $i$ with unobserved price sensitivity $U_i$ from group $g$, $f_t^{(g)}(u)$ is the probability that $U_i = u$ provided that consumer $i$ still exists in period $t$. Because consumers leave the market after purchasing, a consumer would exist in period $t$ if and only if she had chosen not to purchase in all the previous periods given the past product characteristics. In other words, $f_t^{(g)}(u)$ is the probability that $U_i = u$ conditional on that $D_i^{(g)} = 1$ (so $i$ is from group $g$) and consumer $i$ did not purchase from period 1 to $t - 1$ with the past product characteristics $Z_1, \ldots, Z_{t-1}$. That is

$$f_i^{(g)}(u) = \Pr(U_i = u \mid D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_{t-1}).$$

The above conditioning variables just restrict the population to be the remaining consumers after $t - 1$ periods. Because all the remaining consumers in period $t$ face the same the product state variables $Z_t$, we also have the following conditional independence,

$$f_t^{(g)}(u) = \Pr(U_i = u \mid D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_{t-1})$$

$$= \Pr(U_i = u \mid D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_{t-1}, Z_t). \tag{A.2}$$

We now prove eq. (A.1) by the induction. Starting with period 2, we have

$$f_2^{(g)}(u) = \Pr(U_i = u \mid D_i^{(g)} = 1, A_{i1} = 0, Z_1)$$

$$= \frac{f(U_i = u \mid D_i^{(g)} = 1, Z_1) \cdot f(A_{i1} = 0 \mid U_i = u, D_i^{(g)} = 1, Z_1)}{f(A_{i1} = 0 \mid D_i^{(g)} = 1, Z_1)}$$

$$= \phi(u) \times \frac{\sigma_{01}^{(g)}(u)}{S_{01}^{(g)}}.$$
Suppose eq. (A.1) holds for period $t$. We will prove that this equation also holds for period $t + 1$. We have

$$f_{t+1}^{(g)}(u) = \Pr(U_i = u \mid D_t^{(g)} = 1, A_{i,0} = 0, \ldots, A_{i,t-1} = 0, A_{it} = 0, Z_1, \ldots, Z_t)$$

$$= \Pr(U_i = u \mid D_t^{(g)} = 1, A_{i,0} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t)$$

$$\times \frac{f(A_{it} = 0 \mid U_i = u, D_t^{(g)} = 1, A_{i,1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t)}{f(A_{it} = 0 \mid D_t^{(g)} = 1, A_{i,1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t)}$$

$$= f_t^{(g)}(u) \times \frac{f(A_{it} = 0 \mid U_i = u, D_t^{(g)} = 1, A_{i,1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t)}{f(A_{it} = 0 \mid D_t^{(g)} = 1, A_{i,1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t)}$$

by eq. (A.2)

$$= f_t^{(g)}(u) \sigma_0^{(g)}(u) \sigma_0^{(g)}(u) = \phi(u) \times \prod_{s=1}^{t} \frac{S_{0s}^{(g)}(u)}{S_{0s}^{(g)}}.$$  

Note that the purchase choice $A_{it}$ in period $t$ depends only on the payoffs $v_{ijt}$, which are functions of $(U_i, D_t^{(g)}, Z_t)$ only. So that $A_{it} \perp (A_{i,1}, \ldots, A_{it-1}, Z_1, \ldots, Z_t-1) \mid (U_i, D_t^{(g)}, Z_t)$, and the last line follows. This completes the proof.

\[\] 

**Proposition A.1** (Group composition due to attrition). Suppose consumers leave the market after purchasing. Let $\pi_t^{(g)}$ denote the proportion of group $g$ consumers in period $t$, and let $S_{0t}$ denote the share of consumers who choose the outside option (not purchase) in period $t$. We have

$$\pi_t^{(g)} = \pi_1^{(g)} \times \prod_{s=1}^{t-1} \frac{S_{0s}^{(g)}}{S_{0s}}. \tag{A.3}$$

Proof. The proof is similar to the proof of proposition [1] and we keep using the notation defined in that proof. We prove by induction. Starting from period 2, we have the following by definition,

$$\pi_2^{(g)} = \Pr(D_t^{(g)} = 1 \mid A_{i,1} = 0, Z_1)$$

$$= \frac{\Pr(D_t^{(g)} = 1 \mid Z_1) \Pr(A_{i,1} = 0 \mid D_t^{(g)} = 1, Z_1)}{\Pr(A_{i,1} = 0, Z_1)}$$

$$= \pi_1^{(g)} \times \frac{S_{01}^{(g)}}{S_{01}}.$$  

Suppose eq. (A.3) holds for period $t$. We want to show the statement holds for period $t + 1$, we have

$$\pi_{t+1}^{(g)} = \Pr(D_t^{(g)} = 1 \mid A_{i,1} = 0, \ldots, A_{i,t-1} = 0, A_{it} = 0, Z_1, \ldots, Z_t)$$

$$= \Pr(D_t^{(g)} = 1 \mid A_{i,1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t, Z_t)$$

$$\times \frac{\Pr(A_{it} = 0 \mid D_t^{(g)} = 1, A_{i,1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)}{\Pr(A_{it} = 0 \mid A_{i,1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)}$$

$$= \pi_t^{(g)} \frac{S_{0t}^{(g)}}{S_{0t}}.$$  

The last line follows because the vector product characteristics $Z_t$ is the same for different groups of remaining consumers after $t - 1$ periods, so that $\Pr(D_t^{(g)} = 1 \mid A_{i,1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t, Z_t) = \Pr(D_t^{(g)} = 1 \mid A_{i,1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_{t-1}, Z_t) = \pi_t^{(g)}$. This completes the proof.  

\[\]
B  Asymptotics

The asymptotics is built on $T \to \infty$ with the number of products $J$ and the number of groups $G$ being fixed. Let $\theta \equiv (\theta_1, \theta_2)$, where $\theta_1 \equiv (\tau, \omega, \rho, (\delta_2 - \delta_1)/(1 - \beta), \ldots, (\delta_J - \delta_1)/(1 - \beta), \gamma/(1 - \beta), \alpha^{(1)})$, and $\theta_2 \equiv (\beta, \delta_1)$. We decompose $\theta$ into these two parts, because the estimation of $\theta_2$ relies on the estimation of $\theta_1$. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ denote the estimator of $\theta_1$ and $\theta_2$, respectively.

It is easier to derive the asymptotic distribution backwardly from $\theta_2 \equiv (\beta, \delta_1)$. We use eq. (Linear-Reg-2) to estimate $\theta_2$. In the estimation of $\theta_2$, we fix $U$ at certain number. In particular, we let $U = 0$ here to simplify the discussion. For exposition simplicity, we ignore the dependence on $U$ below. Let

$$W^{(g)}(\hat{\theta}_1) \equiv T^{-1} \sum_{t=1}^{T} W_t^{(g)}(U = 0; \hat{\theta}_1), \quad \text{and} \quad \hat{\ell}^{(g)}(\hat{\theta}_1) \equiv T^{-1} \sum_{t=1}^{T} \ln \sigma^{(g)}_{\theta t}(U = 0; \hat{\theta}_1).$$

Then

$$\hat{\theta}_2 = [A(\hat{\theta}_1)'A(\hat{\theta}_1)]^{-1} A(\hat{\theta}_1)' Y_2(\hat{\theta}_1).$$

where $A(\hat{\theta}_1)$ is $G \times 2$ matrix, and $Y_2(\hat{\theta}_1)$ is $G \times 1$ vector defined below:

$$A(\hat{\theta}_1) = \begin{bmatrix} 1 & \hat{W}^{(1)}(\hat{\theta}_1) + \hat{\ell}^{(1)}(\hat{\theta}_1) \\ \vdots & \vdots \\ 1 & \hat{W}^{(G)}(\hat{\theta}_1) + \hat{\ell}^{(G)}(\hat{\theta}_1) \end{bmatrix} \quad \text{and} \quad Y_2(\hat{\theta}_1) = \begin{bmatrix} \hat{W}^{(1)}(\hat{\theta}_1) \\ \vdots \\ \hat{W}^{(G)}(\hat{\theta}_1) \end{bmatrix}.$$ 

They are defined in this way so that $Y_2$ is the “dependent variable” and $A$ is the “design matrix” of eq. (Linear-Reg-2). Because $\hat{\theta}_2$ is analytical function of random variables $\hat{W}^{(g)}(\hat{\theta}_1)$ and $\hat{\ell}^{(g)}(\hat{\theta}_1)$, its variance can be easily obtained by simulation provided that we know the asymptotic distribution of $\hat{W}^{(g)}(\hat{\theta}_1)$ and $\hat{\ell}^{(g)}(\hat{\theta}_1)$.

Now, we derive the distribution of $\hat{W}^{(g)}(\hat{\theta}_1)$ and $\hat{\ell}^{(g)}(\hat{\theta}_1)$, whose definition requires $W_t^{(g)}(0; \hat{\theta}_1)$ and $\sigma^{(g)}_{\theta t}(0; \hat{\theta}_1)$. We have

$$W_t^{(g)}(0; \hat{\theta}_1) = \ln \left[ \frac{\hat{\sigma}^{(g)}_t (0)}{\sigma^{(g)}_{\theta t} (0)} \right] - X'_t \left( \frac{\gamma}{1 - \beta} \right) + (\hat{\alpha}^{(1)} + \hat{\tau}^{(g)}) P_{1t},$$

$$= \ln \left[ \frac{\hat{\sigma}^{(1)}_t (\hat{\tau}^{(g)}/\omega)}{\hat{\sigma}^{(1)\prime}_t (\hat{\tau}^{(g)}/\omega)} \right] - X'_t \left( \frac{\gamma}{1 - \beta} \right) + (\hat{\alpha}^{(1)} + \hat{\tau}^{(g)}) P_{1t},$$

$$= \hat{\rho}_{1t1} + \hat{\rho}_{1t2} \frac{\hat{\tau}^{(g)}}{\omega} + \hat{\rho}_{1t3} \left( \frac{\hat{\tau}^{(g)}}{\omega} \right)^2 - X'_t \left( \frac{\gamma}{1 - \beta} \right) + (\hat{\alpha}^{(1)} + \hat{\tau}^{(g)}) P_{1t},$$

and

$$\sigma^{(g)}_{\theta t}(0; \hat{\theta}_1) = \sigma^{(1)}_{\theta t} (\hat{\tau}^{(g)}/\omega; \hat{\theta}_1)$$

has the series logit expression. Both $W_t^{(g)}(0; \hat{\theta}_1)$ and $\sigma^{(g)}_{\theta t}(0; \hat{\theta}_1)$ are analytical functions of $\hat{\theta}_1$. We then can determine the variance of $\hat{W}^{(g)}(\hat{\theta}_1)$ and $\hat{\ell}^{(g)}(\hat{\theta}_1)$ by randomly drawing samples from the asymptotic distribution of $\hat{\theta}_1$.

Lastly, we derive the distribution of $\hat{\theta}_1$. We estimate $\theta_1$ by

$$\hat{\theta}_1 \equiv \arg \min_{\theta_1 \in \Theta_1} T^{-1} \sum_{t=1}^{T} h_t(\theta)' h_t(\theta)$$

subject to constraints eq. (C.1) below,
where

\[ h_t(\theta_1) = (h_{1t}(\theta_1), h_{2t}(\theta_1), \ldots, h_{Jt}(\theta_1)), \]

and

\[ h_{jt}(\theta_1) = \begin{bmatrix} S_{jt}^{(1)} - GH_{jt}^{(G)}(\tau, \omega, \rho) \\ \\ \\ S_{jt}^{(G)} - GH_{jt}^{(G)}(\tau, \omega, \rho) \\ X_{jt}^{IV} - \frac{\delta_t - \delta_1}{1 - \beta} - (X_{jt} - X_{1t})' \frac{\gamma_t}{1 - \beta} + \alpha(1)(P_{jt} - P_{1t}) \end{bmatrix}. \]

Here \( X_{jt}^{IV} \) is a vector of IV in eq. (Linear-Reg-1) satisfying \( E[X_{jt}^{IV}(\xi_{jt} - \xi_{1t})] = 0 \). This is a standard M-estimation problem, so under the regularity conditions, \( \sqrt{T}(\hat{\theta}_1 - \theta_1) \to_d N(0, \Sigma_1) \). The asymptotic variance \( \Sigma_1 \) is readily reported by most statistical softwares, or computed using numerical score and Hessian matrices.

## C Constraints about CCP

We claim that the dynamic model implies the following constraints about the parameters \( \rho_i \) in the CCP function:

\[ \rho_{j12} - \rho_{112} = -\omega(P_{jt} - P_{1t}) \quad \text{and} \quad \rho_{j13} - \rho_{113} = 0, \quad j = 2, \ldots, J. \quad (C.1) \]

The above constraints eliminate many parameters, leaving the following to estimate in the NLS problem, eq. (11):

\[ \rho_{j11}, \rho_{j12}, \rho_{j13}, \omega, \tau, \quad \text{for} \quad j = 1, \ldots, J. \]

The degree of freedom of the NLS problem is \( JGT - JT - 2T - G \), where \( JGT \) is the number of observations, \( JT \) results from \( \rho_{j11} \) for each \( j = 1, \ldots, J \) and \( t = 1, \ldots, T \), \( 2T \) comes from \( (\rho_{112}, \rho_{113}) \) for each \( t = 1, \ldots, T \), and \( G \) refers to one \( \omega \) plus \((G - 1) \times 1\) vector \( \tau \). The most stringent case is when \( G = 2 \), in which we need at least three products \( (J \geq 3) \) and \((J - 2)T > G \). In practice, such NLS with the above constraints takes very little time and is robust to the choice of initial guess.

Thus, the CCP estimation stage uses the series logit approximation,

\[ \sigma_{jt}^{(1)}(U + \frac{\tau^{(g)}}{\omega} : \rho_t) = \frac{\exp\left[\rho_{jt1} + \rho_{jt2} \left(U + \frac{\tau^{(g)}}{\omega}\right) + \rho_{jt3} \left(U + \frac{\tau^{(g)}}{\omega}\right)^2\right]}{1 + \sum_{k=1}^{J} \exp\left[\rho_{kt1} + \rho_{kt2} \left(U + \frac{\tau^{(g)}}{\omega}\right) + \rho_{kt3} \left(U + \frac{\tau^{(g)}}{\omega}\right)^2\right]}, \]

subject to the above constraints.

To see how we obtain the above constraints, note that in dynamic model, we have \( \ln[\sigma_{jt}^{(g)}(U)/\sigma_{1t}^{(g)}(U)] = v_{jt}^{(g)}(U) - v_{1t}^{(g)}(U) \). Using the definition of the payoffs, we can compute the derivative:

\[ \frac{d\ln\left[\sigma_{jt}^{(g)}(U)/\sigma_{1t}^{(g)}(U)\right]}{dU} = -\omega(P_{jt} - P_{1t}) \]

By the above series logit approximation, we have

\[ \ln\left[\frac{\sigma_{jt}^{(g)}(U)}{\sigma_{1t}^{(g)}(U)}\right] = (\rho_{jt,1} - \rho_{1t,1}) + \left[\left(\frac{\rho_{jt2}}{\omega}\right) - \left(\frac{\rho_{1t2}}{\omega}\right)\right] (\omega U + \tau^{(g)}) + \left[\left(\frac{\rho_{jt3}}{\omega^2}\right) - \left(\frac{\rho_{1t3}}{\omega^2}\right)\right] (\omega U + \tau^{(g)})^2. \quad (C.2) \]
which implies
\[
\frac{\mathrm{d} \ln \left[ \frac{\sigma^{(g)}_{jt}(U)}{\sigma^{(g)}_{1t}(U)} \right]}{\mathrm{d}U} = \omega \left[ \left( \frac{\rho_{jt2}}{\omega} \right) - \left( \frac{\rho_{1t2}}{\omega} \right) \right] + 2 \left[ \left( \frac{\rho_{jt3}}{\omega^2} \right) - \left( \frac{\rho_{1t3}}{\omega^2} \right) \right] (\omega U + \tau^{(g)})\omega.
\]

Equalizing the two formulas of the same derivative gives rise to the conclusion in eq. (C.1).

A useful conclusion is the following. Applying the constraints about \(\omega\)'s to eq. (C.2) for the first group, \(g = 1\), we have
\[
\ln \left[ \frac{\sigma^{(1)}_{jt}(U)}{\sigma^{(1)}_{1t}(U)} \right] = (\rho_{jt,1} - \rho_{1t,1}) - \omega (P_{jt} - P_{1t}).
\]

The dependent variable \(Y_{jt}\) of eq. (Linear-Reg-1), whose definition is copied below, has a simple expression,
\[
Y_{jt} \equiv \int \ln \left[ \frac{\sigma^{(1)}_{jt}(U)}{\sigma^{(1)}_{1t}(U)} \right] \, dF^{(1)}_{jt}(U) + \omega (P_{jt} - P_{1t}) \int \, dF^{(1)}_{jt}(U) = \rho_{jt,1} - \rho_{1t,1}.
\]

This is useful, because the NLS step will estimate \(\rho_{jt,1} - \rho_{1t,1}\). After which, one can estimate \((\delta_j - \delta_1)/(1 - \beta),\gamma/(1 - \beta)\) and \(\alpha^{(1)}\) by running 2SLS of \((\rho_{jt,1} - \rho_{1t,1})\) on \((X_{jt} - X_{1t})\) and \((P_{jt} - P_{1t})\) according to eq. (Linear-Reg-1). This also proves Proposition 2.

Remark C.1 (More structure, more constraints). It is interesting to ask does a more restrictive demand model give rise to more constraints for the NLS problem, eq. (11), of estimating CCP. Here we consider the myopic model by letting \(\beta = 0\), and obtain the restrictions about \(\rho\). We claim that the myopic model implies the following constraints about the parameters \(\rho_t\):
\[
\rho_{jt2} = -\omega P_{jt} \quad \text{and} \quad \rho_{jt3} = 0.
\]

Compared with the dynamic model, we now only need to estimate \(\rho_{jt1}\) for each product \(j\) and period \(t\), in addition to \(\tau\) and \(\omega\). That is 2T less parameters to estimate. It can be shown (see proof below) that the above constraints also imply
\[
\rho_{jt1} = \delta_j + \gamma' X_{jt} - \alpha^{(1)} P_{jt} + \xi_{jt}.
\]

This is an instructive conclusion—it says that \(\rho_{jt1}\) equals the mean value of product \(j\) among the consumers from group 1 as defined in Berry (1994).

In the rest, we show how we obtain the constraints in eq. (C.3). We need to return to the structural model. When \(\beta = 0\), we conclude from the log CCP ratio between product \(j\) and the outside option 0, eq. (5), which is copied below for ease of reference,
\[
\ln \left[ \frac{\sigma^{(g)}_{jt}(U)}{\sigma^{(g)}_{1t}(U)} \right] = \delta_j + \gamma' X_{jt} - \left( \alpha^{(1)} + \tau^{(g)} \right) P_{jt} + \xi_{jt} - \omega U P_{jt}.
\]

From the above equation, we have
\[
\frac{\mathrm{d} \ln \left[ \frac{\sigma^{(g)}_{jt}(U)}{\sigma^{(g)}_{1t}(U)} \right]}{\mathrm{d}U} = -\omega P_{jt}.
\]

45
Note that it does not depend on $U$. On the other hand, it follows from series logit properties that
\[
\ln\left[\frac{\sigma_{jt}(U)}{\sigma_{0t}(U)}\right] = \rho_{jt1} + \left(\frac{\rho_{jt2}}{\omega}\right)(\omega U + \tau^{(g)}) + \left(\frac{\rho_{jt3}}{\omega^2}\right)(\omega U + \tau^{(g)})^2,
\]
which implies an alternative expression of the same derivative:
\[
\frac{d\ln[\sigma_{jt}(U)/\sigma_{0t}(U)]}{dU} = \left(\frac{\rho_{jt2}}{\omega}\right)\omega + 2\left(\frac{\rho_{jt3}}{\omega^2}\right)(\omega U + \tau^{(g)})\omega.
\]
Equalizing these two expressions of the same term, we reach the conclusion in eq. (C.3). Note that under the constraints in eq. (C.3), we have
\[
\ln\left[\frac{\sigma_{jt}(U)}{\sigma_{0t}(U)}\right] = \rho_{jt1} - P_{jt}(\omega U + \tau^{(g)}).
\]
Using this expression and eq. (C.4), it is straightforward to show the formula of $\rho_{jt1}$.

**Initial values**

Good initial values help solve the NLS in eq. (11). We need initial values of $\tau^{(\text{init})} = (\tau^{(2)}, \ldots, \tau^{(G)})$, $\rho^{(\text{init})}$, and $\omega^{(\text{init})}$. We follow two steps to obtain the initial values, and these steps are based on the first order Taylor expansion of CCP function $\sigma_{jt}^{(g)}(U)$.

In the first step, we find $\tau^{(\text{init})}$ by running 2SLS for the following linear regression,
\[
\ln\left(\frac{S_{jt}^{(g)}}{S_{1t}^{(g)}}\right) = \delta_j - \delta_1 + (X_{jt} - X_{1t})\gamma - (\alpha^{(1)} + \tau^{(g)})(P_{jt} - P_{1t}) + \frac{\xi_{jt} - \xi_{1t}}{1 - \beta).
\]
To see the rationale, recall the identity
\[
S_{jt}^{(g)} = E[\sigma_{jt}^{(g)}(U^*)\Gamma_{t}^{(g)}(U^*)], \quad U^* \sim N(0, 1),
\]
and consider the first order Taylor expansion of CCP function $\sigma_{jt}^{(g)}(U^*)\Gamma_{t}^{(g)}(U^*)$ at 0, which is the mean of $U^* \sim N(0, 1)$, for each group $g$. We have
\[
S_{jt}^{(g)} = E[\sigma_{jt}^{(g)}(U^*)\Gamma_{t}^{(g)}(U^*)] \\
\approx \sigma_{jt}^{(g)}(0)\Gamma_{t}^{(g)}(0) + \left(\frac{d\sigma_{jt}^{(g)}(U^*)\Gamma_{t}^{(g)}(U^*)}{dU^*}\right)_{U^*=0}E(U^* - 0) \\
= \sigma_{jt}^{(g)}(0)\Gamma_{t}^{(g)}(0).
\]
The first order Taylor expansion leads to $S_{jt}^{(g)} \approx \sigma_{jt}^{(g)}(0)\Gamma_{t}^{(g)}(0)$. Thus,
\[
\frac{S_{jt}^{(g)}}{S_{1t}^{(g)}} \approx \frac{\sigma_{jt}^{(g)}(0)}{\sigma_{1t}^{(g)}(0)}.
\]
The application of this conclusion to eq. (Linear-Reg-1) when $U = 0$ gives rise to the stated regression.
In the second step, we find $\rho_{jt1,\text{init}}$ for all $j = 1, \ldots, J$, and $(\rho_{1t2}/\omega)_{\text{init}}, (\rho_{1t3}/\omega^2)_{\text{init}}$ by running OLS for the following linear regression,

$$
\ln \left( \frac{S_{jt}}{S_{0t}} \right) + (P_{jt} - P_{1t})\tau(g)_{\text{init}} = \rho_{jt1} + \left( \frac{\rho_{1t2}}{\omega} \right)\tau(g)_{\text{init}} + \left( \frac{\rho_{1t3}}{\omega^2} \right)(\tau(g)_{\text{init}})^2.
$$

We now explain how we got the above regression. It follows from series logit that

$$
\ln \left[ \frac{\sigma_{jt}^{(g)}(0)}{\sigma_{0t}^{(g)}(0)} \right] = \ln \left[ \frac{\sigma_{jt}^{(1)}(\tau(g)/\omega; \rho_t)}{\sigma_{0t}^{(1)}(\tau(g)/\omega; \rho_t)} \right] = \rho_{jt1} + \left( \frac{\rho_{jt2}}{\omega} \right)\tau(g) + \left( \frac{\rho_{jt3}}{\omega^2} \right)(\tau(g))^2
$$

$$
= \rho_{jt1} + \left( \frac{\rho_{1t2}}{\omega} \right)\tau(g) - (P_{jt} - P_{1t})\tau(g) + \left( \frac{\rho_{1t3}}{\omega^2} \right)(\tau(g))^2.
$$

The second line follows from imposing the constraints eq. (C.1). By the approximation,

$$
\ln \left( \frac{S_{jt}}{S_{0t}} \right) \approx \ln \left[ \frac{\sigma_{jt}^{(g)}(0)}{\sigma_{0t}^{(g)}(0)} \right],
$$

we have the stated regression.\(^{24}\)

\section{D Simulation}

In order to determine how well our estimator performs in small samples, we run several simulations that vary the number of products, the number of observed groups, and the degree of within group variation. We designed our numerical experiments to illustrate the applicability of our estimator and to understand the following empirically relevant questions:

- How does the number of products affect the estimation?
- How does the within group variation affect the estimation?
- How does the number of observed groups and the number of periods affect the estimation?
- Does the theory work when there is enormous group difference while there is little within group variation?
- How does the attrition rate affect the estimation?

We address each question in the results section \textbf{D.2} below. We make section \textbf{D.2} self-contained so that readers, who are not interested in the data generating process (DGP) details, can skip the DGP section and jump to the results.

\(^{24}\)We do not have a clever initial value of $\omega_{\text{init}}$. In our simulation, we varied the initial value $\omega_{\text{init}}$ substantially, and our optimization routine seems very robust.
D.1 Data Generating Process

In our DGP, the flow utility function follows the specification in Section 3.1. When consumer $i$ of group $g$ purchases product $j$ in period $t$ in market $m$, she receives the following utility

$$u_{ijtm} = \frac{f(X_{jtm}, \xi_{jtm})}{1 - \beta} - \alpha_i P_{jtm} + \varepsilon_{ijtm},$$

and receives $f(X_{jtm}, \xi_{jtm})$ as flow utility in each period post purchase in period $t$ where

$$\alpha_i = \alpha^{(1)} + \tau^{(2)} D_i^{(2)} + \cdots + \tau^{(G)} D_i^{(G)} + \omega U_i.$$

In all the simulations below (except where noted) we let

$$f(X_{jtm}, \xi_{jtm}) = \delta_j + X_{jtm}' \gamma + \xi_{jtm} = -0.1 + X_{jtm} \times 0.03 + \xi_{jtm},$$

for any product $j$. Thus, $\gamma = 0.03$ and $\delta_j = -0.1$ for any product $j$. For price coefficient $\alpha_i$, let $\alpha^{(1)} = 0.1$, $\tau^{(2)} = 0.05$, $\tau^{(3)} = 0.1$, $\tau^{(4)} = 0.15$, $\tau^{(5)} = 0.2$, $\tau^{(6)} = 0.25$, the within group variation $\omega$ will take one value from $(0.025, 0.05, 0.075)$, and let $U_i$ be a random variable drawn from the standard normal distribution.

Products are differentiated by the observed price, $P_{jtm}$, observed product characteristic $X_{jtm}$ and unobserved characteristics, $\xi_{jtm}$. The discount factor $\beta$ is set to 0.90.

We next describe the data generation process of price, $X_{jtm}$, and the unobserved product characteristics. We specifically account for correlation between $\xi_{jtm}$ and $P_{jtm}$. Such a formulation is motivated by the price endogeneity problem researchers face when employing aggregate data, where firms can observe $\xi_{jtm}$ and then set prices optimally. In practice, we allow for multiple markets where $M = 2$. We use a reduced form price model to characterize this dependence. Specifically,

$$X_{jtm} = r_m + \phi_{X}^r X_{j,t-1,m} + \nu_{jtm}^X,$$

$$\xi_{jtm} = \phi_{\xi} \xi_{j,t-1,m} + \nu_{jtm}^\xi,$$

$$P_{jtm} = c + MC_{jtm} + \nu_{jtm}^P,$$

$$MC_{jtm} = d_j + \phi_{MC}^j MC_{j,t-1,m} + \nu_{jtm}^{MC},$$

where $(\nu_{jtm}^X, \nu_{jtm}^\xi, \nu_{jtm}^P, \nu_{jtm}^{MC})'$ is independent and identically distributed across products, time periods and markets, and follows a multivariate normal distribution,

$$\begin{pmatrix} \nu_{jtm}^X \\ \nu_{jtm}^\xi \\ \nu_{jtm}^P \\ \nu_{jtm}^{MC} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \sigma_X^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & 0 & 0 & 0 \\ 0 & \sigma_\xi^2 & \rho_{\xi \sigma_p} & 0 \\ 0 & \rho_{\xi \sigma_p} & \sigma_p^2 & 0 \\ 0 & 0 & 0 & \sigma_{MC}^2 \end{pmatrix} \right).$$

Here $MC_{jtm}$ denotes the marginal cost of product $j$ at time $t$ in market $m$. We will use $MC_{jtm}$ as the instrumental variable in estimation.
In our simulations, the maximum number of products is 8, and we assign the following parameter values. We let \( c = 3 \), \( (d_1, \ldots, d_8) = (0.21, 0.28, 0.35, 0.42, 0.49, 0.56, 0.63, 0.7) \), \( (r_{m=1}, r_{m=2}) = (0.35, 0.55) \), \( \phi^f = 0 \), \( (\phi^{MC}_1, \ldots, \phi^{MC}_8) = (0.965, 0.94, 0.925, 0.91, 0.895, 0.88, 0.865, 0.85) \) and \( (\phi^r_{m=1}, \phi^r_{m=2}) = (0.35, 0.55) \). For the initial state of \( MC_{jtm} \), we let \( (MC_{1,0,m}, \ldots, MC_{8,0,m}) = (9.5, 9.25, 9.00, 8.75, 8.50, 8.25, 8.00, 7.75) \). Such specification ensures that product marginal cost, \( MC_{jtm} \), has a declining trajectory, which is consistent with durable goods models. As for the \( X \) variable, the initial starting values do not differ across \( j \), but do so across markets with \( (X_{0,m=1}, X_{0,m=2}) = (0.525, 0.825) \). Finally, we let \( \sigma_x = 0.15 \), \( \sigma_\xi = 0.05 \), \( \sigma_p = 0.25 \), \( \sigma_{MC} = 0.1 \) and \( \rho = 1 \).

It is important to note the specified DGP produces own-price elasticities (when all 8 goods are available) in the range of -1 for type 1 consumers to -3.5 for type 6. Additionally, each set of simulations results are based on 50 replications.

### D.2 Results

**Effect of the number of products** \((J = 4 \text{ vs. } J = 6 \text{ vs. } J = 8)\)**

Below we present the results of several Monte Carlo simulations in order to illustrate the performance of our estimator as the number of products, \( J \), increase. Given the time consuming nature of the data generating process we restrict the number of products to be no more than 8. In our empirical exercise, the number of products is 118. Additionally, the number of distinct consumer groups and the within group heterogeneity parameter are held constant at 6 and at a value of \( \omega = 0.075 \), while the number of products in a consumer’s choice set varies from 4 to 8.

We have the following observations based on Table D.1. (a) The estimator has negligible bias, regardless of the number of products. (b) Unlike the other estimators of dynamic discrete choice models, that would suffer from the curse of dimensionality as the number of products increases, a bigger number of products indeed boost the performance of our method by decreasing the standard error. This is more evident for the estimation of the price coefficient \( \alpha^{(1)} \)—when the number of products doubles, the standard error is halved. This observation is very useful because it is common to obtain data about many products. This observation is also new—the big number of products is usually perceived by applied researchers as a source of estimation challenge because it would cause the curse of dimensionality.

**Effect of within group variation:** \((\omega = 0.025 \text{ vs. } \omega = 0.05 \text{ vs. } \omega = 0.075)\)**

The within group variation plays an important role in our theory not only because this parameter itself has important economic interpretation but also because when there is no within group variation at all, different groups of consumers do not overlap in terms of price sensitivity, hence the market share in one group is uninformative about the consumers choice in the other groups. In the end our identification arguments will break.
Table D.1: Simulation Results: Comparison Across Number of Products

DGP: $M = 2$, $T = 12$ and $\omega = 0.075$

| J  | $\delta$     | $\gamma$     | $\alpha^{(1)}$ | $\tau^{(2)}$ | $\tau^{(3)}$ | $\tau^{(4)}$ | $\tau^{(5)}$ | $\omega$ | $\beta$     |
|----|------------|------------|----------------|------------|------------|------------|------------|         |------------|
| 4  | -0.1 -0.1025 (0.0132) | 0.0312 (0.0047) | 0.1001 (0.0090) | 0.0501 (2.35e-5) | 0.1000 (4.16e-5) | 0.1500 (5.67e-5) | 0.2001 (6.99e-5) | 0.0753 (1.09e-4) | 0.8981 (8.44e-4) |
| 6  | -0.1023 (0.0101) | 0.0313 (0.0042) | 0.1004 (0.0069) | 0.0502 (1.68e-5) | 0.1001 (2.82e-5) | 0.1501 (3.73e-5) | 0.2003 (4.56e-5) | 0.0758 (7.75e-5) | 0.8978 (7.51e-4) |
| 8  | -0.1028 (0.0083) | 0.0313 (0.0037) | 0.1004 (0.0055) | 0.0502 (1.63e-5) | 0.1002 (2.72e-5) | 0.1502 (3.50e-5) | 0.2004 (4.09e-5) | 0.0760 (7.46e-5) | 0.8978 (7.20e-4) |

Note: Mean and standard deviation (in parenthesis) for 50 simulations.

The simulation shows that our estimator works very well even when the within group variation is very small. In Table D.2, we consider the estimation when the within group variation $\omega = 0.025$, 0.05, and 0.075. (a) There is no noticeable bias for all estimates, except for $\omega$ itself. Only when $\omega$ is very small, $\omega = 0.025$, there is small bias in estimating $\omega$—the mean of our estimates is 0.0272 as opposed to the true value 0.025. (b) Small $\omega$ indeed only affects the estimation of $\omega$ itself—it does not affect the estimation of all the other parameters, including product fixed effect, price coefficient, and the discount factor.

Effect of the number of groups and the number of periods: (2 Groups vs. 6 Groups; 12 Periods vs. 36 Periods)

The method to estimate CCPs, i.e. eq. (11), is NLS, whose degree of freedom is driven by the number of products, the number of groups $G$, and the number of periods $T$. In Table D.3, we check the performance of our estimator, when $G = 2$ and 6, and $T = 12$ and 36. For the case of $G = 2$, we chose the most challenging case, in which the two selected groups are the most distinct pair in terms of the difference of price coefficients. By checking this “near boundary” case, we show the robustness of the proposed estimator.

We have following observations from Table D.3 (a) Comparing the case ($G = 2, T = 12$) and ($G = 6, T = 12$), we find that small number of groups does not affect the estimation of product fixed effect ($\delta$), coefficients associated with observed product characteristics ($\gamma$), the price coefficient of the base group ($\alpha^{(1)}$),
Table D.2: Simulation Results: Comparison Across Within Heterogeneity

DGP: \( M = 2, T = 12 \) and \( J = 8 \)

<table>
<thead>
<tr>
<th>( \omega )</th>
<th>( \omega = 0.025 )</th>
<th>( \omega = 0.05 )</th>
<th>( \omega = 0.075 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta )</td>
<td>-0.10</td>
<td>-0.1043 (0.0083)</td>
<td>-0.1039 (0.0083)</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.03</td>
<td>0.0317 (0.0037)</td>
<td>0.0316 (0.0037)</td>
</tr>
<tr>
<td>( \alpha^1 )</td>
<td>0.10</td>
<td>0.1003 (0.0055)</td>
<td>0.1003 (0.0055)</td>
</tr>
<tr>
<td>( \tau^2 )</td>
<td>0.05</td>
<td>0.0503 (1.44e-5)</td>
<td>0.0503 (1.27e-5)</td>
</tr>
<tr>
<td>( \tau^3 )</td>
<td>0.10</td>
<td>0.1002 (2.75e-5)</td>
<td>0.1002 (2.18e-5)</td>
</tr>
<tr>
<td>( \tau^4 )</td>
<td>0.15</td>
<td>0.1502 (4.09e-5)</td>
<td>0.1502 (2.96e-5)</td>
</tr>
<tr>
<td>( \tau^5 )</td>
<td>0.20</td>
<td>0.2004 (5.34e-5)</td>
<td>0.2003 (3.59e-5)</td>
</tr>
<tr>
<td>( \tau^6 )</td>
<td>0.25</td>
<td>0.2510 (6.66e-5)</td>
<td>0.2509 (4.22e-5)</td>
</tr>
<tr>
<td>( \omega )</td>
<td></td>
<td>0.0272 (2.40e-4)</td>
<td>0.0512 (9.12e-5)</td>
</tr>
<tr>
<td>( \beta )</td>
<td></td>
<td>0.8968 (7.39e-4)</td>
<td>0.8971 (7.31e-4)</td>
</tr>
</tbody>
</table>

Note: Mean and standard deviation (in parenthesis) for 50 simulations.

and the discount factor (\( \beta \)). (b) When \( G = 2 \) and \( T = 12 \) (one year of monthly data), there is small bias in estimating the within group variation \( \omega \) and price coefficient of the non-base group, and such bias vanishes when \( T \) increases to 36 (three years of monthly data).

**Biased estimation of discount factor caused by nonstationarity when ignore attrition**

In Remark 2, we point out that non-random attrition of consumers will cause estimation bias in two ways—misspecification of the distribution of unobserved heterogeneity and nonstationary. The second (nonstationariness) is more subtle. To enhance this point, we consider using the correctly estimated distribution of unobserved heterogeneity, but in the estimation of the discount factor, we use

\[
\hat{\beta} = \frac{1}{G-1} \sum_{g=2}^{G} \frac{E(\bar{W}_t^{(1)} - E(\bar{W}_t^{(g)}))}{E(\bar{W}_{t+1}^{(1)} - E(\bar{W}_{t+1}^{(g)})) + E[\int \ln \sigma_{0,t+1}^{(1)}(U) dF_{t}^{(1)}(U)] - E[\int \ln \sigma_{0,t+1}^{(2)}(U) dF_{t}^{(2)}(U)],}
\]

where, we use \( T^{-1} \sum_{t=1}^{T} \bar{W}_t^{(g)} \) as the estimator of \( E(\bar{W}_t^{(g)}) \), and use \( T^{-1} \sum_{t=1}^{T} \int \ln \sigma_{0,t+1}^{(1)}(U) dF_{t}^{(1)}(U) \) as the estimator of \( E(\int \ln \sigma_{0,t+1}^{(1)}(U) dF_{t}^{(1)}(U)) \). The non-random attrition makes \( \bar{W}_t^{(g)} \) nonstationary even when \((X_t, P_t, \xi_t)\) is stationary, so that \( T^{-1} \sum_{t=1}^{T} \bar{W}_t^{(g)} \) does not converge.

Table D.4 reports the estimation results using the above procedure. We have two observations. (a) Nonstationarity caused by attrition, if ignored, will bias the estimate of the discount factor, hence bias the estimates of the product fixed effect and preference of observed product characteristics. (b) The bias
Table D.3: Simulation Results: Comparison Across Number of Groups

DGP: $M = 2$, $J = 8$ and $\omega = 0.075$

<table>
<thead>
<tr>
<th></th>
<th>$G = 2$, $T = 12$</th>
<th>$G = 6$, $T = 12$</th>
<th>$G = 2$, $T = 36$</th>
<th>$G = 6$, $T = 36$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>-0.10</td>
<td>-0.1042 (0.0082)</td>
<td>-0.1028 (0.0083)</td>
<td>-0.1046 (0.0032)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.03</td>
<td>0.0316 (0.0037)</td>
<td>0.0314 (0.0037)</td>
<td>0.0308 (0.0020)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.10</td>
<td>0.1006 (0.0055)</td>
<td>0.1004 (0.0055)</td>
<td>0.1011 (0.0027)</td>
</tr>
<tr>
<td>$\tau^2$</td>
<td>0.05</td>
<td>0.0502 (1.63e-5)</td>
<td>0.0502 (1.63e-5)</td>
<td>0.0499 (1.27e-5)</td>
</tr>
<tr>
<td>$\tau^3$</td>
<td>0.10</td>
<td>0.1002 (2.72e-5)</td>
<td>0.1002 (2.72e-5)</td>
<td>0.0997 (2.14e-5)</td>
</tr>
<tr>
<td>$\tau^4$</td>
<td>0.15</td>
<td>0.1502 (3.50e-5)</td>
<td>0.1502 (3.50e-5)</td>
<td>0.1496 (2.70e-5)</td>
</tr>
<tr>
<td>$\tau^5$</td>
<td>0.20</td>
<td>0.2004 (4.09e-5)</td>
<td>0.2004 (4.09e-5)</td>
<td>0.1997 (3.06e-5)</td>
</tr>
<tr>
<td>$\tau^6$</td>
<td>0.25</td>
<td>0.2551 (0.0015)</td>
<td>0.2508 (4.62e-5)</td>
<td>0.2489 (1.31e-4)</td>
</tr>
<tr>
<td>$\omega$</td>
<td>.05</td>
<td>0.0847 (0.0025)</td>
<td>0.0760 (7.46e-5)</td>
<td>0.0736 (2.00e-4)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.90</td>
<td>0.8970 (8.54e-4)</td>
<td>0.8978 (7.20e-4)</td>
<td>0.8983 (5.43e-4)</td>
</tr>
</tbody>
</table>

Note: Mean and standard deviation (in parenthesis) for 50 simulations.

Table D.4: Simulation Results: Bias If Ignore Attrition

DGP: $M = 2$, $T = 12$, $G = 6$ and $J = 8$

<table>
<thead>
<tr>
<th></th>
<th>$\omega = 0.025$</th>
<th>$\omega = 0.05$</th>
<th>$\omega = 0.075$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>-0.10</td>
<td>-0.1295 (0.0088)</td>
<td>-0.1844 (0.0096)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.03</td>
<td>0.0336 (0.0039)</td>
<td>0.0381 (0.0044)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.90</td>
<td>0.8907 (7.10e-4)</td>
<td>0.8758 (7.15e-4)</td>
</tr>
</tbody>
</table>

Note: Mean and standard deviation (in parenthesis) for 50 simulations.

from ignoring attrition will increase with greater within group heterogeneity of price sensitivity. Greater within heterogeneity causes more substantial attrition making the distribution of unobserved consumers heterogeneity change more rapidly over time.

E Multidimensional Unobserved Heterogeneity

In this appendix, we extend our main results to include two dimensional unobserved heterogeneity. For the simplicity of exposition, we limit our attention to myopic model with two dimensional unobserved hetero-
geneity. Using two dimensional unobserved heterogeneity as a concrete case, we are going to make two points. First, our estimation method works for multidimensional unobserved heterogeneity after some modification. Second, a higher dimension of unobserved heterogeneity does not cause the curse of dimensionality for our CCP estimation that involves a series polynomial approximation of CCP as a function of multidimensional unobserved heterogeneity. Particularly, for the case with two dimensional unobserved heterogeneity, the NLS for estimating series CCP has exactly two more parameters (one is the within group dispersion of the new unobserved heterogeneity, and the other is the correlation between the two unobserved heterogeneity) than the NLS of CCP with one dimensional unobserved heterogeneity. This is because the structural model imposes certain restrictions that can eliminate a few parameters of CCP function.

We first setup the model. The original expected payoff of product $j$ in period $t$ is

$$v_{ijt} = \frac{\delta_i + \gamma'X_{ijt} + \xi_{ijt}}{1 - \beta} - \alpha_i P_{jt}.$$  

For simplicity, (1) we let $\beta = 0$ (consumers are myopic), (2) let $\delta_j = 0$, and (3) suppose $X_{ijt}$ is a scalar. The new expected payoff function is

$$v_{ijt} = \gamma_i X_t - \alpha_i P_{jt} + \xi_{ijt}.$$  

We now have one new dimension of unobserved heterogeneity $\gamma_i$ associated with product characteristic $X_t$. Using our group specification, we write

$$(\gamma_i \alpha_i) = \left(\begin{array}{c} \gamma^{(1)} \\ \alpha^{(1)} \end{array}\right) + D^{(2)} \left(\begin{array}{c} \tau^{(2)}_1 \\ \tau^{(2)}_2 \end{array}\right) + \cdots + D^{(G)} \left(\begin{array}{c} \tau^{(G)}_1 \\ \tau^{(G)}_2 \end{array}\right) + \left(\begin{array}{c} \omega_1 U_{i1} \\ \omega_2 U_{i2} \end{array}\right).$$

Below, let $\tau^{(g)} = (\tau^{(g)}_1, \tau^{(g)}_2)'$ and let $U_i = (U_{i1}, U_{i2})'$. Again, let $\tau^{(1)} \equiv (0, 0)'.$ As before, we standardize $U_i$, so that the marginal distributions of $U_{i1}$ and $U_{i2}$ are both standard normal. Their correlation $\zeta = \text{corr}(U_{i1}, U_{i2})$, however, is not restricted and will be estimated, since such correlation is usually of empirical interests. Let $\phi(U; \zeta)$ denote the PDF of a binary normal distribution with unit variance and correlation coefficient $\zeta$.

For a consumer of type-$(g, U_1, U_2)$, the CCP function is

$$\sigma^{(g)}_{jt}(U) = \frac{\exp[(\gamma^{(1)} + \tau^{(g)}_1 + \omega_1 U_{1j})X_{jt} - (\alpha^{(1)} + \tau^{(g)}_2 + \omega_2 U_{2j})P_{jt} + \xi_{jt}]}{1 + \sum_{k=1}^{j} \exp[(\gamma^{(1)} + \tau^{(g)}_1 + \omega_1 U_{1k})X_{kt} - (\alpha^{(1)} + \tau^{(g)}_2 + \omega_2 U_{2k})P_{kt} + \xi_{kt}]}.$$  

This gives rise to

$$\ln \left[ \frac{\sigma^{(g)}_{jt}(U)}{\sigma^{(g)}_{0j}(U)} \right] = (\gamma^{(1)} + \tau^{(g)}_1)X_{jt} - (\alpha^{(1)} + \tau^{(g)}_2)P_{jt} + \xi_{jt} + (\omega_1 U_{1j} X_{jt} - \omega_2 U_{2j} P_{jt}).$$

The market share within group $g$ is as defined before: $S^{(g)}_{jt} = \int \sigma^{(g)}_{jt}(u) \, dF^{(g)}(u)$. The distribution of $U$ satisfies Proposition $\square$ with slight modification—$\phi(u)$ in proposition $\square$ should now be replaced by $\phi(u; \zeta)$.

We claim that we have been able to estimate CCP $\sigma^{(g)}_{jt}(U)$ and the correlation coefficient $\zeta = \text{corr}(U_{1}, U_{2})'$,
then \( \gamma^{(1)}, \alpha^{(1)} \) and \( \tau^{(g)} \) can be estimated by applying 2SLS to the following equation,

\[
\int \ln \left[ \frac{\sigma^{(g)}_{jt}(U)}{\sigma^{(g)}_{0t}(U)} \right] f_t^{(g)}(U) \, dU = (\gamma_1 + \tau_1^{(g)})X_{jt} - (\alpha^{(1)} + \tau_2^{(g)})P_{jt} - \int (\omega_1 U_1 X_{jt} - \omega_2 U_2 P_{jt}) f_t^{(g)}(U) \, dU + \xi_{jt}.
\]

We now start addressing the estimation of CCP function \( \sigma^{(g)}_{jt}(U) \) in this extended model. First of all, our observation about shifting the CCP of one group to obtain the CCP of the other groups still hold. Consider two consumers \( i \) and \( j \); \( i \) is from group 1 and \( j \) is from group 2. If \( U_i = (U_{i1}, U_{i2})' \) and \( U_j = (U_{j1}, U_{j2})' \) satisfy the condition that \( U_{ij} = U_i - \tau_1^{(2)}/\omega_1 \) for \( \ell = 1, 2 \), we have the conclusion that these two consumers have the same expected payoffs of the products. We then have

\[
\sigma^{(2)}_{jt}(U_1, U_2) = \sigma^{(1)}_{jt}(U_1 + \tau_1^{(2)}/\omega_1, U_2 + \tau_2^{(2)}/\omega_2).
\]

Secondly, we need to slightly modify the multinomial series logit approximation. We still write

\[
\sigma^{(1)}_{jt}(U; \rho_t) \equiv L_j \left( R_K(U; \rho_{1t}), \ldots, R_K(U; \rho_{jt}) \right),
\]

where \( L_j \) is a multinomial logit model,

\[
L_j(c_1, \ldots, c_j) = \frac{\exp(c_j)}{1 + \sum_{k=1}^J \exp(c_k)},
\]

and \( R_K(U; \rho_{jt}) \) is a polynomial function of \( U = (U_1, U_2) \),

\[
R_K(U; \rho_{jt}) = R_K(U_1, U_2; \rho_{jt}) = \rho_{j11} + \rho_{j12} U_1 + \rho_{j13} U_1^2 + \rho_{j21} U_2 + \rho_{j22} U_2^2 + \rho_{j31} U_1 U_2 + \ldots
\]

Though such polynomial expansion suggests that there would much more parameters to estimate than the case with one dimensional unobserved heterogeneity, we will show that it is not the case. The estimation is still based on

\[
S^{(g)}_{jt} = \int \sigma^{(1)}_{jt}(u_1 + \frac{\tau_1^{(g)}}{\omega_1}, u_2 + \frac{\tau_2^{(g)}}{\omega_2}) f_t^{(g)}(u) \, d\xi
= \int \sigma^{(1)}_{jt}(u_1 + \frac{\tau_1^{(g)}}{\omega_1}, u_2 + \frac{\tau_2^{(g)}}{\omega_2}) I_t^{(g)}(u_1, u_2) \phi(u_1) \phi(u_2) \, d\xi.
\]

Later, we need to use Gauss–Hermite to approximate the above expectation. In order to make the nodes and weights of Gauss–Hermite independent of the unknown correlation coefficient \( \zeta \), we derive the following result. Using the properties of bivariate normal distribution, we have

\[
S^{(g)}_{jt} = \int \int \sigma^{(1)}_{jt}(u_1 + \frac{\tau_1^{(g)}}{\omega_1}, u_2 + \frac{\tau_2^{(g)}}{\omega_2}) I_t^{(g)}(u_1, u_2) \phi(u_1) \phi(u_2) \, du_1 \, du_2
= \sqrt{1 - \zeta^2} \int \int \sigma^{(1)}_{jt}(u_1 + \frac{\tau_1^{(g)}}{\omega_1}, \sqrt{1 - \zeta^2} u_2 + \zeta u_1 + \frac{\tau_2^{(g)}}{\omega_2}) I_t^{(g)}(u_1, u_2) \phi(u_1) \phi(u_2) \, du_1 \, du_2
= \sqrt{1 - \zeta^2} \mathcal{E} \left[ \sigma^{(1)}_{jt} \left( U_1 + \frac{\tau_1^{(g)}}{\omega_1}, \sqrt{1 - \zeta^2} U_2 + \zeta U_1 + \frac{\tau_2^{(g)}}{\omega_2} \right) I_t^{(g)}(U_1^*, U_2^*) \right],
\]

54
where \((U_1^*, U_2^*)'\) follows the standard bivariate normal distribution. Redefine the Gauss–Hermite approximate below

\[
GH_{jt}^{(g)}(\tau, \omega_1, \omega_2, \zeta, \rho) \equiv \sqrt{1 - \zeta^2} \sum_{i=1}^{n} \eta_i \times \left[ \sigma_j^{(1)}(u_{1i} + \frac{\tau_j^{(g)}}{\omega_1}, \sqrt{1 - \zeta^2} u_{2i} + \zeta u_{1i} + \frac{\tau_j^{(g)}}{\omega_2}) \Gamma_{jt}^{(g)}(u_{1i}, u_{2i}) \right].
\]

The estimation of \(\tau, \omega_1, \omega_2, \zeta, \rho\) is based on the NLS principle.

For the rest, we want to verify the constraints implied by this myopic model with two dimensional unobserved heterogeneity. From the structural model itself, we conclude

\[
\frac{d \ln[\sigma_j^{(g)}(U)/\sigma_0^{(g)}(U)]}{dU_1} = -\omega_1 X_{jt}, \quad \text{and} \quad \frac{d \ln[\sigma_j^{(g)}(U)/\sigma_0^{(g)}(U)]}{dU_2} = -\omega_2 P_{jt}.
\]

Comparing the above derivatives with the resulted derivatives from series logit form, we conclude

\[
\rho_{jt2,X} = -\omega_1 X_{jt}, \quad \rho_{jt2,P} = -\omega_2 P_{jt}, \quad \rho_{jt3,X} = \rho_{jt3,P} = \rho_{jt4,XP} = \cdots = 0.
\]

We reach an interesting conclusion that including two dimensional unobserved heterogeneity indeed does not generate any new coefficients in the series logit model—we still have to and only need to estimate \(\rho_{jt1}\) for each \(j\) and \(t\). The only new parameters involved here are \(\omega_1, \omega_2, \zeta\) and additional \(\tau\). The degree of freedom of the NLS problem now is \(GJT - (JT + 2G + 1)\). So unless the number of products and the periods are very small, we can still estimate the model with only 2 groups.