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, unobserved product characteristics, and non-random attrition of consumers.

Keywords: Dynamic discrete choice, dynamic selection, market shares

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 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 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 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, 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}
mator 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.

The proposed approach relies on our new idea of estimating the conditional choice probability (CCP) functions. 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 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.

Arcidiacono and Miller (2011) made important progress so that the CCP function can depend on an agent’s unobserved discrete type.
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 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.

With having highlighted the paper’s innovations, we believe it is important to

---

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.
discuss the data requirements for implementing our new methodology and its relevance for the on-going discussion of digital privacy. 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 and panel data of product sales for two or more consumer groups (or repeated cross-sectional data of individual consumer purchases where researchers can construct group sales from it), researchers can now account for both needs at no cost. Because customer segmentation is a standard marketing practice, it is easy to obtain panel data at the consumer segment level. For example, the data company NPD provides such consumer segment level panel data for many industries from apparel to video games.

We also believe the need for such a method will grow as consumers demand stronger user-privacy protections and as companies respond to such demand by limiting the collection of individual level panel data. For instance, Google recently announced that its Chrome internet browser will stop supporting third-party cookies (a user-tracking technology) by late 2023 making it very difficult for digital advertising companies to individually target consumers. As an alternative, Google has been testing a new tool called Floc which allows advertisers to follow cohorts of users rather than individuals. We therefore expect our method to be applicable to the data generated by Floc or other similar privacy amicable technologies.

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

---

4Segment level panel data are also substantially cheaper than individual level data if researchers have to purchase data.
variables. The existing identification results (Magnac and Thesmar, 2002; Norets, 2009; Kasahara and Shimotsu, 2009; Arcidiacono and Miller, 2011, 2018; Hu and Shum, 2012; Hu et al., 2017) 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

\footnote{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.}
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 estima-
tors with a variety of approaches. Melnikov (2013) and 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 approx-
imation properties of this approach, and have shown that it is a biased and an
inconsistent estimator when consumers track the full set of state variable. 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 explain the idea
using a BLP-like static model. In section 3, we apply the insights obtained from
the static model to identify and estimate the dynamic discrete demand model. In
section 4, we conduct simulation studies to illustrate the applicability of our method
and to understand a few empirically relevant questions. Section 5 concludes this
paper. The online appendix contains technical proofs/details and some extensions of
the main theory.

2 A Static Discrete Choice Demand Example

2.1 The Static Model

We start with a BLP-like static discrete choice demand model to illustrate the idea.
Suppose that there are $G \geq 2$ demographic groups of consumers (such as income
bracket, age group, sex etc.) in the market facing the same price and products char-
acteristics, observed and unobserved. For a market, we only observe group market
shares, prices and observed characteristics of products over $T$ periods indexed by
$t = 1, \ldots, T$. Consumers choose from products $1, \ldots, J$ with 0 being the outside
A random consumer \(i\) has the following indirect utility of choosing different alternatives in period \(t\):

\[
V_{0t} = \varepsilon_{0t} \quad \text{and} \quad V_{ijt} = \delta_j + \gamma'X_{jt} + \xi_{jt} - \alpha_iP_{jt} + \varepsilon_{ijt}, \quad j = 1, \ldots, J,
\]

where \(X_{jt}\) is a vector of observable product attributes other than price, and \(P_{jt}\) is the price, and \(\xi_{jt}\) is unobserved product characteristics with mean zero. Following Berry [1994], the term \(\delta_j + \xi_{jt}\) can be interpreted as the mean of consumers’ valuation of unobserved product characteristics, such as quality, in one period, and \(\varepsilon_{ijt}\) denotes the individual deviation from this mean. We let \(\xi_{jt}\) has mean zero over \(T\) periods, so \(\delta_j\) is the product fixed effect. Let \(X_t \equiv (X_{1t}, \ldots, X_{Jt})\), and \(P_t\) and \(\xi_t\) are defined similarly.

Assume \(\varepsilon_{it} \equiv (\varepsilon_{i0t}, \ldots, \varepsilon_{ijt})\) follows type 1 extreme value distribution (EVD).

Consumers are heterogeneous in their price coefficient \(\alpha_i\), which depends on discrete demographic groups and unobserved continuous heterogeneity \(U_i\). For expositional simplicity, we limit our attention to the consumer heterogeneity in price coefficient \(\alpha_i\) here and defer the extension to the multidimensional consumer heterogeneity until Appendix D. We use a vector of dummy variables \(D_i \equiv (D_{i}^{(1)}, \ldots, D_{i}^{(G)})'\) to indicate the membership—\(D_{i}^{(g)} = 1\) if consumer \(i\) belongs to group \(g\), and \(D_{i}^{(g)} = 0\) otherwise. Assume \(\sum_{g=1}^{G} D_{i}^{(g)} = 1\). Consider

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

where \(\tau^{(g)}\) captures the between group variation, and \(\omega U_i\) is idiosyncratic unobserved price preference, which captures the within group variation of price coefficient. 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 2 on page 16). We will use the information of the overlapping groups to identify and estimate the model. It will be convenient to define \(\tau^{(1)} = 0\). Hereafter, 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 \(\alpha_i = \alpha^{(1)} + \tau^{(g)} + \omega U\).

---

8To take account of heteroskedasticity (group varying within group variation), it is straightforward to consider the more flexible specification \(\alpha_i = \alpha^{(1)} + \tau^{(2)}D_{i}^{(2)} + \cdots + \tau^{(G)}D_{i}^{(G)} + (\omega^{(1)}D_{i}^{(1)} + \cdots + \omega^{(G)}D_{i}^{(G)})U_i\), where \(\omega^{(g)}\) controls the variation of \(\alpha_i\) within group \(g\).
Let $\sigma_{jt}^{(g)}(U)$ denote the conditional choice probabilities (CCP) of buying product $j$ in period $t$ provided that consumer $i$ is of type-$(g,U)$. It follows from the logit specification that

$$
\sigma_{jt}^{(g)}(U) = \frac{\exp[\delta_j + \gamma' X_{jt} + \xi_{jt} - (\alpha^{(1)} + \tau^{(g)} + \omega U)P_{jt}]}{1 + \sum_{k=1}^{J} \exp[\delta_k + \gamma' X_{kt} + \xi_{kt} - (\alpha^{(1)} + \tau^{(g)} + \omega U)P_{kt}]},
$$

(1)

for $j = 1, \ldots, J$. The CCP $\sigma_{jt}^{(g)}(U)$ can be viewed as a function of $U$ conditional on the current state of market $(X_t, P_t, \xi_t)$.

The “group market share” $S_{jt}^{(g)}$ is defined from averaging unobserved individual CCP $\sigma_{jt}^{(g)}(U)$ over unobserved price sensitivity $U$,

$$
S_{jt}^{(g)} = \int \sigma_{jt}^{(g)}(u) \, dF_{jt}^{(g)}(u),
$$

where $F_{jt}^{(g)}(u)$ is the cumulative distribution function (CDF) of the unobserved price sensitivity $U_i$ of consumers within group $g$ in period $t$. Let $f_{jt}^{(g)}(u)$ denote the respective PDF of $F_{jt}^{(g)}(u)$. In simulating the CCP, which we discuss below, we will match observed group market shares to simulated shares where observed group market share is either directly observed, constructed from observed aggregate sales (to different consumer segments), or constructed from individual panel data by the researchers. Depending on whether or not a consumer would leave the market after purchasing, $f_{jt}^{(g)}(u)$ could vary from group to group and over time. Given our current focus on a static model, we leave the discussion of the group- and time-varying nature of $f_{jt}^{(g)}(u)$ when we discuss the dynamic model.

Assume in the first period the unobserved consumer price sensitivity follows the standard normal distribution. We then can write $f_{jt}^{(g)}(u) = \phi(u)\Gamma_{jt}^{(g)}(u)$, where $\Gamma_{jt}^{(g)}(u) = 1$ if consumers remain the market after purchasing, hence the composition of consumers does not vary (if consumers left the market after purchasing, $\Gamma_{jt}^{(g)}(u)$ is recursively defined by the CCP $\sigma_{01}^{(g)}(u), \ldots, \sigma_{0,t-1}^{(g)}(u)$). Here $\phi(u)$ denotes the PDF of the standard normal distribution, and let $\Phi(u)$ be the respective CDF.

It is easy to see that the log of the CCP ratio is linear in preference parameters:

$$
\ln \left[ \frac{\sigma_{jt}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right] = \delta_j + \gamma' X_{jt} - (\alpha^{(1)} + \tau^{(g)})P_{jt} + \xi_{jt} - \omega U P_{jt}.
$$
Integrating both sides with respect to the distribution of $U$ (which is the standard normal in this static model) and noting that $\text{E}(U) = 0$, we have

$$
\int \ln \left[ \frac{\sigma_{jt}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right] \, d\Phi(U) = \delta_j + \gamma'X_{jt} - (\alpha^{(1)} + \tau^{(g)})P_{jt} + \xi_{jt}. \quad (2)
$$

Provided that the CCP function $\sigma_{jt}^{(g)}(U)$ in the term on the left-hand-side is known, we then identify $\delta_j, \gamma, \alpha^{(1)}, \tau^{(g)}$ using 2SLS. 2SLS is used because price $P_{jt}$ is often correlated with $\xi_{jt}$.

In the actual estimation, we can simplify the procedure by directly estimating $\tau \equiv (\tau^{(2)}, \ldots, \tau^{(G)})'$ and $\omega$ when we estimate the CCP functions $\sigma_{jt}^{(g)}(U)$.

Focus on the first group, and let

$$
Y_{jt}^{\text{static}} \equiv \int \ln \left[ \frac{\sigma_{jt}^{(1)}(U)}{\sigma_{0t}^{(1)}(U)} \right] \, d\Phi(U).
$$

Equation (2) becomes a linear regression,

$$
Y_{jt}^{\text{static}} = \delta_j + \gamma'X_{jt} - \alpha^{(1)}P_{jt} + \xi_{jt}, \quad \text{(Linear-Reg-Static)}
$$

for $j = 1, \ldots, J$. We define $Y_{jt}^{\text{static}}$ because we will estimate this dependent variable as a whole directly during CCP estimation.

### 2.2 CCP Estimation by Using Overlapping Groups of Consumers

This section shows how to directly estimate the CCP $\sigma_{jt}^{(g)}(U)$ as a function of unobserved $U$. We propose a novel method of estimating $\sigma_{jt}^{(g)}(U)$ using observed group market share data and certain constraints implied by the underlying structural model. The estimation only involves solving an NLS with linear constraints. Though the particular form of the constraints depends on whether the model is static or dynamic, the NLS problem remains unchanged when we discuss the estimation of dynamic models.

---

9 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)).
We have $G$ equations from the definition of group market shares $S_{jt}^{(1)}, \ldots, S_{jt}^{(G)}$,

\begin{align*}
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).
\end{align*}

(3)

It seems that there are $G$ unknowns $\sigma_{jt}^{(1)}(U), \ldots, \sigma_{jt}^{(G)}(U)$ in the above $G$ equations, and it appears hopeless to solve $\sigma_{jt}^{(G)}(U)$. Below, we will argue that this is not the case.

### 2.2.1 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 transformations. 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

$$
\delta_j + \gamma'X_{jt} + \xi_{jt} - \left(\alpha^{(1)} + \tau^{(2)}D_{i}^{(2)} + \omega U_i\right)P_{jt}.
$$

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 a high-income group 1, and consumer $\ell$ from a 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.

$$
\sigma^{(2)}_{jt}(U) = \sigma^{(1)}_{jt}(U + \tau^{(2)}/\omega).
$$

Figure 1 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 $\sigma^{(1)}_{jt}(U)$. The application of these restrictions can be better seen after expressing $\sigma^{(1)}_{jt}(U)$ using a series multinomial logit.

We can now rewrite the $G$ group market shares equations about $G$ unknown CCP functions at the beginning, eq. (3), as the following $G$ equations about one unknown $\sigma^{(1)}_{jt}(U)$ by applying our shifting observation $\sigma^{(g)}_{jt}(U) = \sigma^{(1)}_{jt}(U + \tau^{(g)}/\omega)$. Recall that
we defined $\tau^{(1)} = 0$ for convenience. We have

$$S^{(1)}_{jt} = \int \sigma^{(1)}_{jt} \left( u + \frac{\tau^{(1)}}{\omega} \right) d F_t^{(1)}(u)$$

$$\vdots$$

$$S^{(G)}_{jt} = \int \sigma^{(1)}_{jt} \left( u + \frac{\tau^{(G)}}{\omega} \right) d F_t^{(g)}(u)$$

For each period, we now have $J \times G$ equations, but only $J + G$ unknowns, where $J$ refers to the unknown CCP $\sigma^{(1)}_{jt}(U), \ldots, \sigma^{(1)}_{jt}(U)$ 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 $\sigma^{(1)}_{jt}(U)$ (hence the other $\sigma^{(2)}_{jt}(U), \ldots, \sigma^{(G)}_{jt}(U)$ by shifting) from the above equations by parameterizing $\sigma^{(1)}_{jt}(U)$ (so it is known up to finite number of parameters).

### 2.2.2 Series Multinomial Logit Approximation of CCP

We now discuss the details of solving the CCP $\sigma^{(1)}_{jt}(U)$ from eq. 4. The unknown $\sigma^{(1)}_{jt}(U)$ is a continuous function of scalar $U$, whose range is between 0 and 1, and $\sum_{j=0}^J \sigma^{(1)}_{jt}(U) = 1$ for any $U$. We approximate the CCP $\sigma^{(1)}_{jt}(U)$ (as a function of $U$) by interpolation using a “series multinomial logit”, which is a simple extension of the series logit in Hirano, Imbens and Ridder (2003),

$$\sigma^{(1)}_{jt}(U; \rho_t) \equiv L_j(R^*_K(U; \rho_{1t}), \ldots, R^*_K(U; \rho_{Jt})),$$

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 $\rho_{jt} = (\rho_{jt1}, \ldots, \rho_{jtK})^\prime$, and let $\rho_t$ be the collection of $\rho_{1t}, \ldots, \rho_{Jt}$. Lastly,

$$\sigma^{(1)}_{0t}(U; \rho_t) \equiv 1 - \sum_{j=1}^J \sigma^{(1)}_{jt}(U; \rho_t).$$
The idea of a series (multinomial) logit is to use the power series $R_K(U; \rho_{jt})$ to approximate the log odds ratio $\ln[\sigma_{jt}^{(1)}(U)/\sigma_{0t}^{(1)}(U)]$. 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 an economic interpretation—for this static model, $\rho_{jt1}$ is the mean value of product $j$ among the consumers from group 1 as defined in Berry (1994) (see Proposition 1 below). In practice, we found the polynomial of degree 2, i.e. $K = 3$, is sufficient for approximation. Indeed, for the current static model, we can prove that $\rho_{jtk} = 0$ for any $k \geq 3$, though this is not true for the dynamic model. For exposition simplicity, we let $K = 3$ hereafter.

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

For the rest of this section, we will explain the estimation of $(\tau, \omega, \rho)$ leaving a remark about the limitation of our approach 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

$$\sigma_{jt}^{(1)}\left(U + \frac{\tau^{(g)}}{\omega}; \rho_t\right) = \frac{\exp\left[\rho_{jt1} + \rho_{jt2}\left(U + \frac{\tau^{(g)}}{\omega}\right) + \rho_{jt3}\left(U + \frac{\tau^{(g)}}{\omega}\right)^2\right]}{1 + \sum_{k=1}^{J} \exp\left[\rho_{kt1} + \rho_{kt2}\left(U + \frac{\tau^{(g)}}{\omega}\right) + \rho_{kt3}\left(U + \frac{\tau^{(g)}}{\omega}\right)^2\right]}, \quad (6)$$

If consumers remain in the market after purchasing as assumed for our static model, $f_t^{(g)}(U) = \Gamma_t^{(g)}(U)/\phi(U)$, where $\Gamma_t^{(g)}(U) = 1$. The above equation can be rewritten as follows,

$$S_{jt}^{(g)} = E\left[\sigma_{jt}^{(1)}\left(U^* + \frac{\tau^{(g)}}{\omega}; \rho_t\right)\Gamma_t^{(g)}(U^*)\right], \quad U^* \sim \mathcal{N}(0, 1).$$

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:

$$\text{GH}_{jt}^{(g)}(\tau, \omega, \rho) \equiv \sum_{i=1}^{n} \eta_i \times \left[\sigma_{jt}^{(1)}\left(u_i + \frac{\tau^{(g)}}{\omega}; \rho_t\right)\Gamma_t^{(g)}(u_i)\right].$$

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
\[ S^{(g)}_{jt} = \text{GH}^{(g)}_{jt}(\tau, \omega, \rho). \]
We then can estimate the unknown parameters \((\tau, \omega, \rho)\) by NLS:
\[
(\hat{\tau}, \hat{\omega}, \hat{\rho}) \equiv \arg \min_{\tau, \omega, \rho} \sum_{j=1, g=1, t=1}^{J, G, T} \left[ S^{(g)}_{jt} - \text{GH}^{(g)}_{jt}(\tau, \omega, \rho) \right]^2,
\]
subject to
\[
\rho_{jt2} = -\omega P_{jt} \quad \text{and} \quad \rho_{jt3} = 0, \quad j = 1, \ldots, J. \quad \text{(Constraints: Static)}
\]
In Appendix C, we provide the derivation of the constraints eq. ( Constraints: Static) implied by our static demand model. From eq. ( Constraints: Static), we have seen the interpretation of the parameters of series expansion, \(\rho_{jt2}\), 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 1 (Interpretation of series logit parameters \(\rho_{jt1}\) in a static model). Recall that
\[
Y^{\text{static}}_{jt} \equiv \int \ln \left[ \frac{\sigma^{(g)}_{jt}(U)}{\sigma^{(g)}_{0t}(U)} \right] d\Phi(U).
\]
is the dependent variable of our first identification linear regression eq. ( Linear-Reg-Static). We have
\[
\rho_{jt,1} = Y^{\text{static}}_{jt} = \delta_j + \gamma' X_{jt} - \alpha^{(1)} P_{jt} + \xi_{jt}.
\]

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

\footnote{To 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 \(\sigma^{(1)}_{jt}(U + \tau^{(g)}/\omega; \rho_t)\) 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 2: 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 2. Note that $\tau^{(3)}/\omega$ becomes large, when either $\tau^{(3)}$ is large or $\omega$ is small. In this case, though the identity $\sigma^{(3)}_{jt}(U) = \sigma^{(1)}_{jt}(U + \tau^{(3)}/\omega)$ is still valid, the equation about the market share in group 3,

$$S^{(3)}_{jt} = \int \sigma^{(1)}_{jt}(u + \frac{\tau^{(3)}}{\omega}) \ d \Phi(u)$$

has less information for identifying $\sigma^{(1)}_{jt}(U)$. Empirically, this also implies that the observed $S^{(3)}_{jt}$ 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 (Section 4) 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).

### 3 The Dynamic Discrete Demand Model

Up to now, we have made two points. First, by leveraging the information contained in the overlapping groups of consumers, we can simply estimate the CCP function $\sigma^{(g)}_{jt}(U)$ by NLS. Second, knowing $\sigma^{(g)}_{jt}(U)$, the structural parameters in a static BLP model can be estimated by the 2SLS even with the presence of unobserved consumer heterogeneity. In what is to follow, we will show that these two points are still valid when we consider the structural dynamic discrete choice model.

#### 3.1 The Dynamic Model

The timing of our model is the following. In each period $t$, a forward-looking 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

---

11 Although the model is general, it is especially appropriate for durable products, 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. For the simplicity of exposition, we let the product space being fixed. The arguments do not change if we consider time varying choice set.
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 causing non-random attrition of consumers. 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.

The lifetime payoff is a “sum” of discounted per period or flow utilities. We first state the flow utilities. If consumer $i$ does not purchase in period $t$, 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 indirect flow utility during the purchase period $t$ is

$$\delta_j + \gamma'X_{jt} + \xi_{jt} - \alpha_i P_{jt} + \varepsilon_{ijt}. \tag{7}$$

She then receives the identical flow utility $\delta_j + \gamma'X_{jt} + \xi_{jt}$ in each period following her purchase. Let $\Omega_{it} \equiv (X'_{it}, P'_{it}, \xi'_{it}, \varepsilon'_{it})'$. Like the static model, individual price coefficient is

$$\alpha_i = \alpha^{(1)} + \tau^{(2)} D^{(2)}_{i} + \cdots + \tau^{(G)} D^{(G)}_{i} + \omega U_{i}.$$
Assumption 3 (Conditional Independence). For all periods \( t \), we have (i) \( \Omega_{i,t+1} \perp \perp \varepsilon_{it} | (X_t, P_t, \xi_t) \); (ii) \( \varepsilon_{it+1} \perp \perp \Omega_{it} | (X_{t+1}, P_{t+1}, \xi_{t+1}) \); (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 assumes that the transition distribution of the market 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. One interesting feature of our method is that we can identity and estimate the flow utility functions and the discount factor under this weaker assumption of a consumer’s belief.

Consider a consumer \( i \) of type \( (g, U) \), and let \( V_t^{(g)}(\Omega_{it}, U) \) denote her value function.\(^{12}\) Then the option value of waiting to purchase is

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

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

\( ^{12}\) We can write the Bellman equation in terms of the value function \( V_t^{(g)}(\Omega_{it}, U) \) as follows:

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

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 choice of not purchasing in period \( t \) provides flow utility \( \varepsilon_{it} \), 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.
Up to now, we have formalized what the payoffs $v_{it0}, v_{i1t}, \ldots, v_{ijt}$ are. They are $v_{ijt} = v_{ijt}^{(g)}(U)$ for a consumer of type $(g, U)$. Correspondingly, the CCP of type-$(g, U)$ is

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

Lastly, the market share of product $j$ in group $g$ in period $t$ is

$$S_{jt}^{(g)} = \int \sigma_{jt}^{(g)}(u) d F_t^{(g)}(u).$$

### 3.2 Dynamic Selection Problem

Some comments about the composition of consumers in different periods, i.e. $F_t^{(g)}(U)$, are due here. The composition of consumers depends on whether or not consumers 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 $F_t^{(g)}(u)$ does not vary across time as in our static model. 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 $F_t^{(g)}(u)$, depending on the rate of attrition. This is a “selection problem” in dynamic discrete choice. We will show that in order to fix the dynamic selection problem, it is essential to obtain the CCP $\sigma_{jt}^{(g)}(U)$ 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.

First assume that the unobserved price sensitivity follows a normal distribution at the beginning of the sample, which is a fairly standard assumption in the literature.
Depending on whether or not there is attrition, Proposition \ref{prop:attrition} provides a formula of the distribution of price sensitivity for the subsequent periods in terms of the CCP function $\sigma_{jt}^{(g)}(U)$.\footnote{Proposition A.1 in the appendix describes the variation of group composition when there is consumer attrition.}

**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 $F_1^{(g)}(u) = \Phi(u)$.

**Proposition 2** (Distribution of unobserved heterogeneity due to attrition). Suppose Assumptions 1 to 4 holds. Let $f_t^{(g)}(U)$ be the PDF of the unobserved price sensitivity $U$ in period $t$ and group $g$. We have that

$$f_t^{(g)}(u) = \phi(u) \times \Gamma_t^{(g)}(u),$$

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

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

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

$$\Gamma_1^{(g)}(u) = 1, \quad \Gamma_t^{(g)}(u) = \prod_{s=1}^{t-1} \frac{\sigma_{Os}^{(g)}(u)}{s_{0s}}, \quad t \geq 2.$$

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

$$\Gamma_1^{(g)}(u) = 1, \quad \Gamma_{t+1}^{(g)}(u) = \Gamma_t^{(g)}(u) \times \frac{\sigma_{Ot}^{(g)}(u)}{s_{Ot}^{(g)}}.$$

### 3.3 CCP Estimation

Our shifting observation $\sigma_{jt}^{(g)}(U) = \sigma_{jt}^{(1)}(U + \tau^{(g)}/\omega)$ still holds for the dynamic model. To see this, note again that the CCP is determined by comparing expected payoffs of different alternatives. The expected payoff of product $j$ in the dynamic model is

$$v_{ijt} \equiv \frac{\delta_j + \gamma' X_{jt} + \xi_{jt}}{1 - \beta} - \left(\alpha^{(1)} + \tau^{(2)} D_i^{(2)} + \omega U_i\right) P_{jt}.$$
The expected payoff of the outside option depends on the expected future \( v_{ij,t+1}, v_{ij,t+2}, \ldots \) for all products \( j = 1, \ldots, J \). Again, the unknown consumer’s type affects the choice probabilities by altering the expected payoffs, which can be only through the term \( \alpha_i = \alpha^{(1)} + \tau^{(2)} D^{(2)} + \omega U_i \). Using this observation, we can again conclude that \( \sigma_{jt}^{(g)}(U) = \sigma_{jt}^{(1)}(U + \tau^{(g)}/\omega) \). So the focus is to estimate the CCP of one group, \( \sigma_{jt}^{(1)}(U) \), \( \tau \) and \( \omega \). We can again parameterize the CCP \( \sigma_{jt}^{(1)}(U) \) using the series logit and write \( \sigma_{jt}^{(1)}(U; \rho) \).

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

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

but subject to a different set of linear constraints:

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

(Constraints: Dynamic)

In Appendix C, we provide the derivation of the constraints eq. (Constraints: Dynamic) implied by our dynamic model. Comparing the constraints from the static model (eq. (Constraints: Static)) and the ones from the dynamic model, it is interesting to see that a more restrictive demand model (the static model restricts the discount factor \( \beta = 0 \)) give rise to more constraints for the NLS problem of estimating CCP. From eq. (Constraints: Dynamic), 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 3 (Interpretation of series logit parameters \( \rho_{jt1} \) in dynamic model).**

Define

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

We have

\[
Y_{jt}^{\text{dynamic}} = \rho_{jt,1} - \rho_{1t,1}.
\]

---

14 This is most transparent by considering a two period dynamic model. In period 2 (terminal period), \( v_{i,j=0,t+2} = 0 \). The expected optimal payoff in period 2 is \( \ln(1 + \sum_j \exp(v_{ij2})) \). Then the payoff of the outside option in period 1 is \( \beta E[\ln(1 + \sum_j \exp(v_{ij2})) \mid X_1, P_1, \xi_1] \). The statement can be generalized to infinite horizon dynamic programming problem easily.
3.4 Post-CCP Estimation: Structural Parameters

The conclusion we arrive at is that in order to estimate the structural parameters in consumer preferences, including the discount factor, one simply needs to run two linear regressions, eq. (Linear-Reg-1) and eq. (Linear-Reg-2), below. At the end, we provide two remarks 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.

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 observation. Conditional on purchasing in period $t$, a consumer’s choice about which one to buy does not depend on the unknown continuation value $v_{gt}^{(g)}(U)$. We choose product 1 as the reference product and focus on consumer group 1, which results in

$$\ln \left( \frac{\sigma^{(1)}_{jt}(U)}{\sigma^{(1)}_{1t}(U)} \right) = v^{(1)}_{jt}(U) - v^{(1)}_{1t}(U).$$

By the definition of payoff functions $v^{(1)}_{jt}(U)$ in eq. (D.3), and integrating the above display over $U$ with respect its distribution function in period $t$, we have the first condition

$$\int \ln \left( \frac{\sigma^{(1)}_{jt}(U)}{\sigma^{(1)}_{1t}(U)} \right) dF_t^{(g)}(U) + \omega(P_{jt} - P_{1t}) \int U dF_t^{(g)}(U) =$$

$$\frac{\delta_j - \delta_1}{1 - \beta} + (X_{jt} - X_{1t})' \frac{\gamma}{1 - \beta} - \alpha^{(1)}(P_{jt} - P_{1t}) + \xi_{jt} - \xi_{1t} \frac{1}{1 - \beta}. \quad (11)$$

Note that the left-hand-side of the above is exactly $Y^{dynamic}_{jt}$ defined in Proposition 3. Thus, $Y^{dynamic}_{jt} = \rho_{jt,1} - \rho_{1t,1}$ is known after CCP estimation. We conclude that

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

for $j = 2, \ldots, J$. We then identify $(\delta_2 - \delta_1)/(1 - \beta), \ldots, (\delta_J - \delta_1)/(1 - \beta), \gamma/(1 - \beta), \alpha^{(1)}$ using 2SLS.
Model Parameters: Discount Factor and Product Fixed Effect

Identification and estimation of $\beta$ and $\delta_1$ originates from a condition that comes from
\[ \ln(\sigma_{1t}^{(g)}(U)/\sigma_{0t}^{(g)}(U)) = v_{1t}^{(g)}(U) - v_{0t}^{(g)}(U) \] and describes the trade-off from buying now and waiting. By the definition of the payoffs, it becomes
\[
\ln \left[ \frac{\sigma_{1t}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right] = \frac{\delta_1}{1 - \beta} + X'_{1t} \frac{\gamma}{1 - \beta} - (\alpha^{(1)} + \tau^{(g)} + \omega_U)P_{1t} + \frac{\xi_{1t}}{1 - \beta} - \beta E[\bar{V}_{t+1}^{(g)}(U) \mid X_t, P_t, \xi_t]. \tag{12}
\]

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 $\bar{U}$ (e.g. $\bar{U} = 0$),
\[
\mathbb{E}[W_t^{(g)}(\bar{U})] = \delta_1 + \beta \mathbb{E}[W_{t+1}^{(g)}(\bar{U}) + \ln \sigma_{0,t+1}^{(g)}(\bar{U})], \tag{Linear-Reg-2}
\]
where $W_t^{(g)}(\bar{U})$ 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. (12) with four steps. Note that after running 2SLS of eq. (Linear-Reg-1), we already know many parameters including $\sigma_{jt}^{(g)}(U)$, $\tau$, $\omega$, $\gamma/(1 - \beta)$, and $\alpha^{(1)}$. Step 1 is to define $W_t^{(g)}(U)$ by combining the terms that are already known in eq. (12). Let
\[
W_t^{(g)}(U) \equiv \ln \left[ \frac{\sigma_{1t}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right] - \left[ X'_{1t} \frac{\gamma}{1 - \beta} - (\alpha^{(1)} + \tau^{(g)} + \omega_U)P_{1t} \right].
\]

Note that $W_t^{(g)}(U)$ 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 $\bar{V}_{t+1}^{(g)}(U)$ into something we already know using the well known expectation maximization formula for the logit model \cite{Arcidiacono and Miller 2011}: $\bar{V}_{t+1}^{(g)}(U) = v_{1,t+1}^{(g)}(U) - \ln \sigma_{1,t+1}^{(g)}(U)$. We have
\[
\bar{V}_{t+1}^{(g)}(U) = -[W_{t+1}^{(g)}(U) + \ln \sigma_{0,t+1}^{(g)}(U)] + \frac{\delta_1}{1 - \beta} + \frac{\xi_{1,t+1}}{1 - \beta}.
\]

In step 3, we rewrite eq. (12) in terms of $W_t^{(g)}(U)$ and conclude
\[
W_t^{(g)}(U) = \delta_1 + \frac{\xi_{1t}}{1 - \beta} + \beta \mathbb{E}\left[ W_{t+1}^{(g)}(U) + \ln \sigma_{0,t+1}^{(g)}(U) - \frac{\xi_{1,t+1}}{1 - \beta} \mid X_t, P_t, \xi_t \right], \tag{13}
\]
Lastly, in step 4, for a fixed unobserved price sensitivity $\bar{U}$, we take unconditional expectation with respect to $(X_t, P_t, \xi_t)$, and use the condition $E(\xi_t) = E(\xi_{1:t+1}) = 0$ 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 $E[W_t^{(g)}(\bar{U})]$ and $E[\ln \sigma_0^{(g)}(\bar{U})]$ are taken over $(X_t, P_t, \xi_t)$ only with $\bar{U}$ being fixed for each group $g$ and each $t$. This expectation can be estimated by $T^{-1} \sum_{t=1}^{T} W_t^{(g)}(\bar{U})$ when $(X_t, P_t, \xi_t)$ satisfies certain stationarity conditions. With at least two groups (say 1 and 2), we have

$$E[W_t^{(1)}(\bar{U})] = \delta_1 + \beta E[W_{t+1}^{(1)}(\bar{U}) + \ln \sigma_{0,t+1}^{(1)}(\bar{U})]$$
$$E[W_t^{(2)}(\bar{U})] = \delta_1 + \beta E[W_{t+1}^{(2)}(\bar{U}) + \ln \sigma_{0,t+1}^{(2)}(\bar{U})].$$

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

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

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 know $(\delta_j - \delta_1)/(1 - \beta)$.

**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 $U_t$ 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

\[15\] We need the time series $(X_t, P_t, \xi_t)$ is ergodic, and for the chosen fixed $\bar{U}$, $W_t^{(g)}(\bar{U})$, as a function of $(X_t, P_t, \xi_t)$, satisfies certain continuity conditions.
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

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

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

$$
E(\tilde{W}_t^{(g)}) = \delta_1 + \beta E\left[\tilde{W}_{t+1}^{(g)} + \int \ln \sigma_{0,t+1}^{(g)}(U) \, dF_t^{(g)}(U)\right].
$$

We then have an alternative formula of the discount factor,

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

with at least two groups 1 and 2.

The problem is how to estimate $E(\tilde{W}_t^{(g)})$ and $E[\int \ln \sigma_{0,t+1}^{(g)}(U) \, dF_t^{(g)}(U)]$? Taking $E(\tilde{W}_t^{(g)})$ for example, it is tempting to use $T^{-1} \sum_{t=1}^T \tilde{W}_t^{(g)}$ 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, $\tilde{W}_t^{(g)}$ is still non-stationary due to the attrition of consumer. In particular, both

$$
\int \ln \left( \frac{\sigma_{tt}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right) \, dF_t^{(g)}(U) \quad \text{and} \quad \int U \, dF_t^{(g)}(U)
$$

in the definition of $\tilde{W}_t^{(g)}$ are nonstationary. Intuitively, consumers who are less price sensitive purchase and leave the market earlier making the average $\int U \, dF_t^{(g)}(U)$ drift upward over time. Due to this non-stationary property (caused by attrition), the temporal average will not converge in probability to $E(\tilde{W}_t^{(g)})$, which is indeed only well defined for a fixed period. In order to estimate $E(\tilde{W}_t^{(g)})$, 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. $\sigma_{jt}^{(g)}(\tilde{U})$ 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 \( \sigma_j^{(q)}(U) \) 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 \( E[g(X_{t+1}, P_{t+1}, \xi_{t+1}) \mid X_t, P_t, \xi_t] \) (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. (13)), 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. (13). 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. (13) 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.

4 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 4.2 below. We make section 4.2 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.

### 4.1 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,

\[
X_{jtm} = r_m + \phi^x_{m} X_{j,t-1,m} + \nu^x_{jtm}, \\
\xi_{jtm} = \phi^\xi_{j,t-1,m} + \nu^\xi_{jtm}, \\
P_{jtm} = c + MC_{jtm} + \nu^p_{jtm}, \\
MC_{jtm} = d_j + \phi^MC_{j} MC_{j,t-1,m} + \nu^MC_{jtm},
\]

where \((\nu^x_{jtm}, \nu^\xi_{jtm}, \nu^p_{jtm}, \nu^MC_{jtm})'\) is independent and identically distributed across prod-
ucts, time periods and markets, and follows a multivariate normal distribution,

\[
\begin{pmatrix}
\nu^x_{jtm} \\
\nu^\xi_{jtm} \\
\nu^p_{jtm} \\
\nu^MC_{jtm}
\end{pmatrix} \sim \mathcal{N}
\begin{pmatrix}
\sigma^2_x & 0 & 0 & 0 \\
0 & \sigma^2_\xi & \rho \sigma_\xi \sigma_p & 0 \\
0 & \rho \sigma_p \sigma_\xi & \sigma^2_p & 0 \\
0 & 0 & 0 & \sigma^2_{MC}
\end{pmatrix}.
\]

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 fol-
lowing parameter values. We let \( c = 3, (d_1, \ldots, d_8) = (0.21, 0.28, 0.35, 0.42, 0.49, 0.56, 0.63, 0.7), \\
(r_{m=1}, r_{m=2}) = (0.35, 0.55), \phi^x = 0, (\phi^MC_1, \ldots, \phi^MC_8) = (0.965, 0.94, 0.925, 0.91, 0.895, 0.88, 0.865, 0.85) \\
and (\phi^\xi_{m=1}, \phi^\xi_{m=2}) = (0.35, 0.55). \) For the initial state of \( MC_{j0m} \), 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.

4.2 Results

Effect of the number of products ($J = 4$ vs. $J = 6$ 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. 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 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—a larger 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: ($\omega = 0.025$ vs. $\omega = 0.05$ 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 1: Simulation Results: Comparison Across Number of Products
DGP: $M = 2$, $T = 12$ and $\omega = 0.075$

<table>
<thead>
<tr>
<th></th>
<th>$J = 4$</th>
<th>$J = 6$</th>
<th>$J = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>$-0.1$ -0.1025 (0.0132)</td>
<td>-0.1023 (0.0101)</td>
<td>-0.1028 (0.0083)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.03 0.0312 (0.0047)</td>
<td>0.0313 (0.0042)</td>
<td>0.0313 (0.0037)</td>
</tr>
<tr>
<td>$\alpha^{(1)}$</td>
<td>0.10 0.1001 (0.0090)</td>
<td>0.1004 (0.0069)</td>
<td>0.1004 (0.0055)</td>
</tr>
<tr>
<td>$\tau^{(2)}$</td>
<td>0.05 0.0501 (2.35e-5)</td>
<td>0.0502 (1.68e-5)</td>
<td>0.0502 (1.63e-5)</td>
</tr>
<tr>
<td>$\tau^{(3)}$</td>
<td>0.10 0.1000 (4.16e-5)</td>
<td>0.1001 (2.82e-5)</td>
<td>0.1002 (2.72e-5)</td>
</tr>
<tr>
<td>$\tau^{(4)}$</td>
<td>0.15 0.1500 (5.67e-5)</td>
<td>0.1501 (3.73e-5)</td>
<td>0.1502 (3.50e-5)</td>
</tr>
<tr>
<td>$\tau^{(5)}$</td>
<td>0.20 0.2001 (6.99e-5)</td>
<td>0.2003 (4.56e-5)</td>
<td>0.2004 (4.09e-5)</td>
</tr>
<tr>
<td>$\tau^{(6)}$</td>
<td>0.25 0.2503 (8.31e-5)</td>
<td>0.2507 (5.45e-5)</td>
<td>0.2508 (4.62e-5)</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.075 0.0753 (1.09e-4)</td>
<td>0.0758 (7.75e-5)</td>
<td>0.0760 (7.46e-5)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.90 0.8981 (8.44e-4)</td>
<td>0.8978 (7.51e-4)</td>
<td>0.8978 (7.20e-4)</td>
</tr>
</tbody>
</table>

*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 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. (10), 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 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 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 \((\alpha^{(1)})\), 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 (nonstationarity) is more subtle. To enhance this point, we consider using the correctly estimated distribution of unobserved het-

---

### Table 2: Simulation Results: Comparison Across Within Heterogeneity

**DGP: \(M = 2, T = 12\) 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>-0.1043 (0.0083)</td>
<td>-0.1039 (0.0083)</td>
<td>-0.1028 (0.0083)</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>0.0317 (0.0037)</td>
<td>0.0316 (0.0037)</td>
<td>0.0314 (0.0037)</td>
</tr>
<tr>
<td>(\alpha^{(1)})</td>
<td>0.1003 (0.0055)</td>
<td>0.1003 (0.0055)</td>
<td>0.1004 (0.0055)</td>
</tr>
<tr>
<td>(\tau^{2})</td>
<td>0.0503 (1.44e-5)</td>
<td>0.0503 (1.27e-5)</td>
<td>0.0502 (1.63e-5)</td>
</tr>
<tr>
<td>(\tau^{3})</td>
<td>0.1002 (2.75e-5)</td>
<td>0.1002 (2.18e-5)</td>
<td>0.1002 (2.72e-5)</td>
</tr>
<tr>
<td>(\tau^{4})</td>
<td>0.1502 (4.09e-5)</td>
<td>0.1502 (2.96e-5)</td>
<td>0.1502 (3.50e-5)</td>
</tr>
<tr>
<td>(\tau^{5})</td>
<td>0.2004 (5.34e-5)</td>
<td>0.2003 (3.59e-5)</td>
<td>0.2004 (4.09e-5)</td>
</tr>
<tr>
<td>(\tau^{6})</td>
<td>0.2510 (6.66e-5)</td>
<td>0.2509 (4.22e-5)</td>
<td>0.2508 (4.62e-5)</td>
</tr>
<tr>
<td>(\omega)</td>
<td>0.0272 (2.40e-4)</td>
<td>0.0512 (9.12e-5)</td>
<td>0.0760 (7.46e-5)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.8968 (7.39e-4)</td>
<td>0.8971 (7.31e-4)</td>
<td>0.8978 (7.20e-4)</td>
</tr>
</tbody>
</table>

*Note: Mean and standard deviation (in parenthesis) for 50 simulations.*
Table 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>$\alpha_1$</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></td>
<td>0.0502 (1.63e-5)</td>
<td></td>
</tr>
<tr>
<td>$\tau^3$</td>
<td>0.10</td>
<td></td>
<td>0.1002 (2.72e-5)</td>
<td></td>
</tr>
<tr>
<td>$\tau^4$</td>
<td>0.15</td>
<td></td>
<td>0.1502 (3.50e-5)</td>
<td></td>
</tr>
<tr>
<td>$\tau^5$</td>
<td>0.20</td>
<td></td>
<td>0.2004 (4.09e-5)</td>
<td></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>0.075</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.

erogeneity, but in the estimation of the discount factor, we use

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

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

Table 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 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.

33
Table 4: Simulation Results: Bias If Ignore Attrition

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

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

5 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.

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.

---


17 Dropbox once shared its customer data with researchers and incurred ethical concerns. See the report by Emily Dreyfuss, “Was It Ethical for Dropbox to Share Customer Data with Scientists?”, Wired, July 24, 2018, https://www.wired.com/story/dropbox-sharing-data-study-ethics/
References


A Proofs

Proof of Proposition 2. 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

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

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
\[ f_t^{(g)}(u) = \Pr(U_i = u \mid D_t^{(g)} = 1, A_i = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t). \]

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 product state variables \( Z_t \), we also have the following conditional independence,
\[ f_t^{(g)}(u) = \Pr(U_i = u \mid D_t^{(g)} = 1, A_i = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_{t-1}) \]
\[ = \Pr(U_i = u \mid D_t^{(g)} = 1, A_i = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_{t-1}, Z_t). \] (A.2)

We now prove eq. (A.1) by the induction. Starting with period 2, we have
\[ f_2^{(g)}(u) = \Pr(U_i = u \mid D_2^{(g)} = 1, A_i = 0, Z_1) \]
\[ = \frac{f(U_i = u \mid D_2^{(g)} = 1, Z_1)}{f(A_i = 0 \mid D_2^{(g)} = 1, Z_1)} \cdot f(A_i = 0 \mid D_2^{(g)} = 1, Z_1) \]
\[ = \phi(u) \times \frac{\sigma_{01}^{(g)}(u)}{S_0^{(g)}}. \]

The third line used \( f(U_i = u \mid D_2^{(g)} = 1, Z_1) = f(U_i = u \mid D_2^{(g)} = 1) \) because all consumers in group \( g \) face the same product characteristics \( Z_1 \) in the first period.

Suppose eq. (A.1) holds for period \( t \). We will prove that this equation also holds for period \( t + 1 \). We have
\[ f_{t+1}^{(g)}(u) = \Pr(U_i = u \mid D_{t+1}^{(g)} = 1, A_i = 0, \ldots, A_{i,t-1} = 0, A_{i,t} = 0, Z_1, \ldots, Z_t) \]
\[ = \Pr(U_i = u \mid D_{t+1}^{(g)} = 1, A_i = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t) \]
\[ \times \frac{f(A_{i,t} = 0 \mid U_i = u, D_{t+1}^{(g)} = 1, A_i = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)}{f(A_{i,t} = 0 \mid D_{t+1}^{(g)} = 1, A_i = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)} \]
\[ = f_t^{(g)}(u) \times \frac{f(A_{i,t} = 0 \mid U_i = u, D_t^{(g)} = 1, A_i = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)}{f(A_{i,t} = 0 \mid D_t^{(g)} = 1, A_i = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)} \]
\[ = f_t^{(g)}(u) \frac{\sigma_{0t}^{(g)}(u)}{S_0^{(g)}} = \phi(u) \times \prod_{s=1}^{t} \frac{\sigma_{0s}^{(g)}(u)}{S_0^{(g)}}. \]

Note that the purchase choice \( A_{i,t} \) in period \( t \) depends only on the payoffs \( v_{ijt} \), which are functions of \((U_i, D_i^{(g)}, Z_t)\) only. So that \( A_i \perp (A_{i1}, \ldots, A_{i,t-1}, Z_1, \ldots, Z_{t-1})|U_i, D_i^{(g)}, Z_t) \), 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

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

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

$$\pi_2^{(g)} = \Pr(D_i^{(g)} = 1 \mid A_{i1} = 0, Z_1)$$

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

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

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

$$\pi_{t+1}^{(g)} = \Pr(D_i^{(g)} = 1 \mid A_{i1} = 0, \ldots, A_{i,t-1} = 0, A_{it} = 0, Z_1, \ldots, Z_t)$$

$$= \frac{\Pr(A_{it} = 0 \mid D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)}{\Pr(A_{it} = 0 \mid A_{i1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)} \times \frac{S_{0t}^{(g)}}{S_{0t}}.$$  

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 $\Pr(D_i^{(g)} = 1 \mid A_{i1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_{t-1}, Z_t) = \Pr(D_i^{(g)} = 1 \mid A_{i1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_{t-1}) = \pi_i^{(g)}$. 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 $\theta_1 \equiv (\tau, \omega, \rho, (\delta_2 - \delta_1)/(1 - \beta), \ldots, (\delta_J - \delta_1)/(1 - \beta))$. ...
\[ \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

\[ \hat{W}^{(g)}(\hat{\theta}_1) \equiv T^{-1} \sum_{t=1}^{T} W_t^{(g)}(U = 0; \hat{\theta}_1), \quad \text{and} \quad \hat{\ell}^{(g)}(\hat{\theta}_1) \equiv T^{-1} \sum_{t=1}^{T} \ln \sigma_0^{(g)}(U = 0; \hat{\theta}_1). \]

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 \( \sigma_0^{(g)}(0; \hat{\theta}_1) \). We have

\[
W_t^{(g)}(0; \hat{\theta}_1) = \ln \left[ \frac{\hat{\sigma}_{1t}^{(g)}(0)}{\hat{\sigma}_{0t}^{(g)}(0)} \right] - X_{1t}' \left( \frac{\hat{\gamma}}{1 - \hat{\beta}} + (\hat{\alpha}^{(1)} + (\hat{\tau}^{(g)})) \right) P_{1t}
= \ln \left[ \frac{\hat{\sigma}_{1t}^{(1)}(\hat{\tau}^{(g)}/\hat{\omega})}{\hat{\sigma}_{0t}^{(1)}(\hat{\tau}^{(g)}/\hat{\omega})} \right] - X_{1t}' \left( \frac{\hat{\gamma}}{1 - \hat{\beta}} + (\hat{\alpha}^{(1)} + (\hat{\tau}^{(g)})) \right) P_{1t}
= \left[ \hat{\rho}_{j11} + \hat{\rho}_{j12} \frac{\hat{\tau}^{(g)}}{\hat{\omega}} + \hat{\rho}_{j13} \left( \frac{\hat{\tau}^{(g)}}{\hat{\omega}} \right)^2 \right] - X_{1t}' \left( \frac{\hat{\gamma}}{1 - \hat{\beta}} + (\hat{\alpha}^{(1)} + (\hat{\tau}^{(g)})) \right) P_{1t},
\]

and \( \sigma_0^{(g)}(0; \hat{\theta}_1) = \sigma_0^{(1)}(\hat{\tau}^{(g)}/\hat{\omega}; \hat{\theta}_1) \) has the series logit expression. Both \( W_t^{(g)}(0; \hat{\theta}_1) \) and \( \sigma_0^{(g)}(0; \hat{\theta}_1) \) 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

\[
\hat{\theta}_1 \equiv \arg \min_{\theta_1 \in \Theta} \frac{1}{T} \sum_{t=1}^{T} h_t(\theta)'h_t(\theta)
\]

subject to constraints eq. (C.3) below,

where

\[
h_t(\theta_1)' \equiv (h_{1t}(\theta_1)', h_{2t}(\theta_1)', \ldots, h_{Jt}(\theta_1)')
\]

and

\[
h_{jt}(\theta_1) \equiv \begin{bmatrix}
S_{jt}^{(1)} - GH_{jt}^{(G)}(\tau, \omega, \rho) \\
\vdots \\
S_{jt}^{(G)} - GH_{jt}^{(G)}(\tau, \omega, \rho) \\
X_{jt}^{IV} \left[ Y_{jt} - \frac{\delta_j - \delta_1}{1 - \beta} - (X_{jt} - X_{1t})' \frac{\gamma}{1 - \beta} + \alpha(1)(P_{jt} - P_{1t}) \right]
\end{bmatrix}.
\]

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

### C Constraints about CCP

#### C.1 Constraints about CCP from Static Model

We claim that the static model implies the following constraints about the parameters \( \rho_t \):

\[
\rho_{jt2} = -\omega P_{jt} \quad \text{and} \quad \rho_{jt3} = 0. \quad (C.1)
\]

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.1). We need to return to the structural model. For a static model, we conclude from the log CCP ratio between product \( j \) and the outside option 0 that for ease of reference,

\[
\ln \left[ \frac{\sigma_{jt}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right] = \delta_j + \gamma' X_{jt} - (\alpha(1) + \tau^{(g)}) P_{jt} + \xi_{jt} - \omega U P_{jt}. \tag{C.2}
\]

From the above equation, we have

\[
\frac{d \ln[\sigma_{jt}^{(g)}(U)/\sigma_{0t}^{(g)}(U)]}{dU} = -\omega P_{jt}.
\]

Note that it does not depend on \( U \). On the other hand, it follows from series logit properties that

\[
\ln \left[ \frac{\sigma_{jt}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right] = \rho_{jt1} + \left( \frac{\rho_{jt2}}{\omega} \right) (\omega U + \tau^{(g)}) + \left( \frac{\rho_{jt3}}{\omega^2} \right) (\omega U + \tau^{(g)})^2,
\]

which implies an alternative expression of the same derivative:

\[
\frac{d \ln[\sigma_{jt}^{(g)}(U)/\sigma_{0t}^{(g)}(U)]}{dU} = \left( \frac{\rho_{jt2}}{\omega} \right) \omega + 2 \left( \frac{\rho_{jt3}}{\omega^2} \right) (\omega U + \tau^{(g)}) \omega.
\]

Equalizing these two expressions of the same term, we reach the conclusion in eq. (C.1).

Note that under the constraints in eq. (C.1), we have

\[
\ln \left[ \frac{\sigma_{jt}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right] = \rho_{jt1} - P_{jt} (\omega U + \tau^{(g)}).
\]

Using this expression and eq. (C.2), it is straightforward to show the formula of \( \rho_{jt1} \).

### C.2 Constraints about CCP from Dynamic Model

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

\[
\rho_{jt2} - \rho_{1t2} = -\omega (P_{jt} - P_{1t}) \quad \text{and} \quad \rho_{jt3} - \rho_{1t3} = 0, \quad j = 2, \ldots, J. \tag{C.3}
\]

The above constraints eliminate many parameters, leaving the following to estimate in the NLS problem, eq. (10):

\[
(\rho_{11t}, \ldots, \rho_{jt1})', \quad \rho_{1t2}, \quad \rho_{1t3}, \quad \omega, \quad \tau, \quad \text{for } t = 1, \ldots, T.
\]
The degree of freedom of the NLS problem is $JGT - JT - 2T - G$, where $JGT$ is the number of observations, $JT$ results from $\rho_{jt1}$ for each $j = 1, \ldots, J$ and $t = 1, \ldots, T$, $2T$ comes from $(\rho_{1t2}, \rho_{1t3})$ 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,

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

subject to the above constraints.

To see how we obtain the above constraints, note that in dynamic model, we have

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

Using the definition of the payoffs, we can compute the derivative:

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

By the above series logit approximation, we have

$$
\ln \left[ \frac{\sigma_{jt}^{(g)}(U)}{\sigma_{1t}^{(g)}(U)} \right] = (\rho_{jt1} - \rho_{1t1}) + \left[ \left( \frac{\rho_{jt2}}{\omega} \right) - \left( \frac{\rho_{1t2}}{\omega} \right) \right] (\omega U + \tau^{(g)}) + \left[ \left( \frac{\rho_{jt3}}{\omega^2} \right) - \left( \frac{\rho_{1t3}}{\omega^2} \right) \right] (\omega U + \tau^{(g)})^2,
$$

which implies

$$
\frac{d \ln \left[ \frac{\sigma_{jt}^{(g)}(U)}{\sigma_{1t}^{(g)}(U)} \right]}{dU} = \omega \left[ \left( \frac{\rho_{jt2}}{\omega} \right) - \left( \frac{\rho_{1t2}}{\omega} \right) \right] + 2 \left[ \left( \frac{\rho_{jt3}}{\omega^2} \right) - \left( \frac{\rho_{1t3}}{\omega^2} \right) \right] (\omega U + \tau^{(g)}) \omega.
$$

Equalizing the two formulas of the same derivative gives rise to the conclusion in eq. (C.3).

A useful conclusion is the following. Applying the constraints about $\omega$’s to eq. (C.4) for the first group, $g = 1$, we have

$$
\ln \left[ \frac{\sigma_{jt}^{(1)}(U)}{\sigma_{1t}^{(1)}(U)} \right] = (\rho_{jt1} - \rho_{1t1}) - \omega(U(P_{jt} - P_{1t})).
$$
The dependent variable $Y_{jt}$ of eq. (Linear-Reg-1), whose definition is copied below, has a simple expression,

$$Y_{jt} \equiv \int \ln \left[ \frac{\sigma_{j1}^{(1)}(U)}{\sigma_{1t}^{(1)}(U)} \right] dF_{1t}^{(1)}(U) + \omega(P_{jt} - P_{1t}) \int U dF_{1t}^{(1)}(U) = \rho_{jt,1} - \rho_{1t,1}.$$ 

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 3.

**Initial values**

Good initial values help solve the NLS in eq. (10). Here we focus on the dynamic model, which is of our primary interests. We need initial values of $\tau_{init} = (\tau_{init}^{(2)}, \ldots, \tau_{init}^{(G)})'$, $\rho_{init}$, and $\omega_{init}$. 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 $\tau_{init}$ by running 2SLS for the following linear regression,

$$\ln \left( \frac{S_{jt}^{(g)}}{S_{1t}^{(g)}} \right) = \frac{\delta_j - \delta_1}{1 - \beta} + \frac{(X_{jt} - X_{1t})'}{1 - \beta} - (\alpha^{(1)} + \tau^{(g)})(P_{jt} - P_{1t}) + \frac{\xi_{jt} - \xi_{1t}}{1 - \beta}.$$ 

To see the rationale, recall the identity

$$S_{jt}^{(g)} = E[\sigma_{jt}^{(g)}(U^*) \Gamma_t^{(g)}(U^*), \quad U^* \sim \mathcal{N}(0, 1),$$

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 $U^* \sim \mathcal{N}(0, 1)$, for each group $g$. We have

$$S_{jt}^{(g)} \approx E[\sigma_{jt}^{(g)}(U^*) \Gamma_t^{(g)}(U^*)] = \sigma_{jt}^{(g)}(0) \Gamma_t^{(g)}(0).$$

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 $\rho_{j1, \text{init}}$ 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,

$$
\ln \left( \frac{S^{(g)}_{jt}}{S^{(g)}_{0t}} \right) + (P_{jt} - P_{1t})\tau^{(g)}_{\text{init}} = \rho_{j1} + \left( \frac{\rho_{1t2}}{\omega} \right)\tau^{(g)}_{\text{init}} + \left( \frac{\rho_{1t3}}{\omega^2} \right) (\tau^{(g)}_{\text{init}})^2.
$$

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

$$
\ln \left[ \frac{\sigma^{(g)}_{jt}(0)}{\sigma^{(g)}_{0t}(0)} \right] = \ln \left[ \frac{\sigma^{(1)}_{jt}(\tau^{(g)}/\omega; \rho_t)}{\sigma^{(1)}_{0t}(\tau^{(g)}/\omega; \rho_t)} \right] = \rho_{j1} + \left( \frac{\rho_{1t2}}{\omega} \right)^{\tau^{(g)}} + \left( \frac{\rho_{1t3}}{\omega^2} \right) (\tau^{(g)})^2
$$

The second line follows from imposing the constraints eq. (C.3). By the approximation,

$$
\ln \left( \frac{S^{(g)}_{jt}}{S^{(g)}_{0t}} \right) \approx \ln \left[ \frac{\sigma^{(g)}_{jt}(0)}{\sigma^{(g)}_{0t}(0)} \right],
$$

we have the stated regression.

### D Multidimensional Unobserved Heterogeneity

In this appendix, we extend our main results to include multidimensional unobserved heterogeneity. We have two observations. First, our estimation method works for multidimensional unobserved heterogeneity after some modification in the stage of CCP estimation. 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. This is because the structural model imposes certain restrictions that can eliminate a large number of parameters in CCP function.

---

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 Static Demand Model

The new expected payoff function is

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

We now have the new dimension of unobserved heterogeneity \( \gamma_i \) associated with product characteristic \( X_{jt} \). Particularly, when \( X_{jt} \) includes the product dummy variable, the above specification says that consumers could have heterogeneous valuation about the unobserved product characteristics (e.g. advertising)\(^4\). This will be useful if researchers choose the consumers residing in different locations (California and Pennsylvania) as different groups of consumers.

Using our group specification, we write

\[
\begin{pmatrix}
\gamma_i \\
\alpha_i
\end{pmatrix} = \begin{pmatrix}
\gamma^{(1)}_i \\
\alpha^{(1)}_i
\end{pmatrix} + D_i^{(2)} \begin{pmatrix}
\tau_1^{(2)} \\
\tau_2^{(2)}
\end{pmatrix} + \cdots + D_i^{(G)} \begin{pmatrix}
\tau_1^{(G)} \\
\tau_2^{(G)}
\end{pmatrix} + \omega \begin{pmatrix}
U_{i1} \\
U_{i2}
\end{pmatrix}. \tag{D.1}
\]

Below, let \( \tau^{(g)} = (\tau_1^{(g)}, \tau_2^{(g)})' \) and let \( U_i = (U_{i1}', U_{i2})' \). Use \( q \) to denote the dimension of \( (X'_{jt}, P_{jt})' \). Again, let \( \tau^{(1)} = 0 \). For the simplicity of exposition, we let the dispersion of within group heterogeneity be the same, i.e. \( \omega \), and assume \( U_i \sim \mathcal{N}(0, I) \), where \( I \) is an \( q \times q \) identity matrix\(^5\). Let \( \phi(U) \) denote the PDF of the multivariate normal distribution \( \mathcal{N}(0, I) \).

For a consumer of type-\((g, U)\), the CCP function is

\[
\sigma^{(g)}_{jt}(U) = \frac{\exp[\delta_j + (\gamma^{(1)} + \tau^{(g)} + \omega U_1)' X_{jt} - (\alpha^{(1)} + \tau_2^{(g)} + \omega U_2) P_{jt} + \xi_{jt}]}{1 + \sum_{k=1}^J \exp[\delta_k + (\gamma^{(1)} + \tau^{(g)} + \omega U_1)' X_{kt} - (\alpha^{(1)} + \tau_2^{(g)} + \omega U_2) P_{kt} + \xi_{kt}]}.
\]

\(^4\)For example, suppose \( X_{jt} \) is just the product dummy variable that equals 1 for product \( j \) and 0 otherwise. The expected payoff of product \( j \) in period \( t \) then reads \( v_{ijt} = \delta_j + \gamma_i + \xi_{jt} - \alpha_i P_{jt} \), and \( \gamma_i \) serves as random effect that explains consumer heterogeneity in the valuation of unobserved product characteristics. This will be useful if researchers choose the consumers residing in different locations (California and Pennsylvania) as groups of consumers.

\(^5\)In general, we can write

\[
\begin{pmatrix}
\gamma_i \\
\alpha_i
\end{pmatrix} = \begin{pmatrix}
\gamma^{(1)}_i \\
\alpha^{(1)}_i
\end{pmatrix} + D_i^{(2)} \begin{pmatrix}
\tau_1^{(2)} \\
\tau_2^{(2)}
\end{pmatrix} + \cdots + D_i^{(G)} \begin{pmatrix}
\tau_1^{(G)} \\
\tau_2^{(G)}
\end{pmatrix} + \Omega \begin{pmatrix}
U_{i1} \\
U_{i2}
\end{pmatrix},
\]

with unknown invertible \( q \times q \) matrix \( \Omega \). It can be verified that we still have the shifting formula:

\[
\sigma^{(g)}_{jt}(U) = \sigma^{(1)}_{jt}(U + \Omega^{-1} \tau^{(g)}). \]

Then the estimation method still follows.
This gives rise to
\[
\ln \left[ \frac{\sigma_{jt}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)} \right] = \delta_j + (\gamma^{(1)} + \tau_1^{(g)})'X_{jt} - (\alpha^{(1)} + \tau_2^{(g)})P_{jt} + \xi_{jt} + \omega(U_1'X_{jt} - U_2P_{jt}). \tag{D.2}
\]

The market share within group \(g\) is as defined before: \(S_{jt}^{(g)} = \int \sigma_{jt}^{(g)}(u) \, dF_t^{(g)}(u)\). The distribution of \(U\) satisfies Proposition 2 with slight modification—\(\phi(u)\) in proposition 2 should now be replaced by the multivariate normal density \(\phi(u)\).

We claim that we have been able to estimate CCP \(\sigma_{jt}^{(g)}(U)\), the dispersion coefficient \(\omega\) and \(\tau\), then \(\delta_j\), \(\gamma^{(1)}\) and \(\alpha^{(1)}\) can be estimated by applying 2SLS to the following equation,
\[
\int \ln \left[ \frac{\sigma_{jt}^{(1)}(U)}{\sigma_{0t}^{(1)}(U)} \right] f_t^{(1)}(U) \, dU + \omega \int (U_1'X_{jt} - U_2P_{jt})f_t^{(1)}(U) \, dU = \\
\delta_j + \gamma^{(1)}'X_{jt} - \alpha^{(1)}P_{jt} + \xi_{jt}.
\]

We now start addressing the estimation of CCP function \(\sigma_{jt}^{(g)}(U)\) 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. It is easy to verify that
\[
\sigma_{jt}^{(g)}(U) = \sigma_{jt}^{(1)}(U + \omega^{-1}\tau^{(g)}).
\]
Secondly, we need to slightly modify the multinomial series logit approximation. We still write
\[
\sigma_{jt}^{(1)}(U; \rho_t) \equiv L_j \left( R_K(U; \rho_{1t}), \ldots, R_K(U; \rho_{Jt}) \right),
\]
where \(L_j\) is a multinomial logit model as defined before \((L_j(c_1, \ldots, c_J) \equiv \exp(c_j)/(1 + \sum_{k=1}^J \exp(c_k)))\), and \(R_K(U; \rho_{jt})\) is a polynomial function of \(U\),
\[
R_K(U; \rho_{jt}) = \rho_{jt1} + \rho_{jt2}'U + U'\rho_{jt3}U + \cdots,
\]
where \(\rho_{jt2}\) is a \(q \times 1\) vector, and \(\rho_{jt3}\) is a \(q \times q\) matrix. Though such polynomial expansion suggests that there would be much more parameters to estimate than the case with unidimensional unobserved heterogeneity, we will show that it is not the case because the structural demand model imposes many restrictions. The estimation is still based on
\[
S_{jt}^{(g)} = \int \sigma_{jt}^{(1)}(u + \omega^{-1}\tau^{(g)}; \rho_t) \Gamma_t^{(g)}(u) \phi(u) \, du
\]
\[
= \int \sigma_{jt}^{(1)}(u + \omega^{-1}\tau^{(g)}; \rho_t) \Gamma_t^{(g)}(u) \phi_1(u) \, du.
\]
We then use Gauss–Hermite to approximate the above expectation, and estimate \( \tau, \omega, \rho \) by the NLS principle.

For the rest, we want to find the constraints implied by this myopic model with multidimensional unobserved heterogeneity. From the structural model itself (eq. (D.2)), we conclude

\[
\frac{d \ln[\sigma_{jt}^{(g)}(U)/\sigma_{0t}^{(g)}(U)]]}{dU} = \begin{pmatrix} \omega X_{jt} \\ -\omega P_{jt} \end{pmatrix},
\]

and any higher order derivatives are zeros. Comparing the above derivatives with the resulted derivatives from series logit form, we conclude

\[
\rho_{jt2} = \begin{pmatrix} \omega X_{jt} \\ -\omega P_{jt} \end{pmatrix}, \quad \rho_{jt3} = \rho_{jt4} = \cdots = 0.
\]

(Constraints: Static Mult-Dim Heterogeneity)

We reach an interesting conclusion that including multidimensional 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 \( \omega \) and additional \( \tau \). The degree of freedom of the NLS problem now is \( GJT - (JT + q(G - 1) + 1) \) (recall \( q \) is the dimension of \((X'_{jt}, P_{jt})'\)). So unless the number of products and the periods are very small, we can still estimate the model with only 2 groups.

**D.2 Dynamic Demand Model**

We continue by discussing the estimation of a dynamic model with multidimensional heterogeneity, in which the expected lifetime payoff of purchasing product \( j \) at time \( t \) becomes

\[
v_{jt}^{(g)}(U_i) = \frac{\delta_j + \gamma'_j X_{jt} + \xi_{jt} - \alpha_i P_{jt}}{1 - \beta}, \quad j = 1, \ldots, J,
\]

where the vector \((\gamma'_j, \alpha_i)'\) is as defined before in eq. (D.1). Correspondingly, the CCP of type-\((g, U)\) is

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

Because \(v_{jt}^{(g)}(U) = v_{jt}^{(1)}(U + \omega^{-1}\tau^{(g)})\), we still have shifting formula \(\sigma_{jt}^{(g)}(U) = \sigma_{jt}^{(1)}(U + \omega^{-1}\tau^{(g)})\). So the CCP estimation can be based on the same NLS problem excepting for
the constraints about the series approximation coefficients. Once the CCP functions are known, it will be easy to estimate the model structural parameters using our procedures in the post-CCP estimation section.

Like the above static model, the constraints about $\rho_t$ result from the comparison between the derivatives of the log CCP ratio with respect to $U$ in the structural demand model and the same derivatives in the series approximation. Particularly, we have

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

$$
= \frac{\delta_j - \delta_1}{1 - \beta} + \frac{(\gamma^{(1)} + \tau_1^{(g)})}{1 - \beta}(X_{jt} - X_{1t}) - (\alpha^{(1)} + \tau_2^{(g)})(P_{jt} - P_{1t}) + \frac{\xi_{jt} - \xi_{1t}}{1 - \beta} + \\
\omega \left[ \frac{U'_t}{1 - \beta}(X_{jt} - X_{1t}) - U_2(P_{jt} - P_{1t}) \right].
$$

This gives rise to

$$
d \ln \left[ \frac{\sigma_{jt}(U)}{\sigma_{1t}(U)} \right] \frac{dU}{dU} = \begin{pmatrix}
\frac{\omega}{1 - \beta}(X_{jt} - X_{1t}) \\
-\omega(P_{jt} - P_{1t})
\end{pmatrix},
$$

and any higher order derivatives are zeros. Comparing the above derivatives with the resulted derivatives from series logit form, we conclude

$$
\rho_{jt2} - \rho_{1t2} = \begin{pmatrix}
\frac{\omega}{1 - \beta}(X_{jt} - X_{1t})' \\
-\omega(P_{jt} - P_{1t})'
\end{pmatrix}', \quad (\rho_{jt3} - \rho_{1t3}) = (\rho_{jt4} - \rho_{1t4}) = \cdots = 0.
$$

(Constraints: Dynamic Mult-Dim Heterogeneity)

In the series approximation, we only consider the second order approximation, i.e. letting $\rho_{1t4} = \rho_{1t5} = \cdots = 0$. So for the dynamic model with multidimensional heterogeneity, the CCP estimation stage involves the following unknowns,

$$
(\rho_{1t1}, \ldots, \rho_{jt1})'_{q \times 1}, \quad \rho_{jt2}, \quad \rho_{jt3}, \quad \omega, \quad \frac{\omega}{1 - \beta}, \quad \tau_{q \times 1}, \quad \text{for all } t.
$$

Recall $q$ is the dimension of $(X'_{jt}, P_{jt})'$. The degree of freedom of the NLS problem is $GJT - (JT + (q + q^2)T + 2 + q)$. In order to make this degree of freedom be positive, it is necessary to satisfy $(G - 1)J - (1 + q)q$.\footnote{We write $GJT - (JT + (q + q^2)T + 2 + q) = [(G - 1)J - (q + q^2)]T - 2 - q$. When $T$ is relatively large, $(G - 1)J - (q + q^2) > 0$ will also be sufficient to have positive degree of freedom.}

Depending on the number of dimension
of unobserved heterogeneity, we may or may not need a large number of groups or products. It is interesting to point out that even when the number of groups is small, we can ensure a positive degree of freedom by including a large number of products. This manifests one advantage of our method—the number of products, rather than causing the curse of dimensionality, helps solve the curse of dimensionality if we are willing to assume that purchasing is a terminal action in the dynamic model, which is reasonable for the market of durable goods.