Dynamic Discrete Choice Demand Estimation Leveraging Overlapping Groups of Consumers

Cheng Chou* Tim Derdenger†

July 2022

Abstract

Multiple groups of consumers usually reside in the same market. These groups' preferences often "overlap"—for a consumer from a high-income group, there is a consumer from a low-income group that has the same preference (e.g. price) due to within group variation. Leveraging such cross-group matching of consumers, we use each groups' product market shares to estimate the conditional choice probabilities (CCP) as a function of unobserved consumer heterogeneity. Armed with our novel CCP estimator, we develop an approach to identify and estimate a dynamic discrete demand model for durable goods with non-random attrition of consumers and continuous unobserved consumer heterogeneity but without the usual need of value function approximation or reducing the dimension of state space by ad hoc behavioral assumptions. We illustrate the empirical value of our method by estimating consumer demand for electric vehicles in the state of Washington during the period of 2016–2019. We further ascertain the impact of a different federal tax credit based upon a car's MPGe rating rather than battery size, which was the existing policy during the data period.

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 pref-

*Cheng Chou is an Assistant Professor of Economics at School of Business, University of Leicester. Email: cchou@le.ac.uk

†Tim Derdenger is an Associate Professor of Marketing and Strategy at the Tepper School of Business, Carnegie Mellon University. Email:derdenge@andrew.cmu.edu.

erences 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 consumer 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 to 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.[1] The implementation of our new estimator

[1]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 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.

only involves nonlinear least squares (NLS) and 2 stage least squares (2SLS). Particularly, one does not need to solve or simulate the dynamic programming discrete choice model. The estimation simplicity allows researchers to estimate multiple model specifications at little computational cost. With the absence of the curse of dimensionality, it also makes the dynamic demand model more applicable to markets with many differentiated products. Indeed, including more products improves the efficiency of our estimation rather than causing the curse of dimensionality.

We illustrate the empirical value of our method by estimating consumer demand for electric vehicles (EV) in the state of Washington during the period of 2016–2019. Our estimates determine that demand for EVs is driven in part by price, miles per gallon equivalent (MPGe), gas prices and the network size associated with electric charging stations. We also determine there is significant levels of unobserved consumer heterogeneity associated with EV price and MPGe. After estimation, we leverage our estimates to further ascertain the impact of a different federal tax credit based upon a car's MPGe rating rather than battery size, which was the existing policy during the data period.[2] We conclude a simple tax credit of \$76.53 per MPGe measure reduces CO2 omissions by 13.86% and with 7 fewer electric vehicles on the road at no additional cost relative to the existing policy.

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.[3] 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 observed 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

[2]The proposed policy is treated as a temporary (1 period) and unexpected change.

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

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. 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' preferences usually "overlap" statistically—for a consumer from one group there could be a consumer from another group where they both have the same preference (e.g. price) 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.[4] 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 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

[4] The 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.

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.[5] For example, the data company NPD provides such consumer segment level panel data for many industries from apparel to video games.[6]

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.[7] 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 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).[8] The following exceptions involving persistent unobservables are worth noting. Hu and Shum (2012) study dynamic binary choice models with continuous unobserved state variables, but

[5] Segment level panel data are also substantially cheaper than individual level data if researchers have to purchase data.

[6] See "Checkout Segmentation & Survey Insights," NPD.com, accessed September 7, 2021, `https://www.npd.com/products/checkout-segmentation-survey-insights/`.

[7] Patience Haggin, Sam Schechner and Suzanne Vranica, "Google Delays Cookie Removal to Late 2023", The Wall Street Journal, June 24, 2021, `https://www.wsj.com/articles/google-delays-cookie-removal-to-late-2023-11624542064?st=sund6fzua5hdf76&reflink=desktopwebshare_permalink`.

[8] 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.

their identification result is limited to the conditional choice probabilities and state transition distribution functions, not to model primitives like flow utility functions and the discount factor. Norets (2009) does include a serially correlated unobservable *idiosyncratic error*, which is individual-specific rather than an aggregate product shock like in our case. Arcidiacono and Miller (2011) model persistent unobservables, but limit them to a discrete set of values.

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

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

Extending the BLP models to a dynamic setting with forward-looking agents is challenging. Some researchers either do not model persistent unobserved shocks (Song and Chintagunta, 2003), or make them time-invariant (Goettler and Gordon, 2011). Others have focused on improving the computational speed of fixed point estimators with a variety of approaches. Melnikov (2013) and 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 when using aggregate data. Moreover, the formal identification in the paper is not specified. Derdenger and Kumar (2019) have studied the approximation properties of this approach, and have shown that it is a biased and an inconsistent estimator when consumers track the full set of state variable.

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. In Section 5 we provide an application of our methodology by estimating the demand for EV vehicles in the state of Washington. Section 6 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 behind our dynamic estimator in a simple setting. The key observations derived from studying this familiar model will be used to estimate the more challenging dynamic discrete choice model, which is of our primary interest.[9] 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 characteristics, 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 option. A random consumer i has the following indirect utility of choosing different alternatives in period t:

$$V_{i0t} = \varepsilon_{i0t} \quad \text{and} \quad V_{ijt} = \delta_j + \gamma' X_{jt} + \xi_{jt} - \alpha_i P_{jt} + \varepsilon_{ijt}, \quad j = 1, \ldots, J,$$

where X_{jt} is a vector of observable product attributes other than price, and P_{jt} is the price, and ξ_{jt} is unobserved product characteristics . 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 period t. We let ξ_{jt} have mean zero over T periods, so δ_j is the product fixed effect and therefore ε_{ijt} denotes the individual deviation from this mean. Let $X_t \equiv (X_{1t}, \ldots, X_{Jt})$, and P_t and ξ_t are defined similarly. Assume $\varepsilon_{it} \equiv (\varepsilon_{i0t}, \ldots, \varepsilon_{iJt})$ follows type 1 extreme value distribution (EVD).

[9]We understand that a more simple way to estimate the static model than what is presented would be to separately estimate parameters for each group but again the point of this section is to clearly highlight the mechanism for estimating the more complex dynamic model.

Consumers are heterogeneous in their price coefficient α_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 α_i here and defer the extension to the multidimensional consumer heterogeneity until our empirical application in Section 5 with further details in 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 ωU_i is idiosyncratic unobserved price preference, which captures the *within group variation* of the price coefficient. We normalize the variance of U_i to be 1, hence $\omega \geq 0$ controls the size of within group variation.[10] 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 15). 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 $\alpha_i = \alpha^{(1)} + \tau^{(g)} + \omega U$.

Let $\sigma_{jt}^{(g)}(U)$ denote the CCP of buying product j in period t provided that consumer i is of type-(g, U). It follows from the logit specification that

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

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, ξ_t).

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

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

where $F_t^{(g)}(u)$ is the cumulative distribution function (CDF) of the unobserved price sensitivity U_i of consumers within group g in period t. Let $f_t^{(g)}(u)$ denote the respective PDF of $F_t^{(g)}(u)$. The observed group market share is either directly observed, constructed from

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

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, $f_t^{(g)}(u)$ 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 $f_t^{(g)}(u)$ 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 $f_t^{(g)}(u) = \phi(u)\mathit{\Gamma}_t^{(g)}(u)$, where $\mathit{\Gamma}_t^{(g)}(u) = 1$ if consumers remain the market after purchasing, hence the composition of consumers does not vary (if consumers left the market after purchasing, $\mathit{\Gamma}_t^{(g)}(u)$ 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 $\mathrm{E}(U) = 0$, we have

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

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 ξ_{jt}.[11] In the actual estimation, we can simplify the procedure by directly estimating $\boldsymbol{\tau} \equiv (\tau^{(2)}, \ldots, \tau^{(G)})'$ and ω when we estimate the CCP functions $\sigma_{jt}^{(g)}(U)$.

Focus on the first group, for which $\tau^{(1)} = 0$, and let

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

Equation (2) becomes a linear regression,

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

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

[11] 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).

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

$$\begin{aligned} S_{jt}^{(1)} &= \int \sigma_{jt}^{(1)}(u) \; \mathrm{d}\, F_t^{(1)}(u), \\ &\vdots \\ S_{jt}^{(G)} &= \int \sigma_{jt}^{(G)}(u) \; \mathrm{d}\, F_t^{(G)}(u). \end{aligned} \tag{3}$$

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

$$\delta_j + \gamma' X_{jt} + \xi_{jt} - \overbrace{\left(\alpha^{(1)} + \tau^{(2)} D_i^{(2)} + \omega U_i\right)}^{\text{Consumer Heterogeneity}} P_{jt}.$$

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 ℓ from a low-income group 2. Let U_h and U_ℓ be the idiosyncratic price sensitivity relative to their respective group for the two consumers h and ℓ, respectively. Note that if U_h and U_ℓ 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 ℓ because of h's idiosyncratic relatively high price sensitivity. The same price

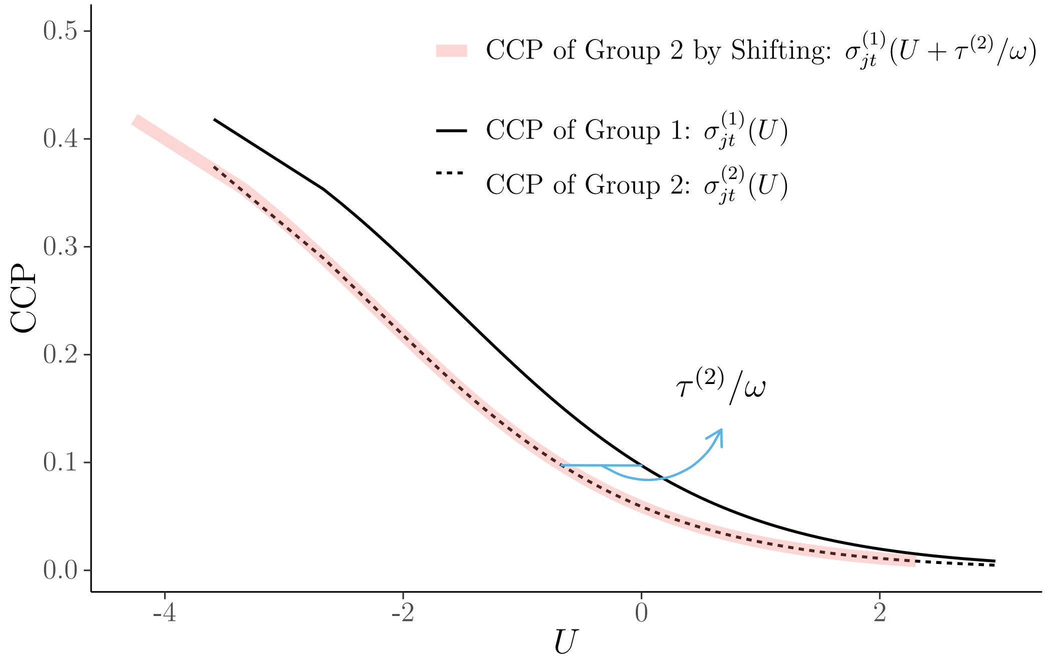

Figure 1: Shift CCP of Group 1 to Obtain CCP of Group 2

coefficient further implies that the two CCPs $\sigma_{jt}^{(2)}(U_\ell) = \sigma_{jt}^{(1)}(U_h)$ when $U_h = U_\ell + \tau^{(2)}/\omega$. In summary, we conclude that two CCPs $\sigma_{jt}^{(1)}(U)$ and $\sigma_{jt}^{(2)}(U)$, viewed as a function U, are essentially identical—we can obtain one by *shifting* the other along the axis of U, i.e.

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

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

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

for convenience. We have

$$
\begin{aligned}
S_{jt}^{(1)} &= \int \sigma_{jt}^{(1)}\Big(u + \frac{\tau^{(1)}}{\omega}\Big)\ \mathrm{d}\, F_t^{(1)}(u) \\
&\ \vdots \qquad\qquad\qquad\qquad\qquad\qquad j = 1, \ldots, J, \quad t = 1, \ldots, T. \\
S_{jt}^{(G)} &= \int \sigma_{jt}^{(1)}\Big(u + \frac{\tau^{(G)}}{\omega}\Big)\ \mathrm{d}\, F_t^{(g)}(u)
\end{aligned}
\tag{4}
$$

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

2.2.2 Series Multinomial Logit Approximation of CCP

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

$$
\sigma_{jt}^{(1)}(U; \rho_t) = L_j(R_K(U; \rho_{1t}), \ldots, R_K(U; \rho_{Jt})), \tag{5}
$$

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 $\rho_{jt} \equiv (\rho_{jt1}, \ldots, \rho_{jtK})'$, and let ρ_t be the collection of $\rho_{1t}, \ldots, \rho_{Jt}$. Lastly,

$$
\sigma_{0t}^{(1)}(U; \rho_t) \equiv 1 - \sum_{j=1}^{J} \sigma_{jt}^{(1)}(U; \rho_t).
$$

The idea of 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 $\boldsymbol{\rho}$ be the $(KJT) \times 1$ vector from stacking ρ_t over all T periods. The coefficients $\boldsymbol{\rho}$ in this series expansion indeed will have an economic interpretation—for this static model, ρ_{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 $(\boldsymbol{\tau}, \omega, \boldsymbol{\rho})$ using an NLS procedure below. Knowing $\boldsymbol{\rho}$, we know group 1 CCP $\sigma_{jt}^{(1)}(U; \rho_t)$ for each alternative j and each period t. Knowing $\boldsymbol{\tau}$ and ω, we know the CCP of the other groups by shifting: $\sigma_{jt}^{(g)}(U) = \sigma_{jt}^{(1)}(U + \tau^{(g)}/\omega)$.

For the rest of this section, we will explain the estimation of $(\boldsymbol{\tau}, \omega, \boldsymbol{\rho})$ leaving a remark about the limitation of our approach at the end. The estimation is based on

$$S_{jt}^{(g)} = \int \sigma_{jt}^{(1)}\Big(u + \frac{\tau^{(g)}}{\omega}; \rho_t\Big) f_t^{(g)}(u) \, \mathrm{d}\, u.$$

where

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

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,

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

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:

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

Here $u_1, \ldots, u_n$ are n nodes, and $\zeta_1, \ldots, \zeta_n$ are the respective weights. Both nodes and weights are predetermined known constants.[12] By Gauss–Hermite approximation, we have

$$S_{jt}^{(g)} = \mathrm{GH}_{jt}^{(g)}(\boldsymbol{\tau}, \omega, \boldsymbol{\rho}).$$

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

$$(\hat{\boldsymbol{\tau}}, \hat{\omega}, \hat{\boldsymbol{\rho}}) \equiv \arg\min_{\boldsymbol{\tau}, \omega, \boldsymbol{\rho}} \sum_{j=1, g=1, t=1}^{J,G,T} \Big[S_{jt}^{(g)} - \mathrm{GH}_{jt}^{(g)}(\boldsymbol{\tau}, \omega, \boldsymbol{\rho})\Big]^2,$$

[12] To be clear, let $u_1^*, \ldots, u_n^*$ be the n nodes of Gauss-Hermite quadrature, and let $\zeta_1^*, \ldots, \zeta_n^*$ be the respective weights. In our simulation, we used $n = 15$ nodes. The nodes and associated weights are determined by the Hermite polynomial, and they do not depend on the function to be approximated, which is $\sigma_{jt}^{(1)}(U + \tau^{(g)}/\omega; \rho_t)$ herein. For $i = 1, \ldots, n$, define $u_i = \sqrt{2}u_i^*$ and $\zeta_i = \zeta_i^*/\sqrt{\pi}$.

subject to

$$\rho_{jt2} = -\omega P_{jt} \qquad \text{and} \qquad \rho_{jt3} = 0, \qquad j = 1, \ldots, J. \qquad \text{(Constraints: Static)}$$

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, ρ_{jt2}, in terms of the structural parameters. The next proposition, whose proof is also in Appendix C, provides the interpretation of ρ_{jt1} using the structural parameters. The proposition states that our NLS procedure and the recovery of $\rho_{jt,1}$ is akin to the contraction mapping of Berry, Levinsohn and Pakes (1995).

Proposition 1 (Interpretation of series logit parameters ρ_{jt1} in a static model)**.** *Recall that*

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

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

$$\rho_{jt,1} = Y_{jt}^{static} = \delta_j + \gamma' X_{jt} - \alpha^{(1)} P_{jt} + \xi_{jt}.$$

Remark 1 (Overlapping Groups)**.** *One key step is the transformation between $\sigma_{jt}^{(1)}(U)$ and $\sigma_{jt}^{(2)}(U)$. Our observation is that for a consumer ℓ from observed group 2 with idiosyncratic U_ℓ, 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.*

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 ω 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 ω is small. In this case, though the identity $\sigma_{jt}^{(3)}(U) = \sigma_{jt}^{(1)}(U + \tau^{(3)}/\omega)$ is still valid, the equation about the market share in group 3,

$$S_{jt}^{(3)} = \int \sigma_{jt}^{(1)}\Big(u + \frac{\tau^{(3)}}{\omega}\Big)\, \mathrm{d}\,\Phi(u)$$

has less information for identifying $\sigma_{jt}^{(1)}(U)$. Empirically, this also implies that the observed $S_{jt}^{(3)}$ is also close to zero. Intuitively, when the two groups are very different ($\tau^{(3)}$ is large) or there is little variation (ω 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).

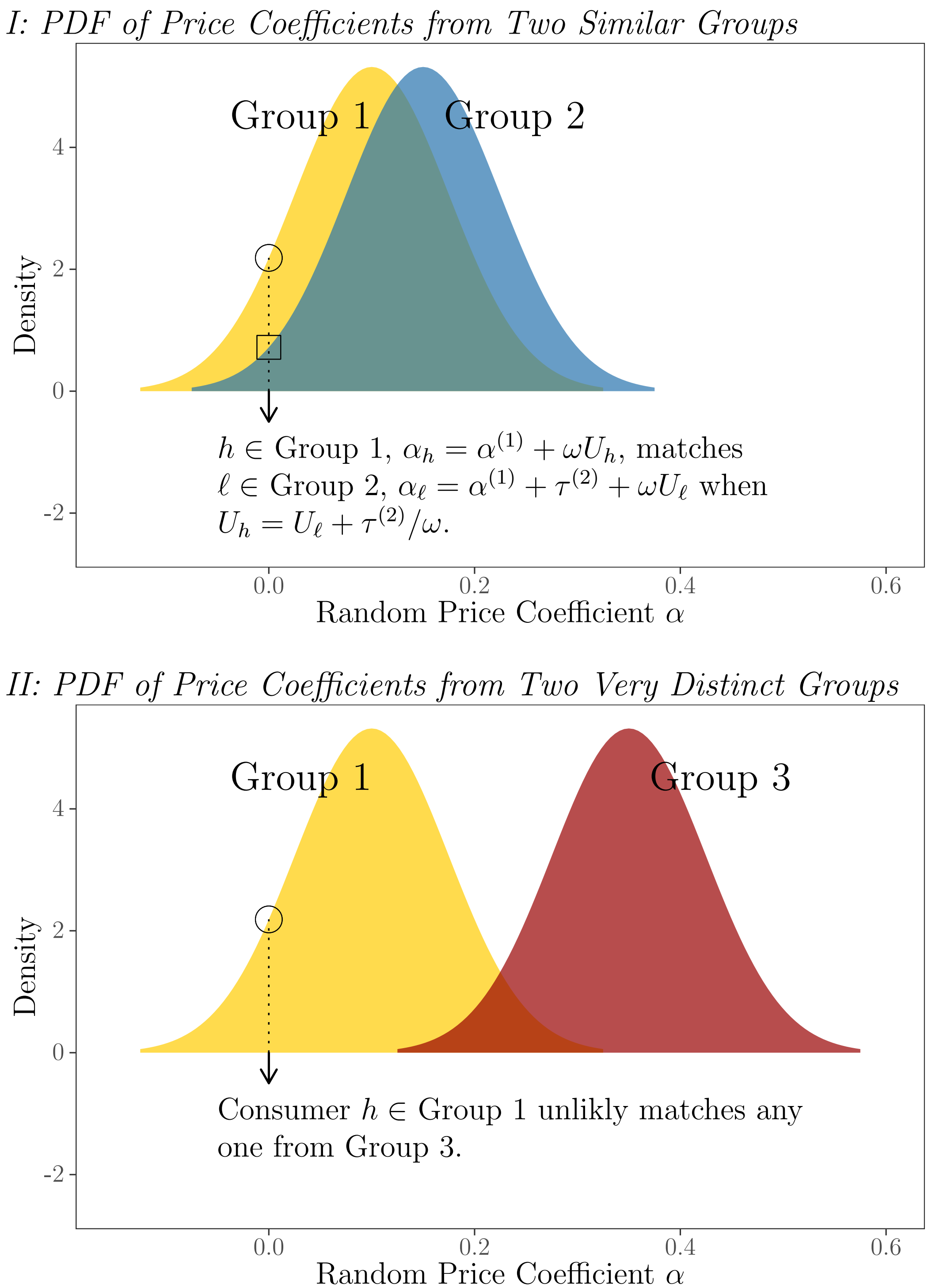

Figure 2: PDF of Price Coefficients from Different Groups

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 $\sigma_{jt}^{(g)}(U)$ by NLS. Second, knowing $\sigma_{jt}^{(g)}(U)$, 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. It is worth pointing out that our estimation strategy below does not require the usual value function approximation or a reduction in the state space by assuming that consumers only consider some function of the state variables (e.g. the inclusive value). Rather, the estimator allows for the consumer to track all state variables individually.

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 Ω_{it} and considers whether or not to purchase a durable product from the available goods $1, \ldots, J$.[13] The associated expected *lifetime* payoffs are $v_{i1t}, \ldots, v_{iJt}$. If she decides to purchase, she then chooses which to buy by comparing payoffs $v_{i1t}, \ldots, v_{iJt}$. Once a consumer has purchased a product, she *exits* the market completely, hence purchasing is a terminal action in our model 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 ε_{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

[13] 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.

purchase. Let $\mathit{\Omega}_{it} \equiv (X_t', P_t', \xi_t', \varepsilon_{it}')'$. Like the static model, individual price coefficient is

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

For a consumer i of type-(g, U), we write $v_{jt}^{(g)}(U)$ to denote her expected lifetime payoffs v_{ijt} from product j. Given the fact that consumers exit the market after the purchase of any product, a consumer's expected lifetime payoff can be written as the sum of the current period t utility and the stream of utilities in periods following purchase:

$$v_{jt}^{(g)}(U) = \frac{\delta_j + \gamma' X_{jt} + \xi_{jt}}{1 - \beta} - (\alpha^{(1)} + \tau^{(g)} + \omega U) P_{jt}, \qquad j = 1, \ldots, J. \tag{8}$$

The discount factor is $\beta \in [0, 1)$. Recall $\alpha^{(1)} + \tau^{(g)} + \omega U$ is the price coefficient of a consumer of type (g, U). To formalize the option value of "waiting" (choosing the outside option), we make the following assumptions, so a consumer's decision becomes a dynamic programming problem.

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

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

Assumption 3 (Conditional Independence)**.** *For all periods t, we have (i) $\mathit{\Omega}_{i,t+1} \perp\!\!\!\perp \varepsilon_{it} | (X_t, P_t, \xi_t)$; (ii) $\varepsilon_{i,t+1} \perp\!\!\!\perp \mathit{\Omega}_{it} | (X_{t+1}, P_{t+1}, \xi_{t+1})$; (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, ξ_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, ξ_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, ξ_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 $V_t^{(g)}(\varOmega_{it}, U)$ denote her value function.[14] Then the option value of waiting to purchase is

$$v_{0t}^{(g)}(U) = \beta \, \mathrm{E}\Big[V_{t+1}^{(g)}(\varOmega_{i,t+1}, U) \;\Big|\; \varOmega_{it}\Big] = \beta \, \mathrm{E}\Big[\bar{V}_{t+1}^{(g)}(X_{t+1}, P_{t+1}, \xi_{t+1}, U) \;\Big|\; X_t, P_t, \xi_t\Big].$$

The second identity follows from applying Assumption 3.

Up to now, we have formalized what the payoffs $v_{i0t}, v_{i1t}, \ldots, v_{iJt}$ are. They are $v_{ijt} = v_{jt}^{(g)}(U)$ for a consumer of type (g, U). Correspondingly, the CCP of type-(g, U) is

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

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

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

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

[14] We can write the Bellman equation in terms of the value function $V_t^{(g)}(\varOmega_{it}, U)$ as follows:

$$V_t^{(g)}(\varOmega_{it}, U) = \max\left(\varepsilon_{i0t} + \beta \, \mathrm{E}\Big[V_{t+1}^{(g)}(\varOmega_{i,t+1}, U) \;\Big|\; \varOmega_{it}\Big], \max_{j \in \{1,\ldots,J\}} v_{jt}^{(g)}(U) + \varepsilon_{ijt}\right).$$

The first term within brackets is the present discounted utility associated with the decision to not purchase any product, i.e. choosing the outside option $j = 0$, in period t. The choice of not purchasing in period t provides flow utility ε_{i0t}, and a term that captures expected future utility conditional on the current state being $\varOmega_{it}$. 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.

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 (Nevo, 2011). Depending on whether or not there is attrition, Proposition 2 provides a formula of the distribution of price sensitivity for the subsequent periods in terms of the CCP function $\sigma_{jt}^{(g)}(U)$.[15]

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 hold. Let $f_t^{(g)}(U)$ be the PDF of the unobserved price sensitivity U in period t and group g. We have that*

$$f_t^{(g)}(u) = \phi(u) \times \varGamma_t^{(g)}(u), \tag{9}$$

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

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

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

$$\varGamma_1^{(g)}(u) = 1, \qquad \varGamma_t^{(g)}(u) = \prod_{s=1}^{t-1} \frac{\sigma_{0s}^{(g)}(u)}{S_{0s}^{(g)}}, \qquad t \geq 2.$$

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

$$\varGamma_1^{(g)}(u) = 1, \qquad \varGamma_{t+1}^{(g)}(u) = \varGamma_t^{(g)}(u) \times \frac{\sigma_{0t}^{(g)}(u)}{S_{0t}^{(g)}}.$$

[15] Proposition A.1 in the appendix describes the variation of group composition when there is consumer attrition.

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

$$v_{ijt} \equiv \frac{\delta_j + \gamma' X_{jt} + \xi_{jt}}{1-\beta} - \overbrace{\left(\alpha^{(1)} + \tau^{(2)} D_i^{(2)} + \omega U_i\right)}^{\text{Consumer Heterogeneity}} P_{jt}.$$

The expected payoff of the outside option depends on the expected future $v_{ij,t+1}, v_{ij,t+2}$, $\ldots$ for all products $j = 1, \ldots, J$.[16] Again the unknown consumer's type affects the choice probabilities by altering the expected payoffs, which can be only through the term $\alpha_i = \alpha^{(1)} + \tau^{(2)} D_i^{(2)} + \omega U_i$. 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, $\sigma_{jt}^{(1)}(U)$, $\boldsymbol{\tau}$ and ω. We can again parameterize the CCP $\sigma_{jt}^{(1)}(U)$ using the series multinomial logit and write $\sigma_{jt}^{(1)}(U; \rho_t)$.

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

$$(\hat{\boldsymbol{\tau}}, \hat{\omega}, \hat{\boldsymbol{\rho}}) \equiv \arg\min_{\boldsymbol{\tau}, \omega, \boldsymbol{\rho}} \sum_{j=1, g=1, t=1}^{J, G, T} \left[S_{jt}^{(g)} - \mathrm{GH}_{jt}^{(g)}(\boldsymbol{\tau}, \omega, \boldsymbol{\rho}) \right]^2 \tag{10}$$

but subject to a different set of linear constraints:

$$\rho_{jt2} - \rho_{1t2} = -\omega(P_{jt} - P_{1t}) \qquad \text{and} \qquad \rho_{jt3} - \rho_{1t3} = 0, \qquad j = 2, \ldots, J. \tag{Constraints: Dynamic}$$

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 $\beta = 0$) gives rise to more constraints for the NLS problem of estimating CCP.

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,

[16] 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 $\beta\,\mathrm{E}[\ln(1 + \sum_j \exp(v_{ij2})) \mid X_1, P_1, \xi_1]$. The statement can be generalized to infinite horizon dynamic programming problem easily.

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 $v_{0t}^{(g)}(U)$. We choose product 1 as the reference product and focus on consumer group 1, which results in

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

By the definition of payoff functions $v_{jt}^{(1)}(U)$ in eq. (8), and integrating the above display over U with respect to its distribution function in period t, we have the *first condition*

$$\int \ln\left[\frac{\sigma_{jt}^{(1)}(U)}{\sigma_{1t}^{(1)}(U)}\right] \mathrm{d}\, F_t^{(g)}(U) + \omega(P_{jt} - P_{1t}) \int U \ \mathrm{d}\, F_t^{(g)}(U) = \\ \frac{\delta_j - \delta_1}{1-\beta} + (X_{jt} - X_{1t})' \frac{\gamma}{1-\beta} - \alpha^{(1)}(P_{jt} - P_{1t}) + \frac{\xi_{jt} - \xi_{1t}}{1-\beta}. \quad (11)$$

The next proposition, whose proof is in Appendix C, provides the interpretation of ρ_{jt1} using the structural parameters. Note that the left-hand-side of the above is exactly $Y_{jt}^{dynamic}$ defined in Proposition 4.

Proposition 3 (Interpretation of series logit parameters ρ_{jt1} in dynamic model)**.** *Define*

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

We have

$$Y_{jt}^{dynamic} = \rho_{jt,1} - \rho_{1t,1}.$$

Thus, $Y_{jt}^{dynamic} = \rho_{jt,1} - \rho_{1t,1}$ is known after CCP estimation. We conclude that

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

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 β and δ_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

$$\ln\left[\frac{\sigma_{1t}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)}\right] = \frac{\delta_1}{1-\beta} + X_{1t}'\frac{\gamma}{1-\beta} - (\alpha^{(1)} + \tau^{(g)} + \omega U)P_{1t} + \frac{\xi_{1t}}{1-\beta} - \beta\,\mathrm{E}[\bar{V}_{t+1}^{(g)}(U) \mid X_t, P_t, \xi_t]. \quad (12)$$

We will identify the discount factor β and product fixed effect δ_1 using the second condition. Before that, we will first show that for any *fixed* unobservable price sensitivity $\tilde{U}$ (e.g. $\tilde{U} = 0$),

$$\mathrm{E}[W_t^{(g)}(\tilde{U})] = \delta_1 + \beta\,\mathrm{E}[W_{t+1}^{(g)}(\tilde{U}) + \ln\sigma_{0,t+1}^{(g)}(\tilde{U})], \quad \text{(Linear-Reg-2)}$$

where $W_t^{(g)}(\tilde{U})$ is estimable and defined below. This equation will give rise to an estimable formula of $(\beta, \delta_1)'$.

We obtain eq. (Linear-Reg-2) from eq. (12) with four steps. Note that after running 2SLS of eq. (Linear-Reg-1), we already know many parameters including $\sigma_{jt}^{(g)}(U)$, $\boldsymbol{\tau}$, ω, $\gamma/(1-\beta)$, 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 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

$$W_t^{(g)}(U) = \delta_1 + \frac{\xi_{1t}}{1-\beta} + \beta\,\mathrm{E}\left(W_{t+1}^{(g)}(U) + \ln\sigma_{0,t+1}^{(g)}(U) - \frac{\xi_{1,t+1}}{1-\beta} \,\middle|\, X_t, P_t, \xi_t\right), \quad (13)$$

Lastly, in *step 4*, for a fixed unobserved price sensitivity $\tilde{U}$, we take *unconditional expectation* with respect to (X_t, P_t, ξ_t), and use the condition $\mathrm{E}(\xi_{1t}) = \mathrm{E}(\xi_{1,t+1}) = 0$ and the law of iterated expectation to reach eq. (Linear-Reg-2).

We now show how to identify the discount factor β and product fixed effect δ_1 using eq. (Linear-Reg-2). The expectations $\mathrm{E}[W_t^{(g)}(\tilde{U})]$ and $\mathrm{E}[\ln \sigma_{0t}^{(g)}(\tilde{U})]$ are taken over (X_t, P_t, ξ_t) only with $\tilde{U}$ being fixed for each group g and each t. This expectation can be estimated by $T^{-1}\sum_{t=1}^{T} W_t^{(g)}(\tilde{U})$ when (X_t, P_t, ξ_t) satisfies certain stationarity conditions.[17] With at least two groups (say 1 and 2), we have

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

We can solve the discount factor β from the above linear system of equations, and obtain

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

The discount factor can be estimated by the sample analog of above formula. Knowing the discount factor β, we recover the fixed effect δ_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_i in the first period, i.e. $F_1^{(1)}(u) = \Phi(u)$. Reading the definition of the dependent variable Y_{jt} of eq.* (Linear-Reg-1)*, it is apparent that ignoring the attrition will misspecify the $F_t^{(1)}(U)$, causing bias in estimating preference parameters. The second is more subtle. Attrition will create a nonstationarity problem, which can be clearly seen from the estimation of discount factor.*

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

$$\bar{W}_t^{(g)} = \int \ln\left[\frac{\sigma_{1t}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)}\right]\,\mathrm{d}\,F_t^{(g)}(U) - X_{1t}'\frac{\gamma}{1-\beta} + (\alpha^{(1)} + \tau^{(g)})P_{1t} + \omega P_{1t}\int U\,\mathrm{d}\,F_t^{(g)}(U),$$

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

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

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

We then have an alternative formula of the discount factor,

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

with at least two groups 1 *and* 2*.*

The problem is how to estimate $\mathrm{E}(\bar{W}_t^{(g)})$ *and* $\mathrm{E}[\int \ln \sigma_{0,t+1}^{(g)}(U) \ \mathrm{d}\, F_t^{(g)}(U)]$*? Taking* $\mathrm{E}(\bar{W}_t^{(g)})$ *for example, it is tempting to use* $T^{-1}\sum_{t=1}^{T} \bar{W}_t^{(g)}$ *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, ξ_t) *satisfies certain stationarity conditions,* $\bar{W}_t^{(g)}$ *is still nonstationary due to the attrition of consumer. In particular, both*

$$\int \ln\left[\frac{\sigma_{1t}^{(g)}(U)}{\sigma_{0t}^{(g)}(U)}\right] \mathrm{d}\, F_t^{(g)}(U) \qquad \textit{and} \qquad \int U \ \mathrm{d}\, F_t^{(g)}(U)$$

in the definition of $\bar{W}_t^{(g)}$ *are nonstationary. Intuitively, consumers who are less price sensitive purchase and leave the market earlier making the average* $\int U \ \mathrm{d}\, F_t^{(g)}(U)$ *drift upward over time. Due to this non-stationary property (caused by attrition), the temporal average will not converge in probability to* $\mathrm{E}(\bar{W}_t^{(g)})$*, which is indeed only well defined for a fixed period. In order to estimate* $\mathrm{E}(\bar{W}_t^{(g)})$*, 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, ξ_t)*. We then can use the temporal average to estimate them and the discount factor. This again highlights the importance of recovering the CCP* $\sigma_{jt}^{(g)}(U)$ *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* $\mathrm{E}[g(X_{t+1}, P_{t+1}, \xi_{t+1})|X_t, P_t, \xi_t]$ *(here* $g(X_{t+1}, P_{t+1}, \xi_{t+1})$ *denotes a generic function of* $(X_{t+1}, P_{t+1}, \xi_{t+1})$*). For example, a consumer's belief about the state transition may not be that of rational expectation.*

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

The intuition for why we can estimate the discount factor without knowing a consumer's belief regarding the law of state transition is less transparent. The key step to identify the discount factor is to take the 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 expectations 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) *ignores the information about the current market state* (X_t, P_t, ξ_t)*, so that consumers belief about* $(X_{t+1}, P_{t+1}, \xi_{t+1})$ *given* (X_t, P_t, ξ_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:

(i) How does the number of products affect the estimation?

(ii) How does the within group variation affect the estimation?

(iii) How does the number of observed groups and the number of periods affect the estimation?

(iv) Does the theory work when there is enormous group difference while there is little within group variation?

(v) 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 α_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 ω 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, ξ_{jtm}. The discount factor β 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 ξ_{jtm} and P_{jtm}. Such a formulation is motivated by the price endogeneity problem researchers face when employing aggregate data, where firms can observe ξ_{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,

$$\begin{aligned}
X_{jtm} &= r_m + \phi_m^x X_{j,t-1,m} + \nu_{jtm}^x, \\
\xi_{jtm} &= \phi^\xi \xi_{j,t-1,m} + \nu_{jtm}^\xi, \\
P_{jtm} &= c + MC_{jtm} + \nu_{jtm}^p, \\
MC_{jtm} &= d_j + \phi_j^{MC} MC_{j,t-1,m} + \nu_{jtm}^{MC},
\end{aligned}$$

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

$$\begin{pmatrix} \nu_{jtm}^x \\ \nu_{jtm}^\xi \\ \nu_{jtm}^p \\ \nu_{jtm}^{MC} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & \sigma_\xi^2 & \rho\sigma_\xi\sigma_p & 0 \\ 0 & \rho\sigma_p\sigma_\xi & \sigma_p^2 & 0 \\ 0 & 0 & 0 & \sigma_{MC}^2 \end{pmatrix}\right).$$

Here MC_{jtm} denotes the marginal cost of product j at time t in market m. We will use MC_{jtm} as the instrumental variable in estimation.

In our simulations, the maximum number of products is 8, and we assign the following parameter values. We let $c = 3$, $(d_1, \ldots, d_8) = (0.21, 0.28, 0.35, 0.42, 0.49, 0.56, 0.63, 0.7)$, $(r_{m=1}, r_{m=2}) = (0.35, 0.55)$, $\phi^{\xi} = 0$, $(\phi_1^{MC}, \ldots, \phi_8^{MC}) = (0.965, 0.94, 0.925, 0.91, 0.895, 0.88, 0.865, 0.85)$ and $(\phi_{m=1}^{x}, \phi_{m=2}^{x}) = (0.35, 0.55)$. 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.

Table 1: Simulation Results: Comparison Across Number of Products
DGP: $M = 2$, $T = 12$ and $\omega = 0.075$

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

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

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.

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 ω itself. Only when ω is very small, $\omega = 0.025$, there is small bias in estimating ω—the mean of our estimates is 0.0272 as opposed to the true value 0.025. (b) Small ω indeed only affects the estimation of ω itself—it does *not* affect the estimation of all the other parameters, including product fixed effect, price coefficient, and the discount factor.

Table 2: Simulation Results: Comparison Across Within Heterogeneity
DGP: $M = 2$, $T = 12$ and $J = 8$

	$\omega = 0.025$	$\omega = 0.05$	$\omega = 0.075$
$\delta = -0.10$	-0.1043 (0.0083)	-0.1039 (0.0083)	-0.1028 (0.0083)
$\gamma = 0.03$	0.0317 (0.0037)	0.0316 (0.0037)	0.0314 (0.0037)
$\alpha^1 = 0.10$	0.1003 (0.0055)	0.1003 (0.0055)	0.1004 (0.0055)
$\tau^2 = 0.05$	0.0503 (1.44e-5)	0.0503 (1.27e-5)	0.0502 (1.63e-5)
$\tau^3 = 0.10$	0.1002 (2.75e-5)	0.1002 (2.18e-5)	0.1002 (2.72e-5)
$\tau^4 = 0.15$	0.1502 (4.09e-5)	0.1502 (2.96e-5)	0.1502 (3.50e-5)
$\tau^5 = 0.20$	0.2004 (5.34e-5)	0.2003 (3.59e-5)	0.2004 (4.09e-5)
$\tau^6 = 0.25$	0.2510 (6.66e-5)	0.2509 (4.22e-5)	0.2508 (4.62e-5)
ω	0.0272 (2.40e-4)	0.0512 (9.12e-5)	0.0760 (7.46e-5)
$\beta = 0.90$	0.8968 (7.39e-4)	0.8971 (7.31e-4)	0.8978 (7.20e-4)

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

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 (δ), coefficients associated with observed product characteristics (γ), the price coefficient of the base group ($\alpha^{(1)}$), and the discount factor (β). (b) When $G = 2$ and $T = 12$ (one year of monthly data), there is small bias in estimating the within group variation ω and price coefficient of the non-base group, and such bias vanishes when T increases to 36 (three years of monthly data).

Table 3: Simulation Results: Comparison Across Number of Groups
DGP: $M = 2$, $J = 8$ and $\omega = 0.075$

	$G = 2$, $T = 12$	$G = 6$, $T = 12$	$G = 2$, $T = 36$	$G = 6$, $T = 36$
$\delta = -0.10$	-0.1042 (0.0082)	-0.1028 (0.0083)	-0.1046 (0.0032)	-0.0977 (0.0033)
$\gamma = 0.03$	0.0316 (0.0037)	0.0314 (0.0037)	0.0308 (0.0020)	0.0300 (0.0020)
$\alpha^1 = 0.10$	0.1006 (0.0055)	0.1004 (0.0055)	0.1011 (0.0027)	0.1003 (0.0027)
$\tau^2 = 0.05$	–	0.0502 (1.63e-5)	–	0.0499 (1.27e-5)
$\tau^3 = 0.10$	–	0.1002 (2.72e-5)	–	0.0997 (2.14e-5)
$\tau^4 = 0.15$	–	0.1502 (3.50e-5)	–	0.1496 (2.70e-5)
$\tau^5 = 0.20$	–	0.2004 (4.09e-5)	–	0.1997 (3.06e-5)
$\tau^6 = 0.25$	0.2551 (0.0015)	0.2508 (4.62e-5)	0.2489 (1.31e-4)	0.2501 (3.40e-5)
$\omega = .075$	0.0847 (0.0025)	0.0760 (7.46e-5)	0.0736 (2.00e-4)	0.0747 (5.10e-5)
$\beta = 0.90$	0.8970 (8.54e-4)	0.8978 (7.20e-4)	0.8983 (5.43e-4)	0.9009 (5.20e-4)

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

Biased estimation of discount factor caused by nonstationarity when attrition is ignored

In Remark 2, we point out that non-random attrition of consumers will cause estimation bias in two ways—misspecifciation 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 heterogeneity, but in the estimation of the discount factor, we use

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

where, we use $T^{-1}\sum_{t=1}^{T} \bar{W}_t^{(g)}$ as the estimator of $\mathrm{E}(\bar{W}_t^{(g)})$, and use $T^{-1}\sum_{t=1}^{T}\int \ln \sigma_{0,t+1}^{(g)}(U)\ \mathrm{d}\, F_t^{(g)}(U)$ as the estimator of $\mathrm{E}(\int \ln \sigma_{0,t+1}^{(1)}(U)\ \mathrm{d}\, F_t^{(1)}(U))$. The non-random attrition makes $\bar{W}_t^{(g)}$ nonstationary even when (X_t, P_t, ξ_t) is stationary, so that $T^{-1}\sum_{t=1}^{T} \bar{W}_t^{(g)}$ 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

Table 4: Simulation Results: Bias If Ignore Attrition

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

	$\omega = 0.025$	$\omega = 0.05$	$\omega = 0.075$
$\delta = -0.10$	-0.1295 (0.0088)	-0.1844 (0.0096)	-0.2693 (0.0111)
$\gamma = 0.03$	0.0336 (0.0039)	0.0381 (0.0044)	0.0453 (0.0052)
$\beta = 0.90$	0.8907 (7.10e-4)	0.8758 (7.15e-4)	0.8525 (7.70e-4)

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

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.

5 Empirical Application: Demand of Electric Vehicles

We illustrate the empirical value of our method by estimating consumer demand for electric vehicles in the state of Washington during the period of 2016–2019. We further ascertain the impact of a different federal tax credit based upon a car's MPGe rating rather than battery size, which was the existing policy during the data period. We conclude a simple tax credit of \$76.53 per MPGe unit is cost neutral to the existing policy and reduces CO2 omissions by 13.86% with 7 fewer electric vehicles on the road.

5.1 Data

The main data originates from new EV registration records from Washington State Department of Licensing. The electric vehicles include battery electric vehicle (BEV) and plug-in hybrid electric vehicle (PHEV). For each vehicle, we observe the first ten digits of vehicle identification number (VIN), from which we collect vehicle characteristics by using a VIN decoder. In the registration records, we also observe the ZIP code of a owner's residence. We then use Internal Revenue Service (IRS) ZIP code income as a proxy of the owner's household income.

In the data, a geographical market is a county, one time period is a quarter of a year, and the group of consumers is defined based on household income. There are three geographical markets in the data: King County ($m = 1$), Snohomish County ($m = 2$), and Pierce County ($m = 3$). Together, these counties account for 77% of the EV market in Washington state.

We observe these geographical markets from quarter 1 (Q1) of 2016 to quarter 4 (Q4) of 2019. There are three income groups in the data. These income groups are based on household income in 2016. Group 1, 2 and 3 consist of households with incomes above the 90-th percentile ($100,966), between the 75-th ($74,353) and 90-th percentile, and below the 75-th percentile, respectively. For each income group and county, we let the initial market size be the half of the number of households in that income group and county in 2016. Recent survey suggests that about half of the respondents consider buying an EV as their new car.[18]

In the data, a product j corresponds to a model car such as the Nissan Leaf. We excluded the models with extremely small market share such as the Smart Fortwo EV. The models we included in the sample account for more than 99% of the total sales of EV in Washington. If two trims of a model are very similar in terms of fuel efficiency and battery range (like BMW 530E and BMW 530E AWD), we collapse them as one model and use their average characteristics as the product characteristics of the collapsed model. In total, we have 29 distinct models from 17 makes.

We obtain additional data from various sources. The gasoline prices by county and period are from Cost of Living Index. The number of electric charging stations in each county and period is from Alternative Fuels Data Center in the US Department of Energy. The number of public facilities in each county and period, which will be used as an instrumental variable of the number of charging stations, is from County Business Patterns from the US Census Bureau. The price is mostly from the transaction price in the registration data. In the registration data, 25.3% of the vehicles do not have valid transaction prices. When the transaction price is unavailable, we use the list price after the deduction of all federal and state-level tax incentives. The available rebates for each model car and the amount of CO2 emission per model car in tons per year are from the US Department of Energy.

Table 5 reports the means of key vehicle characteristics and the sales in different income groups. We note that because electric cars are expensive even after a variety of government financial incentives,[19] buying an EV is more popular among affluent Americans. The top 10% of consumers in terms of income own 47.6% of PHEV and 61.5% of BEV, which is more expensive than PHEV on average, and they own 62.2% and 58.2% of electric cars in 2016 and 2019, respectively. Figure 3 plots the average electric range in miles, the number of charging

[18] Olivier Blanchard, "New Research: Purchase Intent For Hybrids and EVs Has Reached An Inflection Point in North America," Creative Strategies, accessed May 2, 2022, `https://creativestrategies.com/new-research-purchase-intent-for-hybrids-and-evs-has-reached-an-inflection-point-in-north-america/`

[19] The average new car prices in 2016 and 2019 were $34,077 and $36,718, according to Edmunds.

Table 5: Means of Key EV Characteristics

Variable	PHEV	BEV	2016	2019
Price (thousand of $)	42.7	48.0	42.9	47.0
	(17.3)	(28.7)	(21.9)	(23.2)
Federal Tax Credit ($)	4831	6789	6015	5312
	(1354)	(1633)	(1579)	(1817)
Electric Range (miles)	21.1	155.2	69.6	107.5
	(8.75)	(73.59)	(67.25)	(97.92)
Miles per gallon equivalent (MPGe)	48.7	100.9	72.4	75.1
	(18.4)	(14.8)	(31.0)	(31.0)
Horsepower	213	236	212	245
	(88.1)	(138.3)	(89.5)	(136.4)
Tailpipe CO2 emission (ton/year)	1.254	-	0.915	0.737
	(0.728)	-	(0.898)	(0.805)
Sales: all groups	5318	24921	4422	9120
% sales: income below 75th percentile	28.1	17.1	16.7	20.1
% sales: income between 75th and 90th percentile	21.4	21.4	21.1	21.7
% sales: income greater 90th percentile	47.6	61.5	62.2	58.2

Note: CO2 emission of PHEV is based on 15,000 miles driving distance a year, and 37% electric driving in real life (Plötz et al., 2020). The standard deviation is in the parenthesis.

stations in the three counties, average federal tax credit, and sales across time. We observe substantial improvement of electric range and rapid deployment of EV charging stations across time. On the other hand, we see a declining federal tax credit in Figure 3. This is because the tax credit is phased out over time after a manufacturer reaches a total sales (200,000) milestone since 2010.[20] Intuitively, the shrinking tax credit incentivizes consumers to buy EV earlier than later. On the other hand, longer electric range and growing network of charging stations provide the motivation to delay an EV purchase.

Lastly, our data contain only the sales of electric vehicles. We do not observe the sales of the conventional combustion vehicles. This creates a potential issue in our model specification, in which the outside option in a period is to delay the purchase, rather than to

[20] Taking Tesla Model 3 for example, its credit was $7,500 before 12/31/2018, then $3,750 from 1/1/2019 to 6/30/2019, and lastly $1,875 from 7/1/2019 to 12/31/2019.

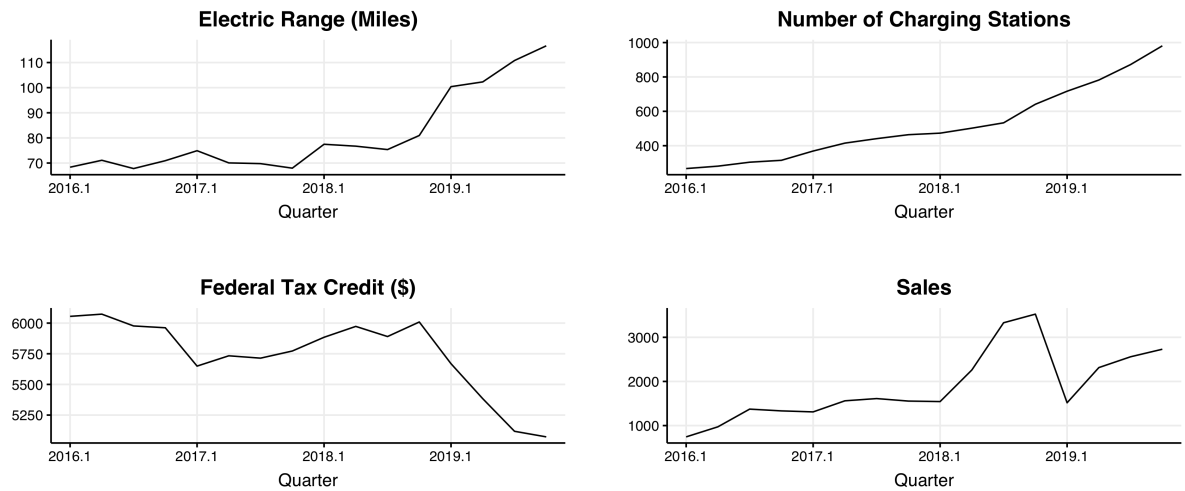

Figure 3: Time Series of Key Variables in the EV Market

purchase a gasoline car. This potential issue might have limited impact on our empirical analysis based on a survey of electric car owners in California conducted by the Center for Sustainable Energy (Johnson et al., 2017). According to this survey, when consumers start searching for a new vehicle, 72 percent of the EV owners stated that they were at least very interested in a PHEV or BEV, and 94 percent of the owners said they at least had some interest in electric cars.

5.2 Consumer Dynamic Demand Model for EV

Our empirical model follows closely to the theoretical framework with a few changes to address specific empirical issues. Most notable is the inclusion of multidimensional unobserved heterogeneity. For a household of type $(g, \mathbf{U}_i \equiv (U_{i1}, U_{i2}))$, assume their expected life-time payoff of buying an EV $j = 1, \ldots, J$ at time t is

$$v_{jt}^{(g)}(\mathbf{U}_i) = \frac{\sum_{k=1}^{K} \delta_j Make_{jk} + \gamma_1' X_{jt}^{ev} + \gamma_2' X_{jt}^{cnty} + \eta_i MPGe_{jt} + \xi_{jt}}{1-\beta} - \alpha_i P_{jt},$$

$$\begin{pmatrix} \eta_i \\ \alpha_i \end{pmatrix} = \begin{pmatrix} \eta^{(1)} \\ \alpha^{(1)} \end{pmatrix} + D_i^{(2)} \begin{pmatrix} \tau_1^{(2)} \\ \tau_2^{(2)} \end{pmatrix} + D_i^{(3)} \begin{pmatrix} \tau_1^{(3)} \\ \tau_2^{(3)} \end{pmatrix} + \mathit{\Omega} \begin{pmatrix} U_{i1} \\ U_{i2} \end{pmatrix},$$

where $\mathit{\Omega}$ is a diagonal matrix defined as

$$\mathit{\Omega} \equiv \begin{pmatrix} \omega_1 & \\ & \omega_2 \end{pmatrix}.$$

For simplicity of reporting the results and for estimation, we report $\frac{\eta^{(1)}}{1-\beta}$, $\frac{\tau^{(2)}}{1-\beta}$, $\frac{\tau^{(2)}}{1-\beta}$, and $\omega_1 \equiv \frac{\tilde{\omega}_1}{1-\beta}$.

Here $Make_{jk}$, for $k = 1, \ldots, K$, is a dummy variable for the make k. If EV j is manufactured by make k, $Make_{jk} = 1$ and $= 0$ otherwise. There are $K = 17$ makes in the sample. Here X^{ev}_{jmt} consists of various product characteristics including horsepower, a dummy variables for BEV, a dummy variable of sport utility vehicle (SUV), the interaction between BEV and SUV indicator variables, the interaction term between the dummy variable of BEV and the gasoline price in county m, and the interaction term between the dummy variable of BEV and the log number of charging stations in county m. The vector X^{cnty}_{mt} consists of county market characteristics including gasoline price and the log number of EV charging stations in the county. The price P_{jmt} is the final price after deducting all federal and state tax incentives. Lastly, we let Nissan Leaf be the base product 1, because it appears in all markets, time periods and groups. Also let Nissan be make 1.

The structural estimation first relies on[21]

$$Y^{dynamic}_{jmt} = \sum_{k=1}^{K} \frac{\delta_{make,k} - \delta_{make,1}}{1-\beta} Make_{jk} + (X^{ev}_{jmt} - X^{ev}_{1mt})' \frac{\gamma_1}{1-\beta} - \alpha^{(1)}(P_{jmt} - P_{1mt}) + \frac{\eta^{(1)}}{1-\beta}[MPGe_{jmt} - MPGe_{1mt}] + \frac{\xi_{jmt} - \xi_{1mt}}{1-\beta}. \quad \text{(EV-Linear-Reg-1)}$$

Note that γ_2 does not appear in and cannot be identified from the above equation. This is because the county market variables do not vary across the products in the county. From this first stage, we obtain the estimates of the make effect $(\delta_{make,2} - \delta_{make,1})/(1-\beta), \ldots, (\delta_{make,K} - \delta_{make,1})/(1-\beta)$, $\gamma_1/(1-\beta)$, $\eta^{(1)}/(1-\beta)$ and price coefficient $\alpha^{(1)}$. In this regression, car price P_{jmt} and the interaction term between BEV dummy variable and the log number of charging stations are endogenous. The IV for price of product j in our estimates is the battery pack cost multiplied by the electric range, which can be viewed as a proxy of the battery size of an EV.[22] The average lithium battery pack price in 2019 was about half of its price in 2016 according to Bloomberg.[23] For this reason we also include the battery pack price and its squared terms as additional instruments. For the number of charging stations

[21] Appendix D collects the details of estimating CCP function with multidimensional unobserved heterogeneity.

[22] Note, the IV employed above accounts for the difference between product 1 and product j—battery pack cost in period t times the difference in electric range between products 1 and j in period t.

[23] See "Battery Pack Prices Fall to an Average of \$132/kWh, But Rising Commodity Prices Start to Bite," BloombergNEF, accessed May 3, 2022, `https://about.bnef.com/blog/battery-pack-prices-fall-to-an-average-of-132-kwh-but-rising-commodity-prices-start-to-bite/`

times BEV, we follow Zhou and Li (2018) and use the log number of public facilities (such as universities, hospitals, supermarkets etc.). This comes from the observation that many EV charging stations locate in the parking lots of these public facilities, which were built long before there was an EV market.

The next step is to use

$$\ln\left[\frac{\sigma^{(g)}_{1mt}(\mathbf{U})}{\sigma^{(g)}_{0mt}(\mathbf{U})}\right] = \frac{\delta_{make,1}}{1-\beta} + X^{ev}_{1mt}{}'\frac{\gamma_1}{1-\beta} + X^{cnty\prime}_{mt}\frac{\gamma_2}{1-\beta} + \left(\frac{\eta^{(1)}}{1-\beta} + \frac{\tau^{(g)}_1}{1-\beta} + \omega_1 U_1\right)(MPGe_{1mt}) \\ - (\alpha^{(1)} + \tau^{(g)}_2 + \omega_2 U_2)P_{1mt} + \frac{\xi_{1mt}}{1-\beta} + \beta\,\mathrm{E}[\bar{V}^{(g)}_{t+1}(U) \mid X_{mt}, P_{mt}, \xi_{mt}]. \quad (14)$$

to recover $\delta_{make,1}$, γ_2 and the discount factor β.

Letting

$$W^{(g)}_{mt}(\mathbf{U}) \equiv \ln\left[\frac{\sigma^{(g)}_{1mt}(\mathbf{U})}{\sigma^{(g)}_{0mt}(\mathbf{U})}\right] - \\ \left[X^{ev}_{1mt}{}'\frac{\gamma_1}{1-\beta} + \left(\frac{\eta^{(1)}}{1-\beta} + \frac{\tau^{(g)}_1}{1-\beta} + \omega_1 U_1\right) MPGe_{1mt} - (\alpha^{(1)} + \tau^{(g)} + \omega_2 U_2)P_{1mt}\right], \quad (15)$$

we have the conclusion following the similar arguments in section 3.4:

$$W^{(g)}_{mt}(\tilde{\mathbf{U}}) = \delta_{make,1} + X^{cnty\prime}_{mt}\frac{\gamma_2}{1-\beta} + \frac{\xi_{1mt}}{1-\beta} + \\ \beta\,\mathrm{E}\left(W^{(g)}_{m,t+1}(\tilde{\mathbf{U}}) - X^{cnty}_{m,t+1}{}'\frac{\gamma_2}{1-\beta} + \ln\sigma^{(g)}_{0,m,t+1}(\tilde{\mathbf{U}}) - \frac{\xi_{1,m,t+1}}{1-\beta} \;\middle|\; X_{mt}, P_{mt}, \xi_{mt}\right), \quad (16)$$

for a fixed $\tilde{\mathbf{U}}$.

To estimate the discount factor β and γ_2, one possible approach is to consider the unconditional expectation again and to have the following conclusion:

$$\mathrm{E}[W^{(g)}_{mt}(\tilde{\mathbf{U}})] = \delta_{make,1} + \beta\,\mathrm{E}[W^{(g)}_{m,t+1}(\tilde{\mathbf{U}}) + \ln\sigma^{(g)}_{0,m,t+1}(\tilde{\mathbf{U}})] - \mathrm{E}[(X^{cnty}_{mt} - \beta X^{cnty}_{m,t+1})']\frac{\gamma_2}{1-\beta} = 0,$$

for a fixed $\tilde{\mathbf{U}}$. Then we can solve β and $\gamma_2/(1-\beta)$ from the above equation by the variation of groups.[24] For this particular application, this approach is infeasible because some of our state

[24] When there are two groups, we have the following formula for β:

$$\beta = \frac{\mathrm{E}[W^{(1)}_{mt}(\tilde{\mathbf{U}})] - \mathrm{E}[W^{(2)}_{mt}(\tilde{\mathbf{U}})]}{\mathrm{E}[W^{(1)}_{m,t+1}(\tilde{\mathbf{U}})] - \mathrm{E}[W^{(2)}_{m,t+1}(\tilde{\mathbf{U}})] + \mathrm{E}[\ln\sigma^{(1)}_{0,m,t+1}(\tilde{\mathbf{U}})] - \mathrm{E}[\ln\sigma^{(2)}_{0,m,t+1}(\tilde{\mathbf{U}})]}.$$

Knowing β, it is straightforward to obtain γ_2 from

$$\mathrm{E}[W^{(g)}_{mt}(\tilde{\mathbf{U}})] - \beta\,\mathrm{E}[W^{(g)}_{m,t+1}(\tilde{\mathbf{U}}) + \ln\sigma^{(g)}_{0,m,t+1}(\tilde{\mathbf{U}})] = \delta_{make,1} + \mathrm{E}[(\beta X^{cnty}_{m,t+1} - X^{cnty}_{mt})']\frac{\gamma_2}{1-\beta},$$

for all g and m.

variables are nonstationary as illustrated by the time series plot in Figure 3. For example, the number of charging stations, which is one element of X_{mt}^{cnty}, is clearly nonstationary. Due to the nonstationarity of the state variables, we cannot use $T^{-1}\sum_{t=1}^{T} X_{mt}^{cnty}$ to estimate $\mathrm{E}(X_{mt}^{cnty})$, which is not well defined for a given county.

As an alternative, we note that in eq. (16), for a fixed $\tilde{\mathbf{U}}$, the random variable $W_{mt}^{(g)}(\tilde{\mathbf{U}})$ is a function of the vector $(X_{mt}, P_{mt}, \xi_{mt})$. So we have a conditional moment condition:

$$\mathrm{E}\Bigg(W_{mt}^{(g)}(\tilde{\mathbf{U}}) - \beta W_{m,t+1}^{(g)}(\tilde{\mathbf{U}}) - \delta_{make,1} - \beta \ln \sigma_{0,m,t+1}^{(g)}(\tilde{\mathbf{U}}) - (X_{mt}^{cnty} - \beta X_{m,t+1}^{cnty})' \frac{\gamma_2}{1-\beta} - \frac{\xi_{1mt} - \beta \xi_{1,m,t+1}}{1-\beta} \Bigg| X_{mt}, P_{mt}, \xi_{mt}\Bigg) = 0.$$

Suppose there is a vector of $X_{mt,IV}$, which are elements or functions of the conditioning variables $(X_{mt}, P_{mt}, \xi_{mt})$ and satisfy

$$\mathrm{E}(X_{mt,IV}\xi_{1mt}) = \mathrm{E}(X_{mt,IV}\xi_{1m,t+1}) = 0.$$

We have the moment conditions of

$$\mathrm{E}\Bigg[X_{mt,IV}\Bigg(W_{mt}^{(g)}(\tilde{\mathbf{U}}) - \beta W_{m,t+1}^{(g)}(\tilde{\mathbf{U}}) - \delta_{make,1} - \beta \ln \sigma_{0,m,t+1}^{(g)}(\tilde{\mathbf{U}}) - (X_{mt}^{cnty} - \beta X_{m,t+1}^{cnty})' \frac{\gamma_2}{1-\beta}\Bigg)\Bigg] = 0.$$

We then can estimate $\delta_{make,1}$, β and γ_2 from the above equation using the moment estimator. In our empirical estimates, we let $\tilde{\mathbf{U}} = 0$ and define $X_{mt,IV}$ as a vector of ones, $gas_{mt}^{cnty} - \beta gas_{m,t+1}^{cnty}$, the number of public charging stations in period t, and the log battery price interacted with a county indicator variable in period t.

5.3 Empirical Results

Given our methodology focuses on estimating unobserved heterogeneity, the most important parameter estimates are those linking to MPGe and price, which recover respective sensitivities $(\eta^{(1)}, \tau_1^{(g)}, \alpha^{(1)}, \tau_2^{(g)})$ for each consumer type as well as Ω which again measures the variation of unobserved consumer heterogeneity within each group.

Our estimates in Table 6 indicate the price sensitivity for those whose income reside in the 90-th percentile or higher is 1.0667, followed by group 2 (income between 75-th and 90-th percentile) whose price sensitivity is $\alpha^{(1)} + \tau^{(g=2)} = 1.1780$. On average, Group 3 consumers whose income is below the 75-th percentile are the most price sensitive consumers with $(1.5775 = 1.0667 + 0.5108)$. Additionally, there is a statistically significant estimate of within group unobserved heterogeneity ($\omega = 0.5491$) in price sensitivity. In order to provide economic meaning to these price estimates, we determine that a temporary unexpected 10% price increase in period 1 for all EVs would lead to a fall in sales by 37.24% in period 1. With respect to $MPGe_{jt}$, we find sensitivities decline as incomes fall. Consumer segment 1 whose income falls above the 90-th percentile is statistically not different from zero. Consumers in group 2 and 3 negatively value MPGe, with sensitivity estimates of (-0.1375) and (0.1143), respectively. There too is statistically significant unobserved heterogeneity with respect to MPGe ($\omega_{mpge} = 0.8387$). The goodness of fit statistic of our first stage nonlinear least squares model is 0.9899.

We include several other observable product characteristics in our estimation besides price and MPGe. We report and discuss these results in terms of their lifetime effect $(\gamma/(1-\beta))$. Most notably, we include make fixed effects in addition to indicator variables for BEV, SUV and the interaction for BEV and SUV. The indicator for BEV is found to be negative and significant. This suggests that after controlling the price difference between BEVs and PHEVs, Americans prefer PHEVs over BEVs. This is possibly due to lack of EV charging stations, especially DC fast charging stations. Because PHEV has a much smaller battery and can drive on gasoline, consumers can simply fully charge their PHEV in their garages over the night, which is impossible for a BEV without fast-charging technology at home. Though BEV now has good electric range, consumers can still feel range anxiety without convenient charging network. This is related to our estimates about the effects of charging stations on the adoption of EV.

We find that demand is driven by the network of charging station in the residing county but only for BEVs (0.4198) as the county effect is insignificant (-0.1654). This effect indicates that consumers value the size and existence of the network of electric charging stations for BEVs. Our estimates also indicate a strong driver of BEV adoption is the current price of gas with an estimate of 0.5187, since the general term for all EVs is statistically insignificant. We add economic meaning to these results by determining the change in total EV sales in period 1 from a temporary unexpected increase in gas prices by 10%. We conclude EV sales would increase by 14.25% in period 1 with BEV sales increasing by 18.21% and PHEV sales by 0%. Additionally, consumers have a preference for larger vehicles (SUV=1.0349) but not for large BEV SUVs as this latter parameter is insignificant. Consumers of EVs also illustrate little utility to horsepower as the parameter estimate is statistically insignificant.

Finally, we discuss our discount factor estimate. The first note of interest is that we are able to precisely estimate it. Additionally, the magnitude (0.9146) is inline with the standard assumption of a monthly discount factor of $\beta = 0.975$. In order to put this estimate into perspective our estimate of $\beta = 0.9146$ equates to consumers valuing utility 15 years in the future—by the 60^{th} quarter or 15^{th} year after the purchase date, the associated flow utility is valued at near zero by the consumer in period $t = 1$ ($\beta^{4\times 15} = 0.0047$). How reasonable is this estimate? It is quite sensible according to the leading industry magazine, *Consumer Reports* reports EV battery packs last 13-17 years, which is based upon a total of $200,000$ miles.[25]

5.4 Policy Simulation

Since 2010, the federal government has incentivized the purchase of new electric vehicles with federal tax credits. To date, many BEV and PHEV buyers have benefited from this program. With consumer preferences estimates in hand, we are able to compare the existing federal tax credit based upon battery size to a new policy that employs MPGe. In order to make a fair comparison of these two policies, we simulate the impact on sales and CO2 emissions for only period 1 in our data, as the new policy must be treated as temporary unexpected change to the existing policy as we have not estimated any state transitions. Moreover, we hold the cost of the new policy equal to the existing cost of the policy in period 1 across all three counties.

Before we discuss the details of our new policy and its results, it is important to understand how the tax credit works, and which cars are eligible. First, the vehicle must be new and its must be purchased rather than leased. The IRS tax credit for 2016 ranged from $\$2,500$ to $\$7,500$ per new electric vehicle purchased. The exact credit amount is based upon the *size* of the EV battery. For instance, the there is a base payment of $\$2,500$ for any battery size. For EVs with larger batteries, the policy provides an additional credit of \$417 per kilowatt hour for batteries that are in excess of 5 kilowatt hours and is capped at $\$5,000$. The total federal tax credit is therefore limited to $\$7,500$.[26] [27]

Our new policy leverages battery efficiency to incentivize EV adoption, rather than battery capacity.[28] For our policy simulation we propose a new federal tax incentive that is a linear function of MPGe: Tax Credit (\$) = MPGe$\times$\$76.53. We also hold the total cost

[25] Ceyhan Cagatay, "How Long Should An Electric Car's Battery Last?", MyEV, accessed May 16 2022, `https://www.myev.com/research/ev-101/how-long-should-an-electric-cars-battery-last`.

[26] https://www.efile.com/electric-vehicle-car-tax-credits/

[27] There are other details of the policy that are omitted as they are not relevant to the current counterfactual simulation (e.g. phasing out of the tax credit over time after a manufacturer reaches a total of 200,000 BEV or PHEV vehicles sold nationally since 2010).

[28] MPGe is the official metric that the EPA uses to measure the efficiency of alternative-fuel.

Table 6: Estimation of EV Demand

Variables	Estimate	SE
Price	1.0667***	(0.4439)
Price*I[Medium Income]	0.1113***	(0.0108)
Price*I[Low Income]	0.5108***	(0.2157)
MPGe	0.3914	(0.6875)
MPGe*I[Medium Income]	−0.5289***	(0.0483)
MPGe*I[Low Income]	−0.5057***	(0.1231)
BEV	−3.1782***	(0.6140)
SUV	1.0349***	(0.3105)
BEV × SUV	1.2284	(0.8555)
Horsepower	0.6744	(0.4707)
Gas Price	6.1302	(4.3634)
Gas Price × BEV	0.5187***	(0.2499)
ln(EV Stations)	0.8350	(9.6096)
ln(EV Stations) × BEV	0.4198***	(0.1040)
ω_1 —MPGe	0.8387***	(0.1512)
ω_2 —Price	0.5491***	(0.0108)
Discount Factor (β)	0.9146***	(0.2250)

Note: Make fixed effects are not reported to save space.

Price is scaled by 1/10,000. MPGe and HP are scaled by 1/100.

Estimates for observed charac. report values $\gamma/(1-\beta)$.

*** 95 percent significance; ** 90 percent significance.

of the proposed policy fixed at the period 1 (2016-Q1) level under the existing federal tax credit policy for the three included counties in our analysis. This latter measure is found by aggregating the tax credits from all EVs sold during Q1 of 2016 across the 3 counties of Washington state.

The new proposed policy rewards more efficient cars in terms of MPGe compared with the existing policy. To put this into perspective, a Tesla Model S in 2016 had an MPGe of 100 and a 95kWh battery and received a federal tax credit of $\$7,500$. Under our proposed policy that tax credit would increase slightly to $\$7,653$. Contrast that with a BMW X5 PHEV with an MPGe of 29 and 9.1 kWh battery in 2016. Under the current policy, the consumer received a $\$4,668$ tax credit whereas under our policy a consumer would only receive $\$2,219$—a roughly 52% reduction in the federal tax credit.

In order to simulate the results of our new temporary policy, we assume the change in the tax rebate was unexpected to the consumers and is believed to be temporary (one period). We leverage the structural parameter estimates along with the series multinomial logit estimates of the CCP (ρ_{jmt1}, $\boldsymbol{\rho}_{jmt2}$, and $\boldsymbol{\rho}_{jmt3}$) to simulate new choice probabilities under the new proposed tax policy. (See eq. (D.3) in Appendix D for the definition of series approximation of CCP with multidimensional unobserved heterogeneity.) We show below that only ρ_{jmt1} and $\boldsymbol{\rho}_{jmt2}$ change from their initial estimates. Note, policies that are not temporary cannot leverage the series multinomial logit CCP parameters to simulate new CCPs as such policies would require a change in the beliefs and/or evolution of the state space.

Proposition 4 (Transformation of the series multinomial logit parameters for a temporary and unexpected policy change)**.** *Define $\tilde{\rho}_{jmt1}$, $\tilde{\boldsymbol{\rho}}_{jmt2}$ and $\tilde{\boldsymbol{\rho}}_{jmt3}$ as CCP series multinomial logit parameters under the new policy. We have*

$$\tilde{\rho}_{jmt1} = \rho_{jmt1} - \alpha^{(1)}\Delta P_{jmt}, \quad \tilde{\boldsymbol{\rho}}_{jmt2} \equiv \begin{pmatrix} \tilde{\rho}_{jmt2,1} \\ \tilde{\rho}_{jmt2,2} \end{pmatrix} = \begin{pmatrix} \rho_{jmt2,1} \\ \rho_{jmt2,2} - \Delta P_{jmt}\omega_2 \end{pmatrix}, \quad \tilde{\boldsymbol{\rho}}_{jmt3} = \boldsymbol{\rho}_{jmt3},$$

where ΔP_{jmt} is the price difference between the new and old policy for product j in period t and county m.

The results of our proposed policy are reported in Table 7, and highlight a sizable reduction in CO2 emissions (13.86%) with a (-0.81%) reduction in the total number of EVs sold (7 car). This indicates under the new policy consumers substituted away from PHEV vehicles to more efficient BEV vehicle (15 more BEVs were sold). In Q1 of 2016, the original policy sold 889 units whereas the the new policy sold 7 fewer cars at 882 cars. Given our policy emphasizes battery efficiency rather than capacity, we analyze the policy effect on CO2

Table 7: Policy Simulation Results

	Existing Policy (kWh)	Proposed Policy (MPGe)	% change
Sales (units)	889	882	-0.81%
BEV Sales (units)	705	720	2.07%
PHEV Sales (units)	184	162	-11.90%
CO2 (tons)	362.82	312.54	-13.86%
Total Cost ($)	6,101,169.10	6,101,140.61	4.6696e-04%

emissions. Leveraging data from the US Department of Energy we are able to determine the change in CO2 due to the change in tax policy. It is important to note that the results that pertain to CO2 emissions are only from the use of EVs and do not include any emissions from the generation of electricity to charge the EVs. We determine our proposed policy reduces CO2 emissions by 13.86% or 50.28 tons of CO2 from 362.82 tons to 312.54 tons. In summary, we show a cost neutral temporary policy that emphasizes battery efficiency over storage can reduce CO2 emissions by incentivizing consumers to switch from PHEVs to BEVs, and in particular lighter more efficient BEVs.

6 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[29]). 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.[30] As a result, these prospects make our data requirement (and methodology) more desirable in the future.

[29] See for example: "Manufacturers Place Orders When Economy Improves," United States Census Bureau, accessed December 24, 2020, https://www.census.gov/library/stories/2018/07/manufacturers-durable-goods.html

[30] 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

An, Yonghong, Yingyao Hu, and Ruli Xiao. 2020. “Dynamic Decisions under Subjective Expectations: A Structural Analysis.” *Journal of Econometrics*, Forthcoming.

Arcidiacono, P., and R.A. Miller. 2011. “Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity.” *Econometrica*, 79(6): 1823–1867.

Arcidiacono, Peter, and Robert Miller. 2018. “Identifying Dynamic Discrete Choice Models off Short Panels.” Duke University Memo.

Bajari, Patrick, Chenghuan Sean Chu, Denis Nekipelov, and Minjung Park. 2016. “Identification and Semiparametric Estimation of a Finite Horizon Dynamic Discrete Choice Model with a Terminating Action.” *Quantitative Marketing and Economics*, 14(4): 271–323.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. “Automobile Prices in Market Equilibrium.” *Econometrica*, 63(4): 841–890.

Berry, Steven T. 1994. “Estimating Discrete-Choice Models of Product Differentiation.” *The RAND Journal of Economics*, 25(2): 242–262.

Berry, Steven T, and Philip A Haile. 2014. “Identification in Differentiated Products Markets Using Market Level Data.” *Econometrica*, 82(5): 1749–1797.

Berry, Steven T, and Philip A Haile. 2020. “Nonparametric Identification of Differentiated Products Demand Using Micro Data.” Yale University Memo.

Daljord, Øystein, Denis Nekipelov, and Minjung Park. 2018. “A Simple and Robust Estimator for Discount Factors in Optimal Stopping Dynamic Discrete Choice Models.” University of Chicago Memo.

Derdenger, Timothy, and Vineet Kumar. 2019. “Estimating Dynamic Discrete Choice Models with Aggregate Data: Small Sample Properties of the Inclusive Value Approximation.” *Quantitative Marking and Economics*, 17: 359–384.

Goettler, Ronald L, and Brett R Gordon. 2011. “Does AMD Spur Intel to Innovate More?” *Journal of Political Economy*, 119(6): 1141–1200.

Gowrisankaran, Gautam, and Marc Rysman. 2012. "Dynamics of Consumer Demand for New Durable Goods." *Journal of Political Economy*, 120(6): 1173–1219.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica*, 71(4): 1161–1189.

Hotz, V Joseph, and Robert A Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *The Review of Economic Studies*, 60(3): 497–529.

Hu, Yingyao, and Matthew Shum. 2012. "Nonparametric Identification of Dynamic Models with Unobserved State Variables." *Journal of Econometrics*, 171(1): 32–44.

Hu, Yingyao, Matthew Shum, Wei Tan, and Ruli Xiao. 2017. "A Simple Estimator for Dynamic Models with Serially Correlated Unobservables." *Journal of Econometric Methods*, 6(1).

Johnson, Clair, Brett Williams, Carlos Hsu, and John Anderson. 2017. "The Clean Vehicle Rebate Project: Summary Documentation of the Electric Vehicle Consumer Survey, 2013–2015 Edition." Center for Sustainable Energy, San Diego CA.

Kasahara, Hiroyuki, and Katsumi Shimotsu. 2009. "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices." *Econometrica*, 77(1): 135–175.

Magnac, Thierry, and David Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes." *Econometrica*, 70(2): 801–816.

Melnikov, Oleg. 2013. "Demand for Differentiated Durable Products: The Case of the US Computer Printer Market." *Economic Inquiry*, 51(2): 1277–1298.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2): 307–342.

Nevo, Aviv. 2011. "Empirical Models of Consumer Behavior." *Annual Review of Economics*, 3(1): 51–75.

Norets, Andriy. 2009. "Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables." *Econometrica*, 77(5): 1665–1682.

Plötz, Patrick, Cornelius Moll, Georg Bieker, Peter Mock, and Yaoming Li. 2020. "Real-World Usage of Plug-In Hybrid Electric Vehilces Fuel Consumption, Electric Drivign, and CO2 Emissions." The International Council of Clean Transportation White Paper.

Rust, John. 1994. "Structural Estimation of Markov Decision Processes." In *Handbook of Econometrics*. Vol. 4, Chapter 51, 3081–3143. Elsevier.

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

Zhou, Yiyi, and Shanjun Li. 2018. "Technology Adoption and Critical Mass: The Case of the US Electric Vehicle Market." *The Journal of Industrial Economics*, 66(2): 423–480.

(For Online Publication)

Online Appendix for

Discrete Choice Demand Estimation Leveraging Overlapping Groups of Consumers

Cheng Chou[1] Tim Derdenger[2]

July 2022

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

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

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

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

The above conditioning variables just restrict the population to be the remaining consumers after $t-1$ periods. Because all the remaining consumers in period t face the same the product

[1]Cheng Chou: School of Business, University of Leicester, UK. Email: cchou@le.ac.uk.

[2]Tim Derdenger: Tepper School of Business, Carnegie Mellon University `email`:derdenge@andrew.cmu.edu.

state variables Z_t, we also have the following conditional independence,

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

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

$$\begin{aligned} f_2^{(g)}(u) &= \Pr(U_i = u \mid D_i^{(g)} = 1, A_{i1} = 0, Z_1) \\ &= \frac{f(U_i = u \mid D_i^{(g)} = 1, Z_1) f(A_{i1} = 0 \mid U_i = u, D_i^{(g)} = 1, Z_1)}{f(A_{i1} = 0 \mid D_i^{(g)} = 1, Z_1)} \\ &= \phi(u) \times \frac{\sigma_{01}^{(g)}(u)}{S_{01}^{(g)}}. \end{aligned}$$

The third line used $f(U_i = u \mid D_i^{(g)} = 1, Z_1) = f(U_i = u \mid D_i^{(g)} = 1)$ 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

$$\begin{aligned} f_{t+1}^{(g)}(u) &= \Pr(U_i = u \mid D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{i,t-1} = 0, A_{it} = 0, Z_1, \ldots, Z_t) \\ &= \Pr(U_i = u \mid D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t) \\ &\quad \times \frac{f(A_{it} = 0 \mid U_i = u, D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t)}{f(A_{it} = 0 \mid D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t)} \\ &= f_t^{(g)}(u) \times \frac{f(A_{it} = 0 \mid U_i = u, D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t)}{f(A_{it} = 0 \mid D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{it-1} = 0, Z_1, \ldots, Z_t)} \quad \text{by eq. (A.2)} \\ &= f_t^{(g)}(u) \frac{\sigma_{0t}^{(g)}(u)}{S_{0t}^{(g)}} = \phi(u) \times \prod_{s=1}^{t} \frac{\sigma_{0s}^{(g)}(u)}{S_{0s}^{(g)}}. \end{aligned}$$

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

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

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,

$$\begin{aligned}\pi_2^{(g)} &= \Pr(D_i^{(g)} = 1 \mid A_{i1} = 0, Z_1) \\ &= \frac{\Pr(D_i^{(g)} = 1 \mid Z_1) \Pr(A_{i1} = 0 \mid D_i^{(g)} = 1, Z_1)}{\Pr(A_{i1} = 0, Z_1)} \\ &= \pi_1^{(g)} \times \frac{S_{01}^{(g)}}{S_{01}}.\end{aligned}$$

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

$$\begin{aligned}\pi_{t+1}^{(g)} &= \Pr(D_i^{(g)} = 1 \mid A_{i1} = 0, \ldots, A_{i,t-1} = 0, A_{it} = 0, Z_1, \ldots, Z_t) \\ &= \Pr(D_i^{(g)} = 1 \mid A_{i1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_{t-1}, Z_t) \\ &\quad \times \frac{\Pr(A_{it} = 0 \mid D_i^{(g)} = 1, A_{i1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)}{\Pr(A_{it} = 0 \mid A_{i1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_t)} \\ &= \pi_t^{(g)} \frac{S_{0t}^{(g)}}{S_{0t}}.\end{aligned}$$

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 $\Pr(D_i^{(g)} = 1 \mid A_{i1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_{t-1}, Z_t) = \Pr(D_i^{(g)} = 1 \mid A_{i1} = 0, \ldots, A_{i,t-1} = 0, Z_1, \ldots, Z_{t-1}) = \pi_t^{(g)}$. This completes the proof. ∎

Proof of Proposition 4. For expositional simplicity, we omit the market index m in this proof. Also note that $U = (U_1, U_2)'$ involves two-dimensional unobserved heterogeneity, and U_1 and U_2 are associated with MPGe and price, respectively. Correspondingly, $\tau^{(g)} = (\tau_1^{(g)}, \tau_2^{(g)})'$. We begin by first specifying the log market share ratio under each policy where the current policy follows from eq. (14), and the new policy takes the form of

$$\begin{aligned}\ln\left[\frac{\tilde{\sigma}_{jmt}^{(g)}(\mathbf{U})}{\tilde{\sigma}_{0mt}^{(g)}(\mathbf{U})}\right] &= \frac{\delta_j}{1-\beta} + X_{jt}'\frac{\gamma_i}{1-\beta} - (\alpha^{(1)} + \tau_2^{(g)} + \omega_2 U_2)\tilde{P}_{jt} + \frac{\xi_{jt}}{1-\beta} \\ &\quad - \beta\,\mathrm{E}[\bar{V}_{t+1}^{(g)}(U) \mid X_{mt}, P_{mt}, \xi_{mt}], \quad \text{(A.4)}\end{aligned}$$

where $\tilde{P}_{jt}$ is the new price of the EV vehicle with the proposed tax credit. Its series representation is that

$$\ln\left[\frac{\tilde{\sigma}_{jmt}^{(g)}(\mathbf{U})}{\tilde{\sigma}_{0mt}^{(g)}(\mathbf{U})}\right] = \tilde{\rho}_{jt1} + \tilde{\rho}_{jt2}'\Omega^{-1}\Big(\Omega U + \tau^{(g)}\Big) + \Big(\Omega U + \tau^{(g)}\Big)'\Omega^{-1}\tilde{\rho}_{jt3}\Omega^{-1}\Big(\Omega U + \tau^{(g)}\Big),$$

with unknown coefficients $\tilde{\rho}_{jt1}, \tilde{\rho}_{jt2}$, and $\tilde{\rho}_{jt3}$. Ω is a diagonal matrix whose diagonal elements are ω_1 and ω_2. The objective is to solve for $\tilde{\rho}_{jt1}, \tilde{\rho}_{jt2}$, and $\tilde{\rho}_{jt3}$. Define $\tilde{P}_{jt} = P_{jt} + \Delta P_{jt}$. First note that by eq. (14), we have

$$\begin{aligned}
\ln\left[\frac{\tilde{\sigma}^{(g)}_{jmt}(\mathbf{U})}{\tilde{\sigma}^{(g)}_{0mt}(\mathbf{U})}\right] &= \ln\left[\frac{\sigma^{(g)}_{jmt}(\mathbf{U})}{\sigma^{(g)}_{0mt}(\mathbf{U})}\right] - (\alpha^{(1)} + \tau_2^{(g)} + \omega_2 U_2)\Delta P_{jt} \\
&= \rho_{jt1} + \rho'_{jt2}\Omega^{-1}\Big(\Omega U + \tau^{(g)}\Big) + \Big(\Omega U + \tau^{(g)}\Big)'\Omega^{-1}\rho_{jt3}\Omega^{-1}\Big(\Omega U + \tau^{(g)}\Big) \\
&\quad - (\alpha^{(1)} + \tau_2^{(g)} + \omega_2 U_2)\Delta P_{jt} \\
&= \rho_{jt1} - (\alpha^{(1)} + \tau_2^{(g)})\Delta P_{jt} + \rho'_{jt2}\Omega^{-1}\Big(\Omega U + \tau^{(g)}\Big) + \Big(\Omega U + \tau^{(g)}\Big)'\Omega^{-1}\rho_{jt3}\Omega^{-1}\Big(\Omega U + \tau^{(g)}\Big) \\
&\quad - \omega_2 U_2 \Delta P_{jt}.
\end{aligned}$$

Note that ρ_{jt2} is a two-dimensional vector:

$$\rho_{jt2} = \begin{pmatrix} \rho_{jt2,1} \\ \rho_{jt2,2} \end{pmatrix} \qquad \text{and} \qquad \rho'_{jt2}\Omega^{-1} = (\rho_{jt2,1}\omega_1^{-1}, \rho_{jt2,2}\omega_2^{-1}),$$

where $\rho_{jt2,1}$ is associated with MPGe, and $\rho_{jt2,2}$ is with price. Next, we write $\omega_2 U_2 \Delta P_{jt} = (\omega_2 U_2 + \tau_2^{(g)})\Delta P_{jt} - \tau_2^{(g)}\Delta P_{jt}$ so that

$$\ln\left[\frac{\tilde{\sigma}^{(g)}_{jmt}(\mathbf{U})}{\tilde{\sigma}^{(g)}_{0mt}(\mathbf{U})}\right] = \rho_{jt1} - \alpha^{(1)}\Delta P_{jt} + \begin{pmatrix} \rho_{jt2,1} \\ \rho_{jt2,2} - \Delta P_{jt}\omega_2 \end{pmatrix}\Omega^{-1}\Big(\Omega U + \tau^{(g)}\Big) + \frac{\rho_{jt3}}{\omega^2}\Big(\Omega U + \tau^{(g)}\Big)^2.$$

Thus,

$$\tilde{\rho}_{jt1} = \rho_{jt1} - \alpha^{(1)}\Delta P_{jt}, \qquad \tilde{\rho}_{jt2} \equiv \begin{pmatrix} \tilde{\rho}_{jt2,1} \\ \tilde{\rho}_{jt2,2} \end{pmatrix} = \begin{pmatrix} \rho_{jt2,1} \\ \rho_{jt2,2} - \Delta P_{jt}\omega_2 \end{pmatrix}, \qquad \tilde{\rho}_{jt3} = \rho_{jt3}.$$

■

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 $\theta_1 \equiv (\boldsymbol{\tau}, \omega, \boldsymbol{\rho}, (\delta_2 - \delta_1)/(1-\beta), \ldots, (\delta_J - \delta_1)/(1-\beta), \gamma/(1-\beta), \alpha^{(1)})$, and $\theta_2 \equiv (\beta, \delta_1)$. We decompose θ into these two parts, because the estimation of θ_2 relies on the estimation of θ_1. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ denote the estimator of θ_1 and θ_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 θ_2. In the estimation of θ_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

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

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:

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

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 $\sigma_{0t}^{(g)}(0;\hat{\theta}_1)$. We have

$$\begin{aligned} W_t^{(g)}(0;\hat{\theta}_1) &= \ln\left[\frac{\hat{\sigma}_{1t}^{(g)}(0)}{\hat{\sigma}_{0t}^{(g)}(0)}\right] - X_{1t}'\widehat{\left(\frac{\gamma}{1-\beta}\right)} + (\hat{\alpha}^{(1)} + \hat{\tau}^{(g)})P_{1t} \\ &= \ln\left[\frac{\hat{\sigma}_{1t}^{(1)}(\hat{\tau}^{(g)}/\hat{\omega})}{\hat{\sigma}_{0t}^{(1)}(\hat{\tau}^{(g)}/\hat{\omega})}\right] - X_{1t}'\widehat{\left(\frac{\gamma}{1-\beta}\right)} + (\hat{\alpha}^{(1)} + \hat{\tau}^{(g)})P_{1t} \\ &= \left[\hat{\rho}_{jt1} + \hat{\rho}_{jt2}\frac{\hat{\tau}^{(g)}}{\hat{\omega}} + \hat{\rho}_{jt3}\left(\frac{\hat{\tau}^{(g)}}{\hat{\omega}}\right)^2\right] - X_{1t}'\widehat{\left(\frac{\gamma}{1-\beta}\right)} + (\hat{\alpha}^{(1)} + \hat{\tau}^{(g)})P_{1t}, \end{aligned}$$

and $\sigma_{0t}^{(g)}(0;\hat{\theta}_1) = \sigma_{0t}^{(1)}(\hat{\tau}^{(g)}/\hat{\omega};\hat{\theta}_1)$ has the series logit expression. Both $W_t^{(g)}(0;\hat{\theta}_1)$ and $\sigma_{0t}^{(g)}(0;\hat{\theta}_1)$ 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 θ_1 by

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

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

where

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

and

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

Here X_{jt}^{IV} is a vector of IV in eq. (Linear-Reg-1) satisfying $\mathrm{E}[X_{jt}^{IV}(\xi_{jt} - \xi_{1t})] = 0$. 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 Σ_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 ρ_t:

$$\rho_{jt2} = -\omega P_{jt} \qquad \text{and} \qquad \rho_{jt3} = 0. \tag{C.1}$$

It can be shown (see proof blow) that the above constraints also imply Proposition 1, i.e.

$$\rho_{jt1} = \delta_j + \gamma' X_{jt} - \alpha^{(1)} P_{jt} + \xi_{jt}.$$

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

$$\ln\left[\frac{\sigma_{jmt}^{(g)}(U)}{\sigma_{0mt}^{(g)}(U)}\right] = \delta_j + \gamma' X_{jt} - (\alpha^{(1)} + \tau^{(g)}) P_{jt} + \xi_{jt} - \omega U P_{jt}. \tag{C.2}$$

From the above equation, we have

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

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

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

which implies an alternative expression of the same derivative:

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

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

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

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

C.2 Constraints about CCP from Dynamic Model

We claim that the dynamic model implies the following constraints about the parameters ρ_t in the CCP function:

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

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

$$(\rho_{1t1}, \ldots, \rho_{Jt1})', \quad \rho_{1t2}, \quad \rho_{1t3}, \quad \omega, \quad \boldsymbol{\tau}, \qquad \text{for } t = 1, \ldots, T.$$

The degree of freedom of the NLS problem is $JGT - JT - 2T - G$, where JGT is the number of observations, JT results from ρ_{jt1} for each $j = 1, \ldots, J$ and $t = 1, \ldots, T$, $2T$ comes from (ρ_{1t2}, ρ_{1t3}) for each $t = 1, \ldots, T$, and G refers to one ω plus $(G-1) \times 1$ vector $\boldsymbol{\tau}$. 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.

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

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

By the series logit approximation, we have

$$\ln\left[\frac{\sigma_{jt}^{(g)}(U)}{\sigma_{1t}^{(g)}(U)}\right] = (\rho_{jt1} - \rho_{1t1}) + \left[\left(\frac{\rho_{jt2}}{\omega}\right) - \left(\frac{\rho_{1t2}}{\omega}\right)\right](\omega U + \tau^{(g)}) + \left[\left(\frac{\rho_{jt3}}{\omega^2}\right) - \left(\frac{\rho_{1t3}}{\omega^2}\right)\right](\omega U + \tau^{(g)})^2, \tag{C.4}$$

which implies

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

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

A useful conclusion is the following. Applying the constraints about $\boldsymbol{\rho}$ to eq. (C.4) for the first group, $g = 1$, we have

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

The dependent variable Y_{jt} of eq. (Linear-Reg-1), whose definition is copied below, has a simple expression,

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

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 4.

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 $\boldsymbol{\tau}_{\text{init}} = (\tau_{\text{init}}^{(2)}, \ldots, \tau_{\text{init}}^{(G)})'$, $\boldsymbol{\rho}_{\text{init}}$, and ω_{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 $\boldsymbol{\tau}_{\text{init}}$ by running 2SLS for the following linear regression,

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

To see the rationale, recall the identity

$$S_{jt}^{(g)} = \mathrm{E}[\sigma_{jt}^{(g)}(U^*)\mathit{\Gamma}_t^{(g)}(U^*)], \qquad U^* \sim \mathcal{N}(0,1),$$

and consider the first order Taylor expansion of CCP function $\sigma_{jt}^{(g)}(U^*)\mathit{\Gamma}_t^{(g)}(U^*)$ at 0, which is the mean of $U^* \sim \mathcal{N}(0,1)$, for each group g. We have

$$\begin{aligned} S_{jt}^{(g)} &= \mathrm{E}[\sigma_{jt}^{(g)}(U^*)\mathit{\Gamma}_t^{(g)}(U^*)] \\ &\approx \sigma_{jt}^{(g)}(0)\mathit{\Gamma}_t^{(g)}(0) + \left(\frac{\mathrm{d}\sigma_{jt}^{(g)}(U^*)\mathit{\Gamma}_t^{(g)}(U^*)}{\mathrm{d}U^*}\right)_{U^*=0} \mathrm{E}(U^* - 0) \\ &= \sigma_{jt}^{(g)}(0)\mathit{\Gamma}_t^{(g)}(0). \end{aligned}$$

The first order Taylor expansion leads to $S_{jt}^{(g)} \approx \sigma_{jt}^{(g)}(0)\Gamma_t^{(g)}(0)$. Thus,

$$\frac{S_{jt}^{(g)}}{S_{1t}^{(g)}} \approx \frac{\sigma_{jt}^{(g)}(0)}{\sigma_{1t}^{(g)}(0)}.$$

The application of this conclusion to eq. (Linear-Reg-1) when $U = 0$ gives rise to the stated regression.

In the *second step*, we find $\rho_{jt1,\text{init}}$ for all $j = 1, \ldots, J$, and $(\rho_{1t2}/\omega)_{\text{init}}, (\rho_{1t3}/\omega^2)_{\text{init}}$ by running OLS for the following linear regression,

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

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

$$\begin{aligned}\ln\left[\frac{\sigma_{jt}^{(g)}(0)}{\sigma_{0t}^{(g)}(0)}\right] = \ln\left[\frac{\sigma_{jt}^{(1)}(\tau^{(g)}/\omega;\rho_t)}{\sigma_{0t}^{(1)}(\tau^{(g)}/\omega;\rho_t)}\right] &= \rho_{jt1} + \left(\frac{\rho_{jt2}}{\omega}\right)\tau^{(g)} + \left(\frac{\rho_{jt3}}{\omega^2}\right)(\tau^{(g)})^2 \\ &= \rho_{jt1} + \left(\frac{\rho_{1t2}}{\omega}\right)\tau^{(g)} - (P_{jt} - P_{1t})\tau^{(g)} + \left(\frac{\rho_{1t3}}{\omega^2}\right)(\tau^{(g)})^2.\end{aligned}$$

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

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

we have the stated regression.[3]

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.

[3]We do not have a clever initial value of ω_{init}. In our simulation, we varied the initial value ω_{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 γ_i associated with product characteristics 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

$$\begin{pmatrix} \gamma_i \\ \alpha_i \end{pmatrix} = \begin{pmatrix} \gamma^{(1)} \\ \alpha^{(1)} \end{pmatrix} + D_i^{(2)} \begin{pmatrix} \boldsymbol{\tau}_1^{(2)} \\ \tau_2^{(2)} \end{pmatrix} + \cdots + D_i^{(G)} \begin{pmatrix} \boldsymbol{\tau}_1^{(G)} \\ \tau_2^{(G)} \end{pmatrix} + \mathit{\Omega} \begin{pmatrix} \mathbf{U}_{i1} \\ U_{i2} \end{pmatrix}, \tag{D.1}$$

where $\mathit{\Omega}$ is a diagonal matrix,

$$\mathit{\Omega} \equiv \begin{pmatrix} \mathit{\Omega}_1 & \\ & \omega_2 \end{pmatrix}.$$

The diagonal elements $\mathit{\Omega}_1$, which is also a diagonal matrix, and ω_2 determine the within group variation of γ_i and α_i, respectively. Below, let $\boldsymbol{\tau}^{(g)} \equiv (\boldsymbol{\tau}_1^{(g)\prime}, \tau_2^{(g)})'$ and let $\mathbf{U}_i \equiv (\mathbf{U}_{i1}', U_{i2})'$. Use q to denote the dimension of $(X_{jt}', P_{jt})'$. Again, let $\boldsymbol{\tau}^{(1)} \equiv \mathbf{0}$. Let $\boldsymbol{\phi}(\mathbf{U})$ denote the PDF of the multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

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

$$\sigma_{jt}^{(g)}(\mathbf{U}) = \frac{\exp[\delta_j + (\gamma^{(1)} + \boldsymbol{\tau}_1^{(g)} + \mathit{\Omega}_1 \mathbf{U}_1)' X_{jt} - (\alpha^{(1)} + \tau_2^{(g)} + \omega_2 U_2) P_{jt} + \xi_{jt}]}{1 + \sum_{k=1}^{J} \exp[\delta_k + (\gamma^{(1)} + \boldsymbol{\tau}_1^{(g)} + \mathit{\Omega}_1 \mathbf{U}_1)' X_{kt} - (\alpha^{(1)} + \tau_2^{(g)} + \omega_2 U_2) P_{kt} + \xi_{kt}]}.$$

This gives rise to

$$\ln\left[\frac{\sigma_{jt}^{(g)}(\mathbf{U})}{\sigma_{0t}^{(g)}(\mathbf{U})}\right] = \delta_j + (\gamma^{(1)} + \boldsymbol{\tau}_1^{(g)})' X_{jt} - (\alpha^{(1)} + \tau_2^{(g)}) P_{jt} + \xi_{jt} + (\mathit{\Omega}_1 \mathbf{U}_1)' X_{jt} - \omega_2 U_2 P_{jt}. \tag{D.2}$$

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

[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 γ_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.

We claim that we have been able to estimate CCP $\sigma_{jt}^{(g)}(\mathbf{U})$, the dispersion coefficients $\mathit{\Omega}$ and $\boldsymbol{\tau}$, then δ_j, $\gamma^{(1)}$ and $\alpha^{(1)}$ can be estimated by applying 2SLS to the following equation,

$$\int \ln\left[\frac{\sigma_{jt}^{(1)}(\mathbf{U})}{\sigma_{0t}^{(1)}(\mathbf{U})}\right] f_t^{(1)}(\mathbf{U})\ \mathrm{d}\,\mathbf{U} + \int ((\mathit{\Omega}_1\mathbf{U}_1)' X_{jt} - \omega_2 \mathbf{U}_2 P_{jt}) f_t^{(1)}(\mathbf{U})\ \mathrm{d}\,\mathbf{U} = \\ \delta_j + \gamma^{(1)\prime} X_{jt} - \alpha^{(1)} P_{jt} + \xi_{jt}.$$

We now start addressing the estimation of CCP function $\sigma_{jt}^{(g)}(\mathbf{U})$ 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

$$\sigma_{jt}^{(g)}(\mathbf{U}) = \sigma_{jt}^{(1)}(\mathbf{U} + \mathit{\Omega}^{-1}\boldsymbol{\tau}^{(g)}).$$

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

$$\sigma_{jt}^{(1)}(\mathbf{U}; \rho_t) \equiv L_j\Big(R_K(\mathbf{U}; \rho_{1t}), \ldots, R_K(\mathbf{U}; \rho_{Jt})\Big), \tag{D.3a}$$

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 $R_K(\mathbf{U}; \rho_{jt})$ is a polynomial function of $\mathbf{U}$,

$$R_K(\mathbf{U}; \rho_{jt}) = \rho_{jt1} + \boldsymbol{\rho}_{jt2}'\mathbf{U} + \mathbf{U}'\boldsymbol{\rho}_{jt3}\mathbf{U} + \cdots, \tag{D.3b}$$

where $\boldsymbol{\rho}_{jt2}$ is a $q \times 1$ vector, and $\boldsymbol{\rho}_{jt3}$ is a $q \times q$ triangular 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

$$\begin{aligned} S_{jt}^{(g)} &= \int \sigma_{jt}^{(1)}(\mathbf{u} + \mathit{\Omega}^{-1}\boldsymbol{\tau}^{(g)}; \rho_t) f_t^{(g)}(\mathbf{u})\ \mathrm{d}\,\mathbf{u} \\ &= \int \sigma_{jt}^{(1)}(\mathbf{u} + \mathit{\Omega}^{-1}\boldsymbol{\tau}^{(g)}; \rho_t) \mathit{\Gamma}_t^{(g)}(\mathbf{u}) \boldsymbol{\phi}(\mathbf{u})\ \mathrm{d}\,\mathbf{u}. \end{aligned}$$

We then use Gauss–Hermite to approximate the above expectation, and estimate $\boldsymbol{\tau}, \mathit{\Omega}, \boldsymbol{\rho}$ 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

$$\frac{\mathrm{d} \ln[\sigma_{jt}^{(g)}(\mathbf{U})/\sigma_{0t}^{(g)}(\mathbf{U})]}{\mathrm{d}\mathbf{U}} = \begin{pmatrix} \mathit{\Omega}_1 X_{jt} \\ -\omega_2 P_{jt} \end{pmatrix},$$

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

$$\boldsymbol{\rho}_{jt2} = \begin{pmatrix} \Omega_1 X_{jt} \\ -\omega_2 P_{jt} \end{pmatrix}, \quad \boldsymbol{\rho}_{jt3} = \boldsymbol{\rho}_{jt4} = \cdots = 0. \text{ (Constraints: Static Mult-Dim Heterogeneity)}$$

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 ρ_{jt1} for each j and t. The only new parameters involved here are Ω and additional $\boldsymbol{\tau}$. The degree of freedom of the NLS problem now is $GJT-(JT+qG)$ (recall q is the dimension of $(X'_{jt}, P_{jt})'$, and qG comes from the unknown $\boldsymbol{\tau}$ and Ω). 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. Under the same specification of (γ'_i, α_i) in Equation (D.1), the expected lifetime payoff of purchasing product j at time t becomes

$$v_{jt}^{(g)}(\mathbf{U}_i) = \frac{\delta_j + \gamma'_i X_{jt} + \xi_{jt}}{1-\beta} - \alpha_i P_{jt}, \qquad j = 1, \ldots, J.$$

Correspondingly, the CCP of type-$(g, \mathbf{U})$ is

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

Because $v_{jt}^{(g)}(\mathbf{U}) = v_{jt}^{(1)}(\mathbf{U} + \Omega^{-1}\boldsymbol{\tau}^{(g)})$, we still have shifting formula $\sigma_{jt}^{(g)}(\mathbf{U}) = \sigma_{jt}^{(1)}(\mathbf{U} + \Omega^{-1}\boldsymbol{\tau}^{(g)})$. 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 ρ_t result from the comparison between the derivatives of the log CCP ratio with respect to $\mathbf{U}$ in the structural demand model and the same derivatives in the series approximation. Particularly, we have

$$\ln\left[\frac{\sigma_{jt}^{(g)}(\mathbf{U})}{\sigma_{1t}^{(g)}(\mathbf{U})}\right] = v_{jt}^{(g)}(\mathbf{U}) - v_{1t}^{(g)}(\mathbf{U})$$
$$= \frac{\delta_j - \delta_1}{1-\beta} + \frac{(\gamma^{(1)} + \boldsymbol{\tau}_1^{(g)})'}{1-\beta}(X_{jt} - X_{1t}) - (\alpha^{(1)} + \tau_2^{(g)})(P_{jt} - P_{1t}) + \frac{\xi_{jt} - \xi_{1t}}{1-\beta} +$$
$$((1-\beta)^{-1}\Omega_1 \mathbf{U}_1)'(X_{jt} - X_{1t}) - \omega_2 U_2 (P_{jt} - P_{1t}).$$

This gives rise to

$$\frac{\mathrm{d}\ln[\sigma_{jt}^{(g)}(\mathbf{U})/\sigma_{1t}^{(g)}(\mathbf{U})]}{\mathrm{d}\mathbf{U}} = \begin{pmatrix} (1-\beta)^{-1}\Omega_1(X_{jt} - X_{1t}) \\ -\omega_2(P_{jt} - P_{1t}) \end{pmatrix},$$

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

$$\boldsymbol{\rho}_{jt2} - \boldsymbol{\rho}_{1t2} = \begin{pmatrix} (1-\beta)^{-1}\Omega_1(X_{jt} - X_{1t}) \\ -\omega_2(P_{jt} - P_{1t} \end{pmatrix}, \quad (\boldsymbol{\rho}_{jt3} - \boldsymbol{\rho}_{1t3}) = (\boldsymbol{\rho}_{jt4} - \boldsymbol{\rho}_{1t4}) = \cdots = 0.$$
(Constraints: Dynamic Mult-Dim Heterogeneity)

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

$$(\rho_{1t1}, \ldots, \rho_{Jt1})', \quad \boldsymbol{\rho}_{1t2}, \quad \boldsymbol{\rho}_{1t3}, \quad (1-\beta)^{-1}\Omega_1, \quad \omega_2, \quad \boldsymbol{\tau}, \qquad \text{for all } t.$$

Note that in this step of CCP estimation, we are only able to estimate $(1-\beta)^{-1}\Omega_1$ as a whole and cannot separately estimate β from Ω_1. Recall q is the dimension of $(X_{jt}', P_{jt})'$, and $\boldsymbol{\rho}_{1t3}$ is a $q \times q$ triangular matrix. The degree of freedom of the NLS problem is $GJT - (JT + (3q+q^2)T/2 + 2q)$. In order to make this degree of freedom be positive, it is necessary to satisfy $(G-1)J - (3q+q^2)/2 > 0$.[5] 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.

[5] We write $GJT - (JT + (3q+q^2)T/2 + 2q) = [(G-1)J - (3q+q^2)/2]T - 2q$. When T is relatively large, $(G-1)J - (3q+q^2)/2 > 0$ will also be sufficient to have positive degree of freedom.