Linear Estimation of Aggregate Dynamic Discrete Demand for Durable Goods without the Curse of Dimensionality

Cheng Chou, Tim Derdenger and Vineet Kumar*

February 2019

Latest Version:
https://www.andrew.cmu.edu/user/derdenge/CDKEstimator.pdf.

Abstract

We develop a new method to estimate dynamic discrete choice demand models for durable goods in settings with aggregate market level data. The method is developed for market settings with two or more terminal choices and either a multinomial logit or nested logit structure. It is able to estimate the consumer's discount factor and utility preferences without reducing the dimension of the state space or imposing beliefs on how consumers form expectations. Our estimator also recovers the dynamic state evolution of unobservable product characteristics. Despite all this, the computational complexity of our method is similar to that of a linear regression. We characterize the asymptotic properties of the estimator and demonstrate its performance in the finite sample case. Lastly, we show how it can be implemented in an application where we estimate the demand for smartphones.

*Cheng Chou is Assistant Professor of Economics at the School of Business, University of Leicester. email:cchou@le.ac.uk. Tim Derdenger is Associate Professor of Marketing and Strategy at the Tepper School of Business, Carnegie Mellon University email:derdenge@andrew.cmu.edu. Vineet Kumar is Assistant Professor of Marketing at School of Management, Yale University. email:vineet.kumar@yale.edu. The authors benefited from discussions with Robert Miller, Geert Ridder, Xun Tang, and participants in the econometrics seminar at UC Riverside. All errors remain their own.

1 Introduction

In recent years, dynamic discrete choice (DDC) models have become more prevalent in marketing and economics due to their ability to analyze the dynamic aspects of firms' and consumers' preferences, and the consequent intertemporal tradeoffs across a wide range of substantive contexts. As researchers recognize and seek to incorporate these factors into their modeling, the complexity of estimating such models remains a challenging barrier for research. Specifically, defining a tractable state space for such models is often a difficult task, leading some to adopt ad hoc approximation approaches. The task becomes even more challenging in the absence of an approximation method and when the researcher incorporates multiple dimensions of unobserved state variables, individual and choice specific. In the demand estimation literature these unobservables relate to traditional individual-product specific idiosyncratic errors and unobserved product characteristics, respectively.[1] Estimation is further complicated when the unobserved product characteristics are serially correlated and correlated with observed state variables given that computing the expected value and likelihood functions involve high-dimensional integration over all unobserved state variables (idiosyncratic and product characteristics). This is especially problematic when there are many available products, each with their own unobserved characteristic.

Our paper focuses on identification and estimation of dynamic discrete choice demand models with *continuous unobserved product specific state variables*, in addition to the commonly included individual-product idiosyncratic errors. The unobserved product specific state variables are specified as serially correlated and correlated with the observed product characteristics, including price. We demonstrate that the complexity of estimating models with these specific dimensions of unobserved state variables used with aggregate data can be overcome with results that eliminate the curse of dimensionality and hence the need for high-dimensional integration. We present an approach that linearly estimates model primitives, including the discount factor, without having to specify a consumer belief structure for the dynamic evolution of the state space.

Recent papers have looked at a variety of topics under the area of research that includes models with unobserved state variables (Magnac and Thesmar, 2002; Norets, 2009; Kasahara and Shimotsu, 2009; Arcidiacono and Miller, 2011; Hu and Shum, 2012; Hu, Shum, Tan, and Xiao, 2017). To the best of our knowledge, no paper proposes identification and estimation of structural forward-looking decisions where a subset of unobserved dynamic state variables

[1]The inclusion of the latter unobserved state is necessary to account for the endogeneity of product price or other observed characteristics

are continuous, serially correlated and correlated with other observed state variables without the use of an approximation method.

Our linear estimation method is also related to the other DDC linear estimators in the literature. One such paper is Bajari, Chu, Nekipelov, and Park (2016) (BCNP in short hereafter). BCNP presents a semiparametric estimator of a finite horizon dynamic discrete choice model with a terminating action. The estimator uses a two-step approach and nonparametrically estimates the conditional expectation of the conditional value function to linearly estimate the discount factor and utility preferences in the second step. Yet, BCNP does not model unobserved choice specific unobservables that are continuous, serially correlated and correlated with other observed state variables. Even without such features, their work requires either one of two sets of assumptions to derive their computationally results. First, both sets of assumptions require "time homogeneity of utility preferences" for identification– a common assumption in the marketing literature. In addition, the researcher must assume either the utility of the default (terminal) action is normalized to a constant across all states or "be a function of time-invariant state variables that only influence the terminal choice utility, but not the per-period utilities of the other options" (BCNP). Unfortunately, both of these latter assumptions are quite restrictive. The first implies that the flow utility the default/terminal action in each period does not vary with respect to the values of the state variable x_t, and more importantly with respect to the distribution of x_t. This has a serious consequence, because the future value of a current choice is the discounted expected sum of all future optimal flow utilities by following the optimal strategy. By prohibiting $u_t(a_t = 0, x_t)$ from changing with respect to x_t for each period t, it greatly restricts how the future value of the current choice changes with respect to the distribution of $(x_{t+1}, x_{t+2}, \dots)$ given x_t and action a_t (Chou and Ridder, 2017). Furthermore, when a counterfactual policy is implemented, it is known that this assumption is not innocuous (Norets and Tang, 2014). The ramification of the second assumption is less severe but is still quite strong. This assumption requires that the utility of the terminal action be a constant, but a constant that is a function of time-invariant state variables that only impacts the terminal choice and not the other choices, which eliminates the above problem of a simple constant normalization for all states.

Our paper presents several important innovations to the literature on dynamic discrete choice modeling. We propose a new approach to model and estimate a large class of DDC models for use with market level data that is able to estimate model primitives with OLS and IV regressions, including the discount factor, without having to assume and estimate an individual's belief structure for the dynamic evolution of the state space. More specifically, the approach to estimate preferences is applicable for stationary and non-stationary models and does not require any excessively strong assumptions (as specified in BCNP) to generate the simplicity of the estimator. All that is needed is a setting where there exists two or more terminal choices for the individual and the standard assumptions about a Markov process, conditional independence, and that the individual level unobservables are either iid type 1 or GEV extreme value distributed.

Our method also allows us to separate the estimation of preferences from the estimation of beliefs. To the best of our knowledge, every other dynamic demand estimator requires the specification of beliefs (rational expectation, perfect foresight or something else) in order to estimate preferences. Typically specifying different believes changes the preference parameter estimates. Our approach allows us to estimate preferences without specifying beliefs, and recover true preferences no matter the underlying belief structure–a belief structure only needs to be specified after estimation of preference for use in counterfactual simulations. Furthermore, we are able to accommodate serially correlated time-varying unobservable product characteristics, which maybe correlated with other observed state variables, without the use of an approximation method.

Given the nature of the problem we study, where the unobservable state variables are possibly serially correlated and correlated with observed state variables, we move past the current literature and show the dynamics of the state evolution are identified with several additional assumptions. We also illustrate that the discount factor for the classic exponential discounted utility framework is identified using only two periods. With market level data an estimate of the discount factor requires only a linear IV regression.

Daljord, Nekipelov, and Park (2018) also derives a simple estimator for the discount factor in optimal stopping models. The primary differences are that our setting involves persistent unobservable state variables, whereas those are not present in the above paper. Our approach also does not require an assumption of time homogeneity of utility preferences as is used in Daljord et al. (2018), BCNP and Abbring and Daljord (2017). Miller and Sanders (1997) is also related to our work, which discusses how one might possibly estimate the dynamic discount factor in a setting where all model parameters enter linearly except for the discount factor through the use of CCPs. They conclude their paper with the statement that "further empirical research is required to pin down the value of this key parameter." We should note that while our findings regarding the discount factor are novel, they are not central to the larger discussion of identification and estimation of consumer preferences and the dynamics of the state evolution.

The last innovation of our paper is that our approach to preference estimation is not an approximation method and thus does not rely on the validity of specific approximations like interpolation or other value function approximations. However, it does require that our model assumptions hold, which can be clearly specified. Moreover, our estimator does not exhibit a curse of dimensionality to consistently estimate preferences even though individual beliefs about the dynamics of the state evolution are unspecified. The reason for such is from the fact that the estimator does not require the estimation or approximation of the ex-ante expected value function, as is almost always the case with prior papers. While papers like BCNP do not calculate the ex-ante value function through backwards induction or forward simulation nor include unobserved states outside the idiosyncratic error term, their estimator does exhibits a curse of dimensionality from the fact they must estimate the agent's (borrower's) choice probabilities and the continuation value conditional on the current state and action; each of which requires the use of orthogonal polynomials of the state variables.

There are a number of approaches developed in the literature for the estimation of DDC models as we have discussed earlier. Two papers that were not addressed are Hotz and Miller (1993) (HM) and Berry (1994). HM helped drive the stream of literature on demand estimation by generally proving that individual choice probabilities can be mathematically inverted to obtain an estimate of the difference between choice-specific value functions with the use of individual choice level data. Berry (1994) similarly illustrated that market shares can be inverted to recover differences in consumer static utilities along with the incorporation of unobserved product characteristics. Each of which followed the early work of McFadden (1978), who had shown that the choice probabilities in the well-known logit formulation take the form $p_j = \exp(v_j)/(\sum_k \exp(v_k))$.

While Berry (1994) focus was on demand estimation with market level data, its domain is in a static environment. The literature on dynamic discrete choice demand estimators for market level data is scarce and those that do exist mostly focus on approximation methods that reduce the model state space in order to determine the ex-ante value function in a computationally feasible manner. Extending the models of Berry (1994) and Berry, Levinsohn, and Pakes (1995) to a dynamic setting with forward-looking agents is challenging. First, observe that the state space grows in the number of choices (products or brands). Second, there are observable and unobservable characteristics and in many cases the continuous unobserved state variable is highly correlated with the observed states. Finally, the observed state variables are typically continuous variables, and to capture them in a discrete state space requires a sufficiently large number of points to obtain a reasonable approximation.[2]

One approach around this challenge is to model the market shares (choice shares) using a representative consumer model with idiosyncratic taste preferences, as detailed in Song and Chintagunta (2003). However, it is often helpful to include unobservable product attributes like quality or design in the model that are valued by consumers, but are not captured in the data. These attributes could be modeled simply as fixed effects that are time-invariant as in Goettler and Gordon (2011). We deviate from this approach by assuming such attributes are time varying, similar to the static literature on demand estimation (Berry (1994) and Berry et al. (1995)).

Another commonly used approach is the inclusive value sufficiency assumption (Melnikov, 2013; Gowrisankaran and Rysman, 2012), which collapses the state space into a single dimension capturing a utility measure, rather than keep a more complex and granular state space with unobserved states. Derdenger and Kumar (2018) have studied the approximation properties of this approach, and have shown that in general it is a biased and an inconsistent estimator.

Two other papers Dubé, Fox, and Su (2012) and Sun and Ishihara (2018) also present estimators for DDC choice models with aggregate data. They however take a different approach to estimating these models by focusing on computational tools that eliminate the costly nested fixed point algorithm used in (Melnikov, 2013; Gowrisankaran and Rysman, 2012), and speed up the estimation of model primitives.

The simplicity of our estimator is quite powerful, but does come at a cost–as we have identified four main limitations. First, consumers in our model must face an optimal stopping situation in that their choice is to continue in the market without purchasing ("no purchase") or to purchase a product and forever exit the market (terminal choice). Specifically, the model must have two or more terminal choices for the estimator to be linear in preferences and for preferences to be estimated via instrumental variables. That said, the model and estimator does allow for non-terminal choices where an individual is faced with a choice of say "lease one car" as long as the decision does not affect the future transition of state variables. Thus, environments where choices exhibit state dependence with repeat purchases are omitted. Second, our computationally simple approach applies to a class of models similar to Berry (1994) with no or a restrictive (GEV distributed) pattern of persistent unobservable heterogeneity in preferences. This limitation eliminates any possibility

[2] for J products and N points to approximate each characteristic, the state space would grow in N^{2J}.

of incorporating unobserved consumer heterogeneity in preferences as in Berry et al. (1995). This may be problematic to those interested in understanding policies targeted to heterogeneous populations, though it should be highlighted that our model can incorporate any observable heterogeneity for a finite number of classes. However, it is well known that identifying unobserved consumer heterogeneity using aggregate data is quite difficult in practice. Albuquerque and Bronnenberg (2009) illustrate that, "in isolation neither variable [(market share or brand penetration)] may lead to precise estimates of heterogeneity." Sudhir (2013) also states that "identification of heterogeneity is tough with aggregate data." As a result, we attempt to mitigate the lack of unobserved heterogeneity through the estimation of a GEV model. With that said, any future work in this domain should follow Albuquerque and Bronnenberg (2009), Sudhir (2013) and others and include in practice additional data and moment conditions to precisely pin down the distribution of unobserved consumer heterogeneity.

Third, in our model we generally can only identify the difference between two unobserved product characteristics, a challenge for counterfactual analysis. We attempt to address this concern with three approaches. The first is to simply draw from the identified distribution of only one unobserved state variable a large number of times to provide a confidence interval on the policy experiment. There is no added cost to this method as in practice this is exactly what a researcher does in order to generate a confidence interval (draws from all parameters) even when the levels of the unobserved state variables are identified. Next, if the researcher does not want to provide confidence bounds on the policy experiment, we show that unobserved product characteristics are identified if the correlation between at least one unobserved state variable and price is perfectly correlated. This second option has the benefit of being testable. The last option also relates to the market setting in that the product unobservables are identified if there exists at least one non-terminal choice where the individual's decision does not affect the future transition of state variables.

The last limitation is a required stationarity assumption for the identification and estimation of the dynamic evolution of state variables. In particular, we require the joint distribution of the unobserved and observed product characteristics and price to be time invariant for at least two periods. If such joint distribution changes in every period, the model will not be identified. The intuition is similar to the identification of a linear panel data model where regression coefficients and the unobserved fixed effect are assumed time invariant for at least two periods in order to use a fixed effect or first difference estimator to identify/estimate the model. This is hardly a limitation in application, however. When the number of periods is large and one suspects that the joint distribution of product characteristics and price could have changed, one can split the sample into a few sub-segments, and estimate the preferences and/or the dynamic evolution of state variables for each subsegment, as long as there are enough number of periods in each sample.

After presenting the identification and estimation of our estimator, we illustrate the use of our estimator with data from the cell phone market. Using monthly data from ten different states we estimate consumer preferences for phone hardware including smartphones. We determine Apple had the largest fixed effect and Blackberry had the smallest out of all brands. Additionally, we find the unobserved product characteristics were positively serially correlated for Apple, yet were negatively for Blackberry. After the recovery of consumer preferences, we run several counterfactuals to identify the feature that most impacts consumer adoption. Counterfactual analysis finds that removing Bluetooth or Wi-Fi from phones dramatically changes the within market shares. Without Wi-Fi, Apple's iPhone would lose substantial market share compared to other brands. This is due to Wi-Fi almost exclusively being available only on the iPhone. Moreover, Bluetooth was found to have the largest overall demand on the market, leading to roughly a 20 percentage point increase in the market share of the outside good.

The rest of the paper is structured as follows. In §2, we present the basic modeling approach. In §3, we detail the assumptions, and show the identification for the model parameters, constructively. In §4, we obtain the estimators for preference parameters including the discount factor. In §5, we discuss possible methods to implement counterfactual analysis. In §6, we provide an empirical application of the model in the smartphone hardware market around the introduction of the iPhone. In the counterfactual analysis, we evaluate market outcomes when product characteristics exogenously change. Lastly, we present in appendix A the derivatives for calculating the asympotitc variance, appendix B presents the extension to the GEV model, and appendix C provides Monte Carlo evidence on the performance of the estimators (logit and nested logit).

2 Model

Our model follows the previous literature on dynamic discrete choice models of demand, particularly those that employ market level data. Although the model is general, it is especially appropriate for durable products, 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.

The choice set of a consumer i in period t is $\mathcal{J}_t \subseteq \mathcal{J} \equiv \{0, 1, \ldots, J\}$, where 0 denotes outside good, "no purchase," and $1, \ldots, J$ are products. The possible time varying choice set corresponds to the observed entry-exit of products in the market. In each period t, consumer i considers whether or not to purchase a product from the available products $\mathcal{J}_t \setminus \{0\}$. If he decides to purchase, he then chooses which to buy. Once a consumer has purchased a product, he exits the market completely. Hence, purchasing a product is a terminal action in our model. The consumer decision process is thus equivalent to an optimal stopping problem. The presence of a terminal choice would greatly simplify the identification and estimation because the expected life-time utility of a terminal choice is easy to characterize.

2.1 Consumer Utility

Consumers consider numerous product and market characteristics that may affect their current and future purchase utilities, such as price, age of product and quality. The state can be described as $\Omega_{it} \equiv (x_t, p_t, \xi_t, \varepsilon_{it})$, where p_t denotes the vector of product prices, x_t denotes the vector of the other observable product characteristics, ξ_t denotes the unobserved (to econometrician) product characteristics, and ε_{it} is the vector of individual choice-specific idiosyncratic shocks, which are unobservable to researchers. Denote $m_t \equiv (x_t, p_t, \xi_t)$ the market level state.

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

Typically, in a product choice model, we can include all the product variables in the state space, $x'_t \equiv (x'_{1t}, \ldots, x'_{Jt})$ and $p_t \equiv (p_{1t}, \ldots, p_{Jt})'$, where x_{jt} and p_{jt} denote the vector of observable product characteristics and the price of product j in period t, respectively. There is some abuse of notation because x_{jt} and p_{jt} are indeed not defined if product j does not exist in period t, i.e. $j \notin \mathcal{J}_t$.

We normalize the expected period utility of the outside good to be 0. Hence, if consumer i does not purchase in period t, he receives flow utility

$$u_{i0t} = 0 + \varepsilon_{i0t}.$$

This normalization is only for simplicity of exposition. Our arguments still hold when u_{i0t} is a parametric function of observed characteristics of the outside good and additive in ε_{i0t}. This is useful because it has been shown that unlike the case of static discrete choice models, normalization in dynamic discrete choice models is not innocuous for the purpose of counterfactual predictions (e.g. Norets and Tang, 2014; Kalouptsidi, Scott, and Souza-Rodrigues, 2015).

When consumer i purchases product j at time t, his flow utility *during the purchase period* t is:

$$u_{ijt} = f(x_{jt}, \xi_{jt}) - \alpha p_{jt} + \varepsilon_{ijt}. \tag{1}$$

He then receives the identical flow utility $f(x_{jt}, \xi_{jt})$ in each period $\tau > t$ following his purchase. In particular, let

$$f(x_{jt}, \xi_{jt}) = x'_{jt}\gamma + \delta_j + \xi_{jt}.$$

Let $\delta = (\delta_1, \ldots, \delta_J)'$. The term δ_j is the unobserved product fixed effect. The vector $\xi_t = (\xi_{1t}, \ldots, \xi_{Jt})'$ is unobservable to researchers, and ξ_{jt} is a scalar with $\mathrm{E}(\xi_{jt}) = 0$. One typically views $\delta_j + \xi_{jt}$ as a measure of functional or design quality. Hereafter, we refer ξ_{jt} as the "quality" of product j at time t.

2.2 Dynamic Decision Problem

The consumer makes a trade-off between buying in the current period t and waiting to make a purchase in the next period. The crucial intertemporal trade-off is in the consumer's expectation of how the market level state variables $m_t = (x_t, p_t, \xi_t)$ evolve in the future. For example, if the product characteristics (or price) are expected to improve over time, then the consumer is incentivized to wait.

Consumer i in period t chooses from the set of choices $\mathcal{J}_t$, which includes the option 0 to wait without purchasing any product. However, if the consumer purchases, recall that he exits the market immediately upon purchase.

For a consumer in the market faced with a state Ω_{it} in period t, we can write the Bellman equation in terms of the value function $V_t(\Omega_{it})$ as follows:

$$V_t(\Omega_{it}) = \max\left(\varepsilon_{i0t} + \beta\,\mathrm{E}(V_{t+1}(\Omega_{i,t+1}) \mid \Omega_{it}), \max_{j \in \mathcal{J}_t \setminus \{0\}} v_j(\Omega_{it}) + \varepsilon_{ijt}\right),$$

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

$$v_{jt}(\Omega_{it}) = \frac{f(x_{jt}, \xi_{jt})}{1-\beta} - \alpha p_{jt} = \frac{x'_{jt}\gamma + \delta_j + \xi_{jt}}{1-\beta} - \alpha p_{jt}, \qquad j \in \mathcal{J}_t \setminus \{0\}. \tag{2}$$

We also let

$$v_{0t}(\Omega_{it}) = \beta \, \mathrm{E}(V_{t+1}(\Omega_{i,t+1}) \mid \Omega_{it}). \tag{3}$$

The value function $V_t(\Omega_{it})$ involves consumer i's flow utility shock ε_{it}. Assumption 2(i) below ensures

$$\mathrm{E}(V_{t+1}(\Omega_{i,t+1}) \mid \Omega_{it}) = \mathrm{E}\big(\bar{V}_{t+1}(x_{t+1}, p_{t+1}, \xi_{t+1}) \bigm| x_t, p_t, \xi_t\big),$$

where

$$\bar{V}_{t+1}(x_{t+1}, p_{t+1}, \xi_{t+1}) \equiv \mathrm{E}(V_{t+1}(\Omega_{i,t+1}) \mid x_{t+1}, p_{t+1}, \xi_{t+1}).$$

The expectation in the above display is taken over $\varepsilon_{i,t+1}$.

Assumption 2 (Conditional independence)**.** *For all t, we have*

(i) $\Omega_{i,t+1} \perp\!\!\!\perp \varepsilon_{it} \mid (x_t, p_t, \xi_t)$*;*

(ii) $\varepsilon_{i,t+1} \perp\!\!\!\perp \Omega_{it} \mid (x_{t+1}, p_{t+1}, \xi_{t+1})$*.*

The role of part (ii) will be clear soon. Under Assumption 2, we know that v_j is a function of market level state variables $m_t = (x_t, p_t, \xi_t)$ only. Let s_{jt} be the market share of product j at time t. Given a conditional distribution function $F(\cdot \mid m_t)$ of ε_{it}, we have

$$\begin{aligned} s_{jt}(m_t) &= \Pr(v_{jt}(m_t) + \varepsilon_{ijt} \geq v_{kt}(m_t) + \varepsilon_{ikt}, k \in \mathcal{J}_t \mid m_t) \\ &= \int 1\left(v_{jt}(m_t) + \varepsilon_{ijt} \geq v_{kt}(m_t) + \varepsilon_{ikt}, k \in \mathcal{J}_t\right) F(\mathrm{d}\,\varepsilon_{it} \mid m_t). \end{aligned} \tag{4}$$

Our results below do not require that the value function $V_t(\Omega_{it})$ or the integrated value function $\bar{V}_t(m_t)$ be time invariant. This could be desirable in applications, because the introduction of new products or technology innovation could change the consumer's value function.

3 Identification

We start by clarifying the data and the structural parameters of the model. With the data, we observe market shares s_{jt}, observable product characteristics x_{jt} and prices p_{jt} for $j \in \mathcal{J}_t$.

Structural parameters include consumer preference parameters $\theta_1 = (\alpha, \beta, \gamma', \delta')'$, the state transition distribution function $F(\Omega_{i,t+1} \mid \Omega_{it})$, and the initial distribution function $F(\Omega_{it})$ for some period t. In general, we need to know θ_1, $F(\Omega_{i,t+1} \mid \Omega_{it})$ and $F(\Omega_{it})$ in order to simulate the consumer's dynamic decisions starting from period t and market shares under various counterfactual experiments.

Using conditional independence (Assumption 2), we have

$$F(\Omega_{it}) = F(m_t)F(\varepsilon_{it} \mid m_t), \qquad F(\Omega_{i,t+1} \mid \Omega_{it}) = F(m_{t+1} \mid m_t)F(\varepsilon_{i,t+1} \mid m_{t+1}).$$

Moreover, we will assume that $\varepsilon_{it} \perp\!\!\!\perp m_t$ and $F(\varepsilon_{it})$ are known for all t. We can write $F(m_t) = F(x_t, p_t)F(\xi_t \mid x_t, p_t)$. Thus, the CDF $F(x_t, p_t)$ is identified from observed x_t and p_t. Our focus is then on $F(\xi_t \mid x_t, p_t)$ and $F(m_{t+1} \mid m_t)$. The difficulty is that we do not observe ξ_t. In the remainder of this section, we show how to identify θ_1, $F(\xi_t \mid x_t, p_t)$, and $F(m_{t+1} \mid m_t)$ nonparametrically under mild restrictions.

We give a brief summary of our results in this section. To identify preference θ_1, one only needs to know $F(\varepsilon_{it} \mid m_t)$ and to have instrumental variables that are uncorrelated with unobserved qualities ξ_t. To identify $\mathrm{E}(\xi_{jt} \mid x_t, p_t)$, we further assume that $F(\xi_t \mid x_t, p_t)$ is time invariant. To identify $\mathrm{Var}(\xi_{jt})$ and $\mathrm{Var}(\xi_{jt} \mid x_t, p_t)$, one needs one additional assumption that is to assume that the unobserved qualities $\xi_{1t}, \ldots, \xi_{Jt}$ are independent and homoscedastic conditional on x_t and p_t. To identify $F(\xi_t \mid x_t, p_t)$ nonparametrically, one needs a further assumption that is to assume that the qualities ξ_{jt} have identical distribution except for their conditional mean. To identify $F(m_{t+1} \mid m_t)$ nonparametrically, one needs additional assumptions, among which one would require that ξ_{t+1} is an autoregressive process and $x_{t+1} \perp\!\!\!\perp (\xi_t, \xi_{t+1}) \mid (x_t, p_t)$. Most of our identification results are constructive, hence they can be used as formulas for estimation.

It is well known in the literature of identifying dynamic discrete choice model (e.g. Magnac and Thesmar, 2002) that without assuming that $F(\varepsilon_{it} | m_t)$ is known, the flow utility functions and discount factor are not identified. Our restriction on $F(\varepsilon_{it} \mid m_t)$ is twofold. First, we assume $\varepsilon_{it} \perp\!\!\!\perp m_t$. Second, we know the marginal distribution of ε_{it}, which will be Type I extreme value distribution.

Assumption 3. *Assume that consumer i's utility shocks $\varepsilon_{it} = (\varepsilon_{i0t}, \ldots, \varepsilon_{iJt})'$ are independent of $m_t = (x_t, p_t, \xi_t)$.*

Assumption 4. *Let $\omega \approx 0.5772$ be Euler's constant. Assume that $\varepsilon_{i0t} + \omega, \ldots, \varepsilon_{iJt} + \omega$ are independent identically distributed Type I extreme value with density $f(\varepsilon_{ijt} + \omega = \varepsilon) = \exp[-(\varepsilon + e^{-\varepsilon})]$. Adding Euler's constant is to make $\mathrm{E}(\varepsilon_{ijt}) = 0$.*

Assumption 3 does not allow correlation between market level state variables and unobserved consumer heterogeneity. This can be restrictive in some applications. For example, consumers may be heterogeneous in their preference for design or quality, which is captured by ξ_t in this model. Such consumer preference is unobserved, hence it is denoted by ε_{it}. This implies that ε_{it} and ξ_t are correlated. Allowing for such correlation between ε_{it} and the other state variables in general has been a difficult problem in the literature of dynamic discrete choice model (see Magnac and Thesmar, 2002; Arcidiacono and Miller, 2011). It seems to be harder here since ξ_t in m_t is unobservable.

Assumption 4 is not essential for our arguments, but it greatly simplifies the exposition.

3.1 Consumer Preference

Let $\theta'_{1o} = (\alpha_o, \beta_o, \delta_o, \gamma'_o)$ denote the true values. To make the idea clear, we consider a simple case with two products (1 and 2) in addition to the outside good 0. Both products are always available. It follows from the multinomial logit model that the market share $s_{jt}(m_t)$ has the following formula

$$s_{jt}(m_t) = \exp(v_{jt}(m_t))/\textstyle\sum_{k\in\mathcal{J}_t}\exp(v_{kt}(m_t)).$$

Hence for any two products $j, k \in \mathcal{J}_t$, we have

$$s_{jt}/s_{kt} = \exp(v_{jt}(m_t))/\exp(v_{kt}(m_t)),$$

or

$$\ln(s_{jt}/s_{kt}) = v_{jt}(m_t) - v_{kt}(m_t). \tag{5}$$

The moment conditions used in the identification arguments as well as in the estimation below are from the log shares ratio between two products. To show identification, we will only use the two ratios $\ln(s_{2t}/s_{1t})$ and $\ln(s_{2t}/s_{0t})$.

Equation (5) is similar to Berry (1994). The key difference is that $v_{0t}(m_t)$ in Berry or BLP equals zero, while $v_{0t}(m_t)$ here depends on an unknown value function.

In eq. (5), letting $j = 2, k = 1$, we have $\ln(s_{2t}/s_{1t}) = v_{2t}(m_t) - v_{1t}(m_t)$, that is

$$\ln\left(\frac{s_{2t}}{s_{1t}}\right) = (x_{2t} - x_{1t})'\tilde{\gamma} - \alpha(p_{2t} - p_{1t}) + \frac{\delta_2 - \delta_1}{1-\beta} + \frac{\xi_{2t} - \xi_{1t}}{1-\beta}, \tag{6}$$

with

$$\tilde{\gamma} = \gamma/(1-\beta).$$

Equation (5) explains the relative market share by the difference of product characteristics. Equation (5) resembles a linear regression since we observe $\ln(s_{2t}/s_{1t})$, $(x_{2t} - x_{1t})$ and $(p_{2t} -$

p_{1t}). Let $z_{(2,1),t}$ denote a vector of instruments that are uncorrelated with $\xi_{2t} - \xi_{1t}$. We can identify $\tilde{\gamma}$, α, and $(\delta_2 - \delta_1)/(1-\beta)$ with *one period of data* from the moment equation

$$\mathrm{E}(g_{1,(2,1),t}(\theta_{1o})) = 0,$$

$$g_{1,(2,1),t}(\theta_1) = z_{(2,1),t}\left[\ln\left(\frac{s_{2t}}{s_{1t}}\right) - (x_{2t} - x_{1t})'\tilde{\gamma} + \alpha(p_{2t} - p_{1t}) - \frac{\delta_2 - \delta_1}{1-\beta}\right]. \tag{7}$$

We next show the identification of the discount factor β and product fixed effect δ. Once β is identified, γ is identified from the already identified $\tilde{\gamma} = \gamma/(1-\beta)$. Taking natural logarithms of both sides of eq. (5), and letting $j = 2, k = 0$, we have

$$\ln\left(\frac{s_{2t}}{s_{0t}}\right) = x_{2t}'\tilde{\gamma} - \alpha p_{2t} + \frac{\delta_2}{1-\beta} + \frac{\xi_{2t}}{1-\beta} - \beta\,\mathrm{E}\big(\bar{V}_{t+1}(m_{t+1}) \,\big|\, m_t\big). \tag{8}$$

Define the already identified term y_t:

$$y_t = \ln(s_{2t}/s_{0t}) - x_{2t}'\tilde{\gamma} + \alpha p_{2t}.$$

Note that y_t is a function of m_t only. We rewrite eq. (8) with $y_t(m_t)$,

$$y_t(m_t) = \frac{\delta_2}{1-\beta} + \frac{\xi_{2t}}{1-\beta} - \beta\,\mathrm{E}\big(\bar{V}_{t+1}(m_{t+1}) \,\big|\, m_t\big). \tag{9}$$

By the expectation maximization formula of the multinomial logit model (Arcidiacono and Miller, 2011), we have

$$\bar{V}_t(m_t) = v_2(m_t) - \ln s_{2t}(m_t) = \left(x_{2t}'\tilde{\gamma} - \alpha p_{2t} + \frac{\delta_2}{1-\beta} + \frac{\xi_{2t}}{1-\beta}\right) - \ln s_{2t}(m_t). \tag{10}$$

Define another identified term w_t:

$$w_t = x_{2t}'\tilde{\gamma} - \alpha p_{2t} - \ln s_{2t}(m_t).$$

Clearly, w_t is a function of m_t only. Then we have

$$\bar{V}_t(m_t) = w_t(m_t) + \frac{\delta_2}{1-\beta} + \frac{\xi_{2t}}{1-\beta}, \qquad \text{for all } t.$$

Substituting $\bar{V}_{t+1}(m_{t+1})$ in eq. (9) with the above display, we have the conditional moment restriction

$$y_t(m_t) = \frac{\delta_2}{1-\beta} + \frac{\xi_{2t}}{1-\beta} - \beta\,\mathrm{E}\left(w_{t+1}(m_{t+1}) + \frac{\delta_2}{1-\beta} + \frac{\xi_{2,t+1}}{1-\beta} \,\middle|\, m_t\right). \tag{11}$$

Since $y_t(m_t)$ is a function of m_t only, $\mathrm{E}(y_t \mid m_t) = y_t$. Moreover, $\mathrm{E}(\xi_{2t} \mid m_t) = \xi_{2t}$ because ξ_{2t} is an element of m_t. As a result, the above display implies

$$\mathrm{E}\left(y_t + \beta w_{t+1} - \delta_2 - \frac{1}{1-\beta}\xi_{2t} + \frac{\beta}{1-\beta}\xi_{2,t+1} \,\middle|\, m_t\right) = 0. \tag{12}$$

By this conditional moment condition, we know that for any integrable function $\eta(m_t)$ we have

$$\mathrm{E}\left[\left(y_t + \beta w_{t+1} - \delta_2 - \frac{1}{1-\beta}\xi_{2t} + \frac{\beta}{1-\beta}\xi_{2,t+1}\right)\eta(m_t)\right] = 0. \tag{13}$$

The conditional moment eq. (12) will be very useful. As its first application, we show the identification of product fixed effect δ when the discount factor β is known. Given β, letting $\eta(m_t) = 1$, we have

$$\delta_2 = \mathrm{E}(y_t + \beta w_{t+1}).$$

Here we used $\mathrm{E}(\xi_{2t}) = \mathrm{E}(\xi_{2,t+1}) = 0$. Since we have identified $(\delta_2 - \delta_1)/(1-\beta)$, we identify δ_1 given δ_2 and β. Alternatively, by similar arguments, one can directly identify δ_1 by using the log share ratio $\ln(s_{1t}/s_{0t})$.

As the second application of eq. (12), we show the identification of β. The market level state m_t includes $(x'_{1t}, x'_{2t}, p_{1t}, p_{2t})$. Let $x_{(2,0),t,IV}$ be a vector of functions of m_t such that $\mathrm{cov}(x_{(2,0),t,IV}, \xi_{2t}) = \mathrm{cov}(x_{(2,0),t,IV}, \xi_{2,t+1}) = 0$.[3] We can identify β and δ_2 from

$$\mathrm{E}(g_{2,(2,0),t}(\theta_{1o})) = 0,$$

$$g_{2,(2,0),t}(\theta_1) = \Big(y_t + \beta w_{t+1} - \delta_2,\ (y_t + \beta w_{t+1} - \delta_2)x'_{(2,0),t,IV}\Big)'. \tag{14}$$

For example, if $x_{(2,0),t,IV}$ is a scalar, we explicitly have

$$\beta = -\mathrm{cov}(y_t, x_{(2,0),t,IV})/\mathrm{cov}(w_{t+1}, x_{(2,0),t,IV}), \tag{15}$$

provided that $\mathrm{cov}(w_{t+1}, x_{(2,0),t,IV}) \neq 0$ (corresponding to rank condition). From the definition of w_{t+1}, the rank condition requires that $x_{(2,0),t,IV}$ must be correlated with the next period market level state variables or market share $s_{2,t+1}$. The following proposition is a summary about the identification of consumer preference.

Proposition 1. *Suppose Assumption-4 hold. Let $d_x = \dim x_{jt}$. If there is a vector of instrumental variables $z_{(2,1),t}$ such that $\mathrm{E}[z_{(2,1),t}(\xi_{2t} - \xi_{1t})] = 0$ and $\mathrm{rank}\,\mathrm{E}[z_{(2,1),t}((x_{2t} - x_{1t})', (p_{2t} - p_{1t}))] = d_x + 1$, and there is a vector-valued function $x_{(2,0),t,IV}$ of m_t such that $\mathrm{cov}(x_{(2,0),t,IV}, \xi_{2t}) = 0$, $\mathrm{cov}(x_{(2,0),t,IV}, \xi_{2,t+1}) = 0$ and $\mathrm{cov}(w_{t+1}, x_{(2,0),t,IV}) \neq 0$, we can identify consumer preference parameters α, β and γ and product fixed effect δ with two periods data.*

Moreover, the above constructive identification arguments suggest a simple estimation method for estimating $\theta_1 = (\alpha, \beta, \delta', \gamma')'$. A simple IV regression can estimate $\tilde{\gamma}$, $(\delta_2 - \delta_1)/(1-$

[3] It should be remarked that $x_{(2,0),t,IV}$ does not need to be a component of x_{2t}. For example, $x_{(2,0),t,IV}$ can be $x_{1t} + x_{2t}$ if $\mathrm{cov}(x_{jt}, \xi_{2t}) = \mathrm{cov}(x_{jt}, \xi_{2,t+1}) = 0$ for both $j = 1$ and 2.

β) and α. The discount factor β can be estimated by the sample analog of eq. (15) or the generalized method of moment (GMM) in the overidentification case. Such an estimator does not impose any further distributional assumptions about state transition law besides the first-order Markovian assumption.[4] As a result, there is no "curse of dimensionality" in the estimation of consumer preferences.

Remark 1 (Why can we identify the discount factor?). In dynamic discrete choice models, in order to identify the discount factor, it is usually necessary to have an excluded variable that does not affect current utility but does impact future payoff (e.g. Fang and Wang, 2015). To see why we can identify the discount factor even without the excluded variable, let's assume that there are no unobserved product characteristics ξ_{jt} and $\delta_j = 0$. The key reason is that we can identify the mean value v_j for each product j from relative market shares. Without ξ_{jt}, we have

$$\ln(s_{2t}/s_{1t}) = (x_{2t} - x_{1t})'\tilde{\gamma} - \alpha(p_{2t} - p_{1t}).$$

We identify $\tilde{\gamma}$ and α, hence v_j for every product j. Knowing v_j and the market share s_{jt} from data, we henceforth know the integrated value function $\bar{V}_t$ by (Arcidiacono and Miller, 2011)

$$\bar{V}_t(m_t) = v_j(x_{jt}, p_{jt}) - \ln s_{jt}.$$

Next,

$$\begin{aligned} \ln(s_{2t}/s_{0t}) &= v_2(x_{2t}, p_{2t}) - v_0(m_t) \\ &= v_2(x_{2t}, p_{2t}) - \beta \, \mathrm{E}(\bar{V}_{t+1}(m_{t+1}) \mid m_t) \\ &= v_2(x_{2t}, p_{2t}) - \beta \, \mathrm{E}[v_2(x_{2,t+1}, p_{2,t+1}) - \ln s_{2,t+1} \mid m_t]. \end{aligned} \tag{16}$$

Note that $m_t = (x_{1t}, p_{1t}, x_{2t}, p_{2t})$ here. Because we know v_2, and market shares $s_{2t}, s_{0t}, s_{2,t+1}$ are included in the data, we can identify the conditional expectation term, hence β.

In general, in dynamic discrete choice models, the mean value v_j for each alternative j depends on the unknown value function, hence β cannot be identified from the relative choice probabilities first. Our arguments do not apply to the general dynamic discrete choice model.

With unobserved product characteristics ξ_{jt}, we are required to use $x_{(2,0),t,IV}$, a non-random function of m_t. Taking the conditional expectation of both sides of eq. (16) given

[4] Implicitly, we assumed the unobserved product characteristics are mean stationary.

$x_{(2,0),t,IV}$, we have

$$\mathrm{E}(\ln(s_{2t}/s_{0t}) \mid x_{(2,0),t,IV}) = \mathrm{E}(v_2(x_{2t}, p_{2t}, \xi_{2t}) \mid x_{(2,0),t,IV}) - \\ \beta\,\mathrm{E}[v_2(x_{2,t+1}, p_{2,t+1}, \xi_{2,t+1}) - \ln s_{2,t+1} \mid x_{(2,0),t,IV}].$$

Because the unobserved ξ_{2t} enters in $v_2(x_{2t}, p_{2t}, \xi_{2t})$ additively, ξ_{2t} and $\xi_{2,t+1}$ disappear from the above display by $\mathrm{E}(\xi_{2t} \mid x_{(2,0),t,IV}) = \mathrm{E}(\xi_{2,t+1} \mid x_{(2,0),t,IV}) = 0$. ■

Remark 2 (What if there were only one product on the market?). When there is only one product on the market, one can still identify consumer preferences $(\alpha, \beta, \gamma, \delta)$ with certain rank condition. However, such identification has limited practical relevance.

Suppose product 2 is the only product on the market. By definition of market share, $s_{2t} = 1 - s_{0t}$. For identification, we have only eq. (8). We still have eq. (10) and the conditional moment equation eq. (12). The new issue is that y_t and w_{t+1} have not been identified. Explicitly, eq. (12) reads

$$\mathrm{E}(\text{eq. (17)} \mid m_t) = 0,$$

with

$$\ln\left(\frac{s_{2t}}{s_{0t}}\right) - x_{2t}'\tilde{\gamma} + \alpha p_{2t} + \beta\left(x_{2,t+1}'\tilde{\gamma} - \alpha p_{2,t+1} - \ln s_{2,t+1}\right) - \delta_2 - \\ \frac{1}{1-\beta}\xi_{2t} + \frac{\beta}{1-\beta}\xi_{2,t+1} \quad (17)$$

Here $m_t = (x_{2t}, p_{2t}, \xi_{2t})$. The exogenous observed characteristics can only be derived from x_{2t}. Provided that $\mathrm{E}(\xi_{2t} \mid x_{2t}) = \mathrm{E}(\xi_{2,t+1} \mid x_{2t}) = 0$, we have

$$\mathrm{E}\left[\ln\left(\frac{s_{2t}}{s_{0t}}\right) - x_{2t}'\tilde{\gamma} + \alpha p_{2t} + \beta\left(x_{2,t+1}'\tilde{\gamma} - \alpha p_{2,t+1} - \ln s_{2,t+1}\right) - \delta_2 \,\middle|\, x_{2t}\right] = 0,$$

which can be rearranged as follows,

$$\mathrm{E}\left(\ln\left(\frac{s_{2t}}{s_{0t}}\right) \,\middle|\, x_{2t}\right) - \beta\,\mathrm{E}(\ln s_{2,t+1} \mid x_{2t}) + \\ [\beta\,\mathrm{E}(x_{2,t+1} \mid x_{2t}) - x_{2t}]'\tilde{\gamma} + [\mathrm{E}(p_{2t} \mid x_{2t}) - \beta\,\mathrm{E}(p_{2,t+1} \mid x_{2t})]\alpha - \delta_2 = 0. \quad (18)$$

Viewing $\mathrm{E}(\ln(s_{2t}/s_{0t}) \mid x_{2t})$ as the dependent variable, and $\mathrm{E}(\ln s_{2,t+1} \mid x_{2t})$, $\mathrm{E}(x_{2,t+1} \mid x_{2t})$, x_{2t}, $\mathrm{E}(p_{2t} \mid x_{2t})$, and $\mathrm{E}(p_{2,t+1} \mid x_{2t})$ as the independent variables, the above display is just a linear regression equation. Provided that those regressors are not collinear, one can identify α, β, γ, and δ. The collinearity could happen for example if $\mathrm{E}(x_{2,t+1} \mid x_{2t})$ is linear in x_{2t}.

Even if the identification holds, in practice it may not be as useful. To see this, suppose the discount factor β is known. Since the model is linear regression, the variance of the estimator of $\tilde{\gamma}$ is proportional to the inverse of the variance of the regressor $\beta\,\mathrm{E}(x_{2,t+1}\,|\,x_{2t}) - x_{2t}$. In practice, β is close to one, and $x_{2,t}$ is persistent (in some cases, x_{2t} is time invariant). This implies that the variance of $\beta\,\mathrm{E}(x_{2,t+1}\,|\,x_{2t}) - x_{2t}$ can be very small, hence the variance of estimating $\tilde{\gamma}$ can be very large. The same issue applies to the variance of estimating α when price p_{2t} is also persistent.

The above issue does not occur in a static model. In the static model, which corresponds to $\beta = 0$, eq. (12) becomes

$$\ln(s_{2t}/s_{0t}) = x_{2t}'\gamma - \alpha p_{2t} + \delta_2 + \xi_{2t}.$$

Let $\sigma_x^2 = \mathrm{Var}(x_{2t})$. Then the variance of estimating γ from the above static model is proportional to σ_x^{-2}. Considering eq. (18), suppose β is known, and $\mathrm{E}(x_{2,t+1} \mid x_{2t}) = \rho x_{2t}$. The variance of the regressor $(\beta\,\mathrm{E}(x_{2,t+1} \mid x_{2t}) - x_{2t}) = (\beta\rho - 1)x_{2t}$ is then $(1-\beta\rho)^2\sigma_x^2$, hence the variance of estimating γ from the dynamic model is proportional to $(1-\beta\rho)^{-2}\sigma_x^{-2}$. Clearly, if β and ρ are close to one, $(1-\beta\rho)^{-2}$ will be very large. Given this observation, it is not recommended to run a dynamic model with only one product. ■

3.2 Dynamics of State Evolution

We now focus on identification of the firm side variables, m_t, which in turn impact the state space for the consumer. While the identification of consumer preferences above did not require us to assume stationarity of the state evolution process, below we assume stationarity (Assumption 5).

Assumption 5 (Stationary Markov Process)**.** *The first-order Markov process m_t is stationary. The conditional distribution function $F(m_{t+1}\,|\,m_t)$ is time invariant, and $F(m_t)$ is the stationary distribution of m_t.*

We will first show the identification of marginal distribution function $F(m_t)$ then the conditional distribution function $F(m_{t+1}\,|\,m_t)$.

3.2.1 Identification of $F(m_t)$

We first identify $\mathrm{E}(\xi_{jt} \mid x_t, p_t)$ with the stationary assumption 6 about ξ_t below. Then we show nonparametric identification of $F(\xi_t \mid x_t, p_t)$ with additional restrictions.

Assumption 6. *(i) The marginal distribution function $F(\xi_t)$ and the conditional distribution function $F(\xi_t \mid x_t, p_t)$ are both time invariant.*

(ii) $\xi_{t+1} \perp\!\!\!\perp (x_t, p_t) \mid (x_{t+1}, p_{t+1})$.

By eq. (6) and the identification of β, we can identify $\xi_{2t} - \xi_{1t}$. Such difference will be frequently used latter. Denote

$$d_t = \xi_{2t} - \xi_{1t}. \tag{19}$$

It follows from eq. (6) that

$$d_t = (1-\beta)\ln(s_{2t}/s_{1t}) - (x_{2t} - x_{1t})'\gamma - (\delta_2 - \delta_1) + (1-\beta)\alpha(p_{2t} - p_{1t}).$$

Variable d_t is an identified object.

It is important to note that it is likely that we cannot identify ξ_{jt}, only the difference $\xi_{2t} - \xi_{1t}$. Equation (11) reads

$$\frac{\delta_2}{1-\beta} + \frac{\xi_{2t}}{1-\beta} - \beta\,\mathrm{E}\left(w_{t+1} + \frac{\delta_2}{1-\beta} + \frac{\xi_{2,t+1}}{1-\beta} \,\middle|\, x_t, p_t, \xi_t\right) - y_t = 0.$$

The unknown ξ_{2t} appears both linearly and nonlinearly as conditioning variable in the above display. Recall that $w_{t+1} = x'_{2,t+1}\tilde{\gamma} - \alpha p_{2,t+1} - \ln s_{2,t+1}$ and $y_t = \ln(s_{2t}/s_{0t}) - x'_{2t}\tilde{\gamma} + \alpha p_{2t}$. In general, in order to show identification of ξ_{2t}, one needs to prove that the left-hand-side (LHS) of the above display is globally monotone in ξ_{2t}, whose primitive condition is unclear to us because y_t and w_{t+1} depend on market shares, hence value function. It is expected that $\partial y_t/\partial \xi_{2t} > 0$ and $-\partial w_{t+1}/\partial \xi_{2,t+1} > 0$, because the market share is expected to be increasing in ξ_{2t}. As a result, the sign of the derivative of the LHS of the above display with respect to ξ_{2t} is indeterminate, when $\xi_{2,t+1}$ is positively correlated with ξ_{2t}. Intuitively, the increase in ξ_{2t} can make both purchasing now and waiting to purchase in the future more desirable, hence the market share is not necessarily monotone in ξ_{2t}.

One sufficient yet uninteresting condition is that $(x_{2,t+1}, p_{2,t+1}, s_{2,t+1}, \xi_{2,t+1}) \perp\!\!\!\perp \xi_t \,|\, (x_t, p_t)$. In this condition one can drop ξ_t from the conditioning variables from the conditional expectation and solve ξ_{2t}.[5] However, in practice, unobserved product characteristics ξ_t are serially correlated as seen in our empirical application. Moreover, next period's market share is typically correlated with the current period's unobserved product characteristics despite their price and x_t. For example, if the current ξ_t is high, consumers tend to buy now rather than waiting until next period.

[5] In the solution, one will need $\mathrm{E}(\xi_{2,t+1} \mid x_t, p_t)$. Below, we will identify $\mathrm{E}(\xi_{2,t+1} \mid x_{t+1}, p_{t+1})$. Then we can identify $\mathrm{E}(\xi_{2,t+1} \mid x_t, p_t) = \mathrm{E}[\mathrm{E}(\xi_{2,t+1} \mid x_{t+1}, p_{t+1}) \mid x_t, p_t]$ by assumption 6.

We now show how to identify $\mathrm{E}(\xi_{jt} \,|\, x_t, p_t)$. By $\mathrm{E}(\xi_{1t} \,|\, x_t, p_t) = \mathrm{E}(\xi_{2t} \,|\, x_t, p_t) - \mathrm{E}(d_t \,|\, x_t, p_t)$, we only need to show the identification of $\mathrm{E}(\xi_{2t} \mid x_t, p_t)$. Multiplying both sides of eq. (12) with $(1-\beta)$, we have

$$\mathrm{E}\big[(1-\beta)y_t + \beta(1-\beta)w_{t+1} - (1-\beta)\delta_2 - \xi_{2t} + \beta\xi_{2,t+1} \,\big|\, m_t\big] = 0.$$

Since $x_t, p_t \in m_t$, apply the law of iterated expectation, and we have

$$\mathrm{E}\big[(1-\beta)y_t + \beta(1-\beta)w_{t+1} - (1-\beta)\delta_2 - \xi_{2t} + \beta\xi_{2,t+1} \,\big|\, x_t, p_t\big] = 0. \tag{20}$$

Now define

$$h(x,p) = \mathrm{E}[(1-\beta)y_t + \beta(1-\beta)w_{t+1} - (1-\beta)\delta_2 \mid x_t = x, p_t = p],$$

$$\pi(x,p) = \mathrm{E}(\xi_{2t} \mid x_t = x, p_t = p).$$

The function $h(x,p)$ is nonparametrically identified since we observe y_t, w_{t+1}, x_t and p_t. The unknown function $\pi(x,p)$ is the parameter of interest.

Equation (20) implies an integral equation of $\pi(x,p)$, from which $\pi(x,p)$ is identified. We have from eq. (20) that

$$\begin{aligned} h(x_t,p_t) &= \pi(x_t,p_t) - \beta\,\mathrm{E}(\xi_{2,t+1} \mid x_t, p_t) \\ &= \pi(x_t,p_t) - \beta\,\mathrm{E}[\mathrm{E}(\xi_{2,t+1} \mid x_{t+1}, p_{t+1}, x_t, p_t) \mid x_t, p_t] \\ &= \pi(x_t,p_t) - \beta\,\mathrm{E}[\mathrm{E}(\xi_{2,t+1} \mid x_{t+1}, p_{t+1}) \mid x_t, p_t] \\ &= \pi(x_t,p_t) - \beta\,\mathrm{E}(\pi(x_{t+1}, p_{t+1}) \mid x_t, p_t) \\ &= \pi(x_t,p_t) - \beta \int \pi(x,p) F(\,\mathrm{d}\,x,\, \mathrm{d}\,p \mid x_t, p_t). \end{aligned}$$

The third line follows from $\xi_{t+1} \perp\!\!\!\perp (x_t, p_t) \mid (x_{t+1}, p_{t+1})$ in assumption 6. The fourth line used the stationary assumption about $F(\xi_t | x_t, p_t)$. The conditional CDF $F(x_{t+1}, p_{t+1} | x_t, p_t)$ in the last line is identifiable from data about (x_t, p_t). We then have a Fredholm integral equation of type 2,

$$\pi(x,p) - \beta \int \pi(x',p') F(\,\mathrm{d}\,x',\, \mathrm{d}\,p' \mid x, p) = h(x,p).$$

We know that there would be a unique solution of $\pi(x,p)$. The proof is to view the left-hand-side of the above equation as a definition of a linear operator $I - \beta L$, where I is the identity operator and $[L(\pi)](x,p) = \int \pi(x',p') F(\,\mathrm{d}\,x',\, \mathrm{d}\,p' \,|\, x, p)$. Then the integral equation reads

$$[(I - \beta L)\pi](x,p) = h(x,p).$$

Because $F(x', p' \mid x, p)$ is a CDF, L has unit norm. Since the discount factor $\beta < 1$, we know that the linear operator $(I - \beta L)$ is invertible from the geometric series theorem for linear operators (e.g. see Helmberg, 2008, Theorem 4.23.3). Hence we conclude that

$$\pi(x, p) = \mathrm{E}(\xi_{2t} \mid x_t = x, p_t = p) = [(I - \beta L)^{-1}(h)](x, p)$$

is identified. More explicitly, it can be shown that

$$\pi(x, p) = \sum_{r=0}^{\infty} \beta^r \, \mathrm{E}(h(x_{t+r}, p_{t+r}) \mid x_t, p_t).$$

The proof is similar to the proof of Lemma 2 in Chou and Ridder (2017) and omitted herein. Of course, if one is concerned only about $\pi(p_{2t}) \equiv \mathrm{E}(\xi_{2t} \mid p_{2t})$, one may consider the conditional moment equation

$$\mathrm{E}\big[(1 - \beta)y_t + \beta(1 - \beta)w_{t+1} - (1 - \beta)\delta_2 - \xi_{2t} + \beta\xi_{2,t+1} \,\big|\, p_{2t}\big] = 0, \tag{21}$$

and the identification of $\pi(p_{2t})$ follows from similar arguments. The next proposition outlines this result.

Proposition 2. *In addition to the conditions of Proposition 1, suppose Assumption 5-6 hold. We can identify* $\mathrm{E}(\xi_{jt} \mid x_t, p_t)$ *for each product* $j \in \mathcal{J}_t$.

To identify the conditional variance $\mathrm{Var}(\xi_t \mid x_t, p_t)$, we need additional assumptions.

Assumption 7. *(i) The unobserved qualities* $\xi_{1t}, \ldots, \xi_{Jt}$ *are independent conditional on* (x_t, p_t)*;*

(ii) Assume that $\mathrm{Var}(\xi_{1t} \mid x_t, p_t) = \cdots = \mathrm{Var}(\xi_{Jt} \mid x_t, p_t) = \sigma^2(x_t, p_t)$.

The homoskedasticity assumption is not essential. Remark 3 at the end of this subsection discusses the extension with heteroskedasticity.

Using $d_t = \xi_{2t} - \xi_{1t}$, we have

$$\begin{aligned}
\mathrm{E}(d_t^2 \mid x_t, p_t) &= \mathrm{E}(\xi_{2t}^2 \mid x_t, p_t) + \mathrm{E}(\xi_{1t}^2 \mid x_t, p_t) - 2\,\mathrm{E}(\xi_{2t}\xi_{1t} \mid x_t, p_t) \\
&= \mathrm{E}(\xi_{2t}^2 \mid x_t, p_t) + \mathrm{E}(\xi_{1t}^2 \mid x_t, p_t) - 2\,\mathrm{E}(\xi_{2t} \mid x_t, p_t)\,\mathrm{E}(\xi_{1t} \mid x_t, p_t) \\
&= \mathrm{Var}(\xi_{2t} \mid x_t, p_t) + \mathrm{Var}(\xi_{1t} \mid x_t, p_t) + (\mathrm{E}(\xi_{2t} \mid x_t, p_t) - \mathrm{E}(\xi_{1t} \mid x_t, p_t))^2 \\
&= 2\sigma^2(x_t, p_t) + [\mathrm{E}(\xi_{2t} \mid x_t, p_t) - \mathrm{E}(\xi_{1t} \mid x_t, p_t)]^2.
\end{aligned}$$

Since we have identified $\mathrm{E}(\xi_{1t} \mid x_t, p_t)$ and $\mathrm{E}(\xi_{2t} \mid x_t, p_t)$, we identify $\sigma^2(x_t, p_t)$ from the above display.

As for the unconditional variance, we use

$$\mathrm{Var}(\xi_{jt}) = \mathrm{E}(\xi_{jt}^2) = \mathrm{E}\big[\mathrm{E}(\xi_{jt}^2 \mid x_t, p_t)\big].$$

Moreover, $\mathrm{E}(\xi_{jt}^2 \mid x_t, p_t) = \sigma^2(x_t, p_t) + \mathrm{E}(\xi_{jt} \mid x_t, p_t)^2$. Since we have identified $\sigma^2(x_t, p_t)$ and $\mathrm{E}(\xi_{jt} \mid x_t, p_t)$, we identify $\mathrm{E}(\xi_{jt}^2 \mid x_t, p_t)$ and hence $\mathrm{Var}(\xi_{jt})$.

Proposition 3. *In addition to the conditions of Proposition 2, suppose Assumption 7 holds. We then can identify* $\mathrm{Var}(\xi_{jt} \mid x_t, p_t)$ *and* $\mathrm{Var}(\xi_{jt})$.

The fact that we can identify both the conditional mean and variance of ξ_{jt} given (x_t, p_t) is quite useful. By the conditional independence of quality (Assumption 7(i)), we can write

$$F(\xi_{1t}, \ldots, \xi_{Jt} \mid x_t, p_t) = F(\xi_{1t} \mid x_t, p_t) \cdots F(\xi_{Jt} \mid x_t, p_t).$$

If the conditional distribution of ξ_{jt} given x_t, p_t belongs to the location scale family, the conditional mean and variance will determine the distribution of $F(\xi_t \mid x_t, p_t)$.

For two products j and k, if we assume $F(\xi_{jt} \mid x_t, p_t)$ and $F(\xi_{kt} \mid x_t, p_t)$ are "similar" in the following sense, we indeed can nonparametrically identify $F(\xi_{jt} \mid x_t, p_t)$.

Assumption 8. *For any two products j and k, conditional on (x_t, p_t), ξ_{jt} and ξ_{kt} have identical distribution, except for their conditional mean.*

Let $\tilde{\xi}_{jt} = \xi_{jt} - \mathrm{E}(\xi_{jt} \mid x_t, p_t)$. From eq. (19), we have

$$\tilde{\xi}_{2t} - \tilde{\xi}_{1t} = d_t + \mathrm{E}(\xi_{1t} \mid x_t, p_t) - \mathrm{E}(\xi_{2t} \mid x_t, p_t).$$

The two random variables $\tilde{\xi}_{2t}$ and $\tilde{\xi}_{1t}$ are independent and identically distributed conditional on x_t, p_t. We also identify the conditional distribution $F(\tilde{\xi}_{2t} - \tilde{\xi}_{1t} \mid x_t, p_t) = F\big(d_t + \mathrm{E}(\xi_{1t} \mid x_t, p_t) - \mathrm{E}(\xi_{2t} \mid x_t, p_t) \mid x_t, p_t\big)$ because all d_t, x_t, p_t and the conditional means are identified. The distribution function of $F(\tilde{\xi}_{1t} \mid x_t, p_t)$ or equivalently $F(\tilde{\xi}_{2t} \mid x_t, p_t)$ can be obtained from the deconvolution process, when the distribution function $F(\tilde{\xi}_{1t} \mid x_t, p_t)$ is symmetric at zero. Such a deconvolution is called "constrained deconvolution" in statistics (see e.g. Belomestnyi, 2002; Belomestny, 2003). The constraint is that the individual CDFs $F(\tilde{\xi}_{2t})$ and $F(\tilde{\xi}_{1t})$ are identical. Theorem 3 of Belomestnyi (2002) gives the sufficient conditions for determining $F(\tilde{\xi}_{1t} \mid x_t, p_t)$ from $F(\tilde{\xi}_{2t} - \tilde{\xi}_{1t} \mid x_t, p_t)$.

Proposition 4. *In addition to the conditions of Proposition 3, suppose Assumption 8 holds. Let $\varphi(t; x_t, p_t)$ the characteristic function of ξ_{jt} conditional on x_t, p_t. Conditional on x_t, p_t, if ξ_{jt} has absolute moment of order 2, $|\varphi(t; x_t, p_t)| + |\varphi(t; x_t, p_t)'| + |\varphi(t; x_t, p_t)''| \neq 0$, and $F(\tilde{\xi}_{1t} \mid x_t, p_t)$ is symmetric at zero, $F(\xi_{jt} \mid x_t, p_t)$ and $F(\xi_t \mid x_t, p_t)$ are identified.*

Remark 3 (Heteroskedasticity). When we have 3 or more products, we only need to assume that there are at least two products whose conditional variance $\mathrm{Var}(\xi_{jt}|x_t, p_t)$ is the same. To see this, suppose there are 3 products, and $\mathrm{Var}(\xi_{1t} \mid x_t, p_t) = \mathrm{Var}(\xi_{2t} \mid x_t, p_t)$. We have shown how to identify $\mathrm{Var}(\xi_{1t} \mid x_t, p_t)$. To identify $\mathrm{Var}(\xi_{3t} \mid x_t, p_t)$, we simply use $d_{31,t} = \xi_{3t} - \xi_{1t}$. By eq. (6), we have

$$d_{31,t} = (1-\beta)\log(s_{3t}/s_{1t}) - (x_{3t} - x_{1t})'\gamma - (\delta_3 - \delta_1) + (1-\beta)\alpha(p_{3t} - p_{1t}),$$

which is identified. By the same arguments, we have

$$\mathrm{E}(d_{31,t}^2 \mid x_t, p_t) = \mathrm{Var}(\xi_{3t} \mid x_t, p_t) + \mathrm{Var}(\xi_{1t} \mid x_t, p_t) + [\mathrm{E}(\xi_{3t} \mid x_t, p_t) - \mathrm{E}(\xi_{1t} \mid x_t, p_t)]^2.$$

We then identify $\mathrm{Var}(\xi_{3t} \mid x_t, p_t)$ from this display. ■

3.2.2 $F(m_{t+1} \mid m_t)$

Note that $m_t = (x_t, p_t, \xi_t)$ and ξ_t is $J \times 1$ vector. When $J = 1$, the results in Hu and Shum (2012) can be used to show the identification of $F(m_{t+1} \mid m_t)$ with four periods data and additional technical conditions. It is not straightforward to extend Hu and Shum (2012) results to multidimensional cases, and their technical conditions, involving the completeness of some linear operators, are nontrivial to check. If ξ_{jt} is a fixed effect taking a finite number of values, the problem of identifying $F(x_{t+1}, p_{t+1} | x_t, p_t, \xi)$ from $F(x_{t+1}, p_{t+1} | x_t, p_t)$ is related to the finite mixture model (e.g. Allman, Matias, and Rhodes, 2009; Bonhomme, Jochmans, and Robin, 2016). The conditions, which mostly require that the number of sampling periods should be large enough relative to the number of types, are easier to verify. However, it is not straightforward to estimate $F(x_{t+1}, p_{t+1} \mid x_t, p_t, \xi)$ due to the presence of conditioning variables. We are not going to pursue the identification of $F(m_{t+1} \mid m_t)$ in general here. We instead consider the identification under the following restriction. We will discuss how to relax the assumption at the end of this section.

Assumption 9. *Assume the following decomposition holds*

$$F(m_{t+1} \mid m_t) = F(\xi_{t+1} \mid \xi_t) F(x_{t+1} \mid x_t, p_t) F(p_{t+1} \mid x_{t+1}, \xi_{t+1}).$$

Implicitly, we assume that $\xi_{t+1} \perp\!\!\!\perp x_t, p_t | \xi_t$, $x_{t+1} \perp\!\!\!\perp (\xi_t, \xi_{t+1}) | (x_t, p_t)$ *and* $p_{t+1} \perp\!\!\!\perp (x_t, p_t, \xi_t) | (x_{t+1}, \xi_{t+1})$.

These assumptions can be interpreted as follows. At the beginning of period $t+1$, each manufacturer receives its quality ξ_{t+1}, which depends on ξ_t. Meanwhile, x_{t+1} is generated based only on x_t and p_t. Given the quality ξ_{t+1} and x_{t+1} in period $t+1$, manufacturers then determine their prices for period $t+1$.

The component $F(x_{t+1}|x_t, p_t)$ is directly identified from data. The component $F(p_{t+1}|x_{t+1}, \xi_{t+1})$ is identified, because we have previously identified the joint distribution below:

$$F(m_{t+1}) = F(x_{t+1}, p_{t+1})F(\xi_{t+1} \mid x_{t+1}, p_{t+1}) \tag{22}$$

The last to be identified is $F(\xi_{t+1} \mid \xi_t)$.

Assumption 10. *(i) Assume that $F(\xi_{t+1} \mid \xi_t) = F(\xi_{1,t+1} \mid \xi_{1t}) \dots F(\xi_{J,t+1} \mid \xi_{Jt})$, and $\xi_{j,t+1}$ and ξ_{kt} are uncorrelated for any two distinct products j and k.*

(ii) Assume that

$$\xi_{j,t+1} = \phi_j \xi_{jt} + \nu_{j,t+1},$$

where $\nu_{j,t+1}$ has mean zero and is independent of ξ_{jt}.

To identify ϕ_j, we only need $\mathrm{E}(\xi_{j,t+1}\xi_{jt})$. Using eq. (13) with $\eta(m_t) = \xi_{2t} - \xi_{1t}$, we have

$$\mathrm{E}\left[(y_t + \beta w_{t+1})\xi_{2t} - \xi_{1t} - \delta_2\xi_{2t} - \xi_{1t} - \frac{1}{1-\beta}\xi_{2t}\xi_{2t} - \xi_{1t} + \frac{\beta}{1-\beta}\xi_{2,t+1}\xi_{2t} - \xi_{1t}\right] = 0.$$

By assumption 10,

$$\mathrm{E}(\xi_{2,t+1}\xi_{1t}) = 0.$$

So we have the following formula for $\mathrm{E}(\xi_{2,t+1}\xi_{2t})$:

$$\mathrm{E}(\xi_{2,t+1}\xi_{2t}) = \frac{\mathrm{E}(\xi_{2t}^2) - \mathrm{E}(\xi_{2t}\xi_{1t})}{\beta} - \frac{1-\beta}{\beta}\,\mathrm{E}[(y_t + \beta w_{t+1})d_t].$$

By the conditional independence between ξ_{1t} and ξ_{2t} given x_t, p_t (Assumption 7(i)), we have

$$\mathrm{E}(\xi_{2t}\xi_{1t}) = \mathrm{E}[\mathrm{E}(\xi_{2t} \mid x_t, p_t)\,\mathrm{E}(\xi_{1t} \mid x_t, p_t)].$$

Hence $\mathrm{E}(\xi_{2,t+1}\xi_{2t})$ is identified. Note in general one cannot claim $\mathrm{E}(\xi_{2t}\xi_{1t}) = \mathrm{E}(\xi_{2t})\,\mathrm{E}(\xi_{1t})$ because ξ_{1t} (ξ_{2t}) is correlated with the price p_{1t} (p_{2t}), and the prices p_{1t} and p_{2t} are correlated in general. We can then identify $\mathrm{E}(\xi_{1t}\xi_{1,t+1})$ from $\mathrm{E}[(\xi_{2t} - \xi_{1t})(\xi_{2,t+1} - \xi_{1,t+1})] = \mathrm{E}(d_t d_{t+1})$. Previously, we have identified the distribution $F(\xi_{jt} \,|\, x_t, p_t)$. Hence the marginal distribution $F(\xi_{jt})$ is identified. Under the assumption that $\nu_t \perp\!\!\!\perp \xi_{2t}$, we can identify the distribution $F(\nu_t)$ by convolution.

Proposition 5. *In addition to the conditions of Proposition 4, suppose Assumption 9 and 10 hold. We can identify $F(\xi_{t+1} \mid \xi_t)$, henceforth $F(m_{t+1} \mid m_t)$.*

Remark 4. Suppose we can decompose x_t into two parts x_{1t} and x_{2t}, and that x_{2t} is correlated with ξ_t. We can decompose $F(m_{t+1} \mid m_t)$ with

$$F(m_{t+1} \mid m_t) = F(\xi_{t+1} \mid \xi_t) F(x_{1,t+1} \mid x_t, p_t) F(p_{t+1}, x_{2,t+1} \mid x_{1,t+1}, \xi_{t+1}).$$

$F(p_{t+1}, x_{2,t+1} \mid x_{1,t+1}, \xi_{t+1})$ is not a problem since we have identified the joint distribution $F(m_{t+1})$. ■

Remark 5 (Non-terminal choices). Up to now, we have assumed that a consumer's choice is terminal, e.g. "buy one car then exit the market." We can extend the analysis to allow for non-terminal choice, e.g. "lease one car." For simplicity, assume that there is a third product "lease one car". Let the flow utility of the third product be

$$u_{i3t} = x_{3t}' \gamma - \alpha p_{3t} + \delta_3 + \xi_{3t} + \varepsilon_{i3t}.$$

By buying product 3, the consumer remains on the market. Hence the choice specific value function v_{3t} is as follows,

$$v_{3t} = x_{3t}' \gamma - \alpha p_{3t} + \delta_3 + \xi_{3t} + \beta \, \mathrm{E}(\bar{V}_{t+1}(m_{t+1}) \mid m_t).$$

Recall that $v_{0t} = \beta \, \mathrm{E}(\bar{V}_{t+1}(m_{t+1}) \mid m_t)$. We then have

$$\ln(s_{3t}/s_{0t}) = v_{3t} - v_{0t} = x_{3t}' \gamma - \alpha p_{3t} + \delta_3 + \xi_{3t},$$

which is the standard regression model in the BLP model. It is well known that one can identify ξ_{3t} itself in general (Berry and Haile, 2014). Once one has identified ξ_{3t}, the joint distribution $F(p_{3t}, \xi_{3t})$ and the autocorrelation $\mathrm{corr}(\xi_{3t}, \xi_{3,t+1})$ are identified. So non-terminal choice does not break our arguments.

It should be remarked that the main reason why this works is that the current non-terminal choice, "lease one car" today, does not affect the future market state m_{t+1}. Hence the expected future payoff $\mathrm{E}(\bar{V}_{t+1}(m_{t+1}) \mid m_t)$ does not vary with respect to choice. As a result, the payoff difference between the choice of "lease one car" and the choice of "outside good" is simply the flow utility difference. In most dynamic discrete choice with individual level data, this does not hold because in most applications, the current choice affects the transition of state variables; hence the expected future payoff is also alternative specific. ■

4 Estimation

For exposition simplicity, we focus on the case that the data are from one single market, e.g. the US, over T consecutive periods. Both numerical studies and empirical application show the case with multiple markets. Below, we first describe the estimation routine and postpone the variance estimation until the end. The routine involves only IV and linear regressions.

4.1 Preference

4.1.1 Preference for the observed characteristics and price

Step 1: Estimate $(\tilde{\gamma}' \equiv \gamma'/(1-\beta), \alpha)$ using the following moment equation:

$$\mathrm{E}(g_{1,(j,k),t}(\theta_{1o})) = 0, \qquad \text{for} \qquad 0 < j < k \leq J,$$

$$g_{1,(j,k),t}(\theta_1) = z_{(j,k),t}\left[\ln\left(\frac{s_{jt}}{s_{kt}}\right) - (x_{jt} - x_{kt})'\tilde{\gamma} + \alpha(p_{jt} - p_{kt}) - \frac{\delta_j - \delta_k}{1-\beta}\right].$$

The vector $z_{(j,k),t}$ is a vector of IVs that are uncorrelated with $(\xi_{jt} - \xi_{kt})$. The moment equation follows from eq. (7) in identification.

In practice, one can estimate $\tilde{\gamma}$ and α by an IV regression of $\ln(s_{jt}/s_{kt})$ on $(x_{jt} - x_{kt})$ and $(p_{jt} - p_{kt})$ with IV $z_{(j,k),t}$ using data $t = 1, \ldots, T$ and a set of selected pairs of products (j, k). In real data applications, we found that it is desirable to divide the products into a few clusters based on their prices, e.g. run a k-means clustering by price, and consider only inter-cluster pairs of products. The underlying reason is that price difference $p_{jt} - p_{kt}$ is usually endogenous. If two products are close in their price, e.g. they come from the same cluster, the instrument $z_{(j,k),t}$ is likely to be weak.

Letting $\hat{\tilde{\gamma}}$ and $\hat{\alpha}$ be the obtained estimates, define

$$y_{jt} = \ln\left(\frac{s_{jt}}{s_{0t}}\right) - x_{jt}'\tilde{\gamma} + \alpha p_{jt} \qquad \text{and} \qquad w_{jt} = x_{jt}'\tilde{\gamma} - \alpha p_{jt} - \ln s_{jt},$$

and their estimates

$$\hat{y}_{jt} = \ln\left(\frac{s_{jt}}{s_{0t}}\right) - x_{jt}'\hat{\tilde{\gamma}} + \hat{\alpha} p_{jt} \qquad \text{and} \qquad \hat{w}_{jt} = x_{jt}'\hat{\tilde{\gamma}} - \hat{\alpha} p_{jt} - \ln s_{jt}.$$

4.1.2 Discount factor

Step 2: Estimate β using

$$\mathrm{E}(g_{2,(j,0),t}(\theta_{1o})) = 0, \qquad \text{for} \qquad 0 < j < J,$$

where

$$g_{2,(j,0),t}(\theta_1) = x_{(j,0),t,IV}(y_{jt} + \beta w_{j,t+1} - \delta_j).$$

The above moment equation follows from eq. (14) in identification. In practice, to estimate β, one simply runs an IV regression of $\hat{y}_{jt}$ on $-\hat{w}_{j,t+1}$ using $x_{jt,IV}$ as the IV for $\hat{w}_{j,t+1}$ using data $t = 1, \ldots, T-1$ and $j = 1, \ldots, J$.

4.1.3 Expected unobserved product fixed effect

Step 3: Estimate δ_j using

$$\mathrm{E}(y_{jt} + \beta w_{j,t+1} - \delta_j) = 0,$$

which corresponds to the above moment equation when $x_{(j,0),t,IV} = 1$. In practice, one runs a linear regression for each j of $(\hat{y}_{jt} + \hat{\beta}\hat{w}_{j,t+1})$ on a constant of one using data from $t = 1, \ldots, T-1$.

Define $\hat{d}_{(j,k),t}$, which will be used in the estimation of the other parameters,

$$\hat{d}_{(j,k),t} = (1-\hat{\beta})\left[\ln\left(\frac{s_{jt}}{s_{kt}}\right) - (x_{jt} - x_{kt})'\hat{\gamma} + \hat{\alpha}(p_{jt} - p_{kt}) - \frac{\hat{\delta}_j - \hat{\delta}_k}{1-\hat{\beta}}\right].$$

4.2 $F(m_t)$ and $F(m_{t+1} \mid m_t)$

The full nonparametric estimation of $F(m_t)$ and $F(m_{t+1} \mid m_t)$ would be unreliable in a small sample, which is the case in most applications using market level data. We consider the assumption of normal distribution to simplify the problem while keeping the interesting dynamics and joint dependence among m_t. For exposition simplicity, we assume that the distribution $F(x_t, p_t)$ and $F(x_{t+1} \mid x_t, p_t)$ are known.

Assumption 11. *(i)* $x_t \perp\!\!\!\perp \xi_t \mid p_t$ *and* $\xi_{t+1} \perp\!\!\!\perp p_t \mid p_{t+1}$.

(ii) Assume the necessary conditional independence so that

$$F(\xi_{1t}, \ldots, \xi_{Jt} \mid p_{1t}, \ldots, p_{Jt}) = F(\xi_{1t} \mid p_{1t}) \cdots F(\xi_{Jt} \mid p_{Jt}).$$

In particular, this implies $\xi_{jt} \perp\!\!\!\perp p_{kt} \mid p_{jt}$ *for* $j \neq k$.

(iii) For each product j*, assume that* (p_{jt}, ξ_{jt}) *follows bivariate normal distribution:*

$$\begin{pmatrix} p_{jt} \\ \xi_{jt} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{pjt} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{pjt} & \rho_j \sigma \sigma_{pjt} \\ & \sigma^2 \end{pmatrix}\right).$$

Let

$$\tilde{p}_{jt} = (p_{jt} - \mu_{pjt})/\sigma_{pjt},$$

be the standardised price. The bivariate normal distribution implies that

$$\mathrm{E}(\xi_{jt} \mid p_{jt}) = \rho_j \sigma \tilde{p}_{jt}.$$

This also implies that $\nu_{j,t+1}$ *in the AR(1) process* $\xi_{j,t+1} = \phi_j \xi_{jt} + \nu_{j,t+1}$ *follows a normal distribution.*

Given this assumption, the primary interests are to estimate $\sigma^2 = \mathrm{Var}(\xi_{jt})$, $\rho_j = \mathrm{corr}(p_{jt}, \xi_{jt})$, and ϕ_j. However, it is easier to estimate $\tilde{\rho}_j = \rho_j \sigma$ as a whole parameter. Let $\theta_2 = (\tilde{\rho}_1, \ldots, \tilde{\rho}_J, \sigma, \phi_1, \ldots, \phi_J)'$ and $\theta = (\theta_1', \theta_2')'$.

4.2.1 Correlation between product price and unobserved product characteristic

Step 4: estimate $\tilde{\rho}_j \equiv \rho_j \sigma$ using

$$\mathrm{E}(g_{3,j,t}(\theta_o)) = 0, \qquad \text{for} \qquad 0 < j \leq J,$$

where

$$g_{3,j,t}(\theta) = z_{\rho,jt}(p_{jt}) r_{jt}, \tag{23}$$

$$r_{jt} = (1-\beta)(y_{jt} + \beta w_{j,t+1}) - (1-\beta)\delta_j - \tilde{\rho}_j(\tilde{p}_{jt} - \beta \tilde{p}_{j,t+1})$$

Here $z_{\rho,jt}(p_{jt})$ is a vector of functions of p_{jt}, e.g. $z_{\rho,jt}(p_{jt}) = (p_{jt}, p_{jt}^2, \ldots, p_{jt}^K)'$ for some integer $K \geq 1$. We discuss the optimal choice of $z_{\rho,jt}(p_{jt})$ below. In practice, one can estimate $\tilde{\rho}_j$ for each j by an IV regression of $(1-\hat{\beta})(\hat{y}_{jt} + \hat{\beta}\hat{w}_{j,t+1})$ on $(\tilde{p}_{jt} - \hat{\beta}\tilde{p}_{j,t+1})$ with IV $z_{\rho,jt}$.

It should be remarked that in practice $\tilde{\rho}_j$ is more difficult to estimate than $\theta_1 = (\alpha, \beta, \gamma', \delta')'$ for three reasons. First, to estimate $\tilde{\rho}_j$, one has only $T-1$ number of observations. Second, the sampling error in estimating θ_1 impacts the estimation of $\tilde{\rho}_j$. Third, the variance of $\tilde{\rho}_j$ is proportional to the inverse of the variance of $(\tilde{p}_{jt} - \beta\tilde{p}_{j,t+1})$. When price is persistent over time, the variance of $(\tilde{p}_{jt} - \beta\tilde{p}_{j,t+1})$ is small.

The above moment condition was derived from the following arguments. We have

$$\mathrm{E}\big[(1-\beta)(y_{jt} + \beta w_{j,t+1}) - (1-\beta)\delta_j - \xi_{jt} + \beta\xi_{j,t+1} \,\big|\, p_{jt}\big] = 0,$$

from eq. (21) in identification. By the bivariate normal assumption and the assumption $\xi_{j,t+1} \perp\!\!\!\perp p_{jt} \mid p_{j,t+1}$, we have

$$\mathrm{E}(\xi_{jt} \mid p_{jt}) = \rho_j \sigma \tilde{p}_{jt} = \tilde{\rho}_j \tilde{p}_{jt},$$

$$\mathrm{E}(\xi_{j,t+1} \mid p_{jt}) = \mathrm{E}(\mathrm{E}(\xi_{j,t+1} \mid p_{j,t+1}) \mid p_{jt}) = \mathrm{E}(\rho_j \sigma \tilde{p}_{j,t+1} \mid p_{jt}) = \mathrm{E}(\tilde{\rho}_j \tilde{p}_{j,t+1} \mid p_{jt}).$$

Hence $\mathrm{E}(r_{jt} \mid p_{jt}) = 0$. Given the conditional moment equation, it is well known (see Newey, 1993) that the optimal instrument is

$$z(p_{jt}) = \frac{\tilde{p}_{jt} - \beta\,\mathrm{E}(\tilde{p}_{j,t+1} \mid p_{jt})}{\Sigma_j(p_{jt})},$$

where

$$\Sigma_j(p_{jt}) = \mathrm{E}(r_{jt}^2 \mid p_{jt}).$$

In practice, the optimal instrument can be replaced by its sample analog. First, $\mathrm{E}(\tilde{p}_{j,t+1} \,|\, p_{jt})$ can be replaced by the fitted value of a nonparametric regression or polynomial regression of $\tilde{p}_{j,t+1}$ on p_{jt} depending on the sample size. To estimate $\Sigma_j(p_{jt})$, run a linear regression of $(1-\hat{\beta})(\hat{y}_{jt}+\hat{\beta}\hat{w}_{j,t+1})$ on $(\tilde{p}_{jt}-\hat{\beta}\tilde{p}_{j,t+1})$ for each product j. Denote $\hat{r}_{jt}$ the residuals from such a linear regression. Then $\Sigma_j(p_{jt})$ can be estimated by the fitted value of a nonparametric regression or polynomial regression of $\hat{r}^2_{jt}$ on p_{jt} for each product j. Note, if one is willing to accept that $\mathrm{E}(r^2_{jt} \,|\, p_{jt}) = \mathrm{E}(r^2_{jt})$, $\Sigma_j(p_{jt})$ can then be estimated by the sample variance of $\hat{r}_{jt}$.

4.2.2 Variance of unobserved product characteristic

Step 5: estimate σ using

$$\mathrm{E}(g_{4,(j,k),t}(\theta_o)) = 0, \qquad \text{for} \qquad 0 < j < k \leq J,$$

$$g_{4,(j,k),t}(\theta) = d^2_{(j,k),t}/2 + \tilde{\rho}_j\tilde{\rho}_k\tilde{p}_{jt}\tilde{p}_{kt} - \sigma^2.$$

In practice, one can run a linear regression of $\hat{d}^2_{(j,k),t}/2 + \hat{\tilde{\rho}}_j\hat{\tilde{\rho}}_k\tilde{p}_{jt}\tilde{p}_{kt}$ on a constant one using data $t = 1, \ldots, T$ and all selected pair of products. Knowing $\tilde{\rho}_j = \rho_j\sigma$ and σ, we know the joint distribution of (ξ_{jt}, p_{jt}).

The above moment condition holds because

$$\begin{aligned}
\mathrm{E}\big(d^2_{(j,k),t}\big) = \mathrm{E}\big[(\xi_{jt}-\xi_{kt})^2\big] &= 2\sigma^2 - 2\,\mathrm{E}(\xi_{jt}\xi_{kt}) \\
&= 2\sigma^2 - 2\,\mathrm{E}[\mathrm{E}(\xi_{jt} \mid p_{jt})\,\mathrm{E}(\xi_{kt} \mid p_{kt})] \\
&= 2\sigma^2 - 2\tilde{\rho}_j\tilde{\rho}_k\,\mathrm{E}(\tilde{p}_{jt}\tilde{p}_{kt}).
\end{aligned}$$

From above, we also have

$$\frac{\mathrm{E}\Big(d^2_{(j,k),t}\Big)}{2\sigma^2} = 1 - \frac{\mathrm{E}(\xi_{jt}\xi_{kt})}{\sigma^2}. \tag{24}$$

4.2.3 Serial correlation of unobserved product characteristic

Step 6: estimate ϕ_j using

$$\mathrm{E}(g_{5,(j,k),t}(\theta_o)) = 0, \qquad \text{for} \qquad k \neq j,$$

$$g_{5,(j,k),t}(\theta) = \frac{d^2_{(j,k),t}}{2\beta\sigma^2} - \frac{1-\beta}{\beta}(y_{jt}+\beta w_{j,t+1})\frac{d_{(j,k),t}}{\sigma^2} - \phi_j.$$

In practice, one can run a linear regression for each j of $\hat{d}^2_{(j,k),t}/(2\hat{\beta}\hat{\sigma}^2) - [(1-\hat{\beta})/\hat{\beta}](\hat{y}_{jt} + \hat{\beta}\hat{w}_{j,t+1})\hat{d}_{(j,k),t}/\hat{\sigma}^2$ on a constant one using data $t = 1, \ldots, T-1$ and the selected pairs of products.

The above moment equation follows from

$$\phi_j = \mathrm{cov}(\xi_{j,t+1}, \xi_{jt})/\sigma^2 = \mathrm{E}(\xi_{j,t+1}\xi_{jt})/\sigma^2.$$

We have

$$\mathrm{E}(\xi_{j,t+1}\xi_{jt}) = \frac{\sigma^2 - \mathrm{E}(\xi_{jt}\xi_{kt})}{\beta} - \frac{1-\beta}{\beta}\,\mathrm{E}\big[(y_{jt} + \beta w_{j,t+1})d_{(j,k),t}\big], \quad \text{for} \quad k \neq j.$$

By eq. (24), we have

$$\phi_j = \frac{1}{\beta}\frac{\mathrm{E}\Big(d^2_{(j,k),t}\Big)}{2\sigma^2} - \frac{1-\beta}{\beta}\,\mathrm{E}\bigg((y_{jt} + \beta w_{j,t+1})\frac{d_{(j,k),t}}{\sigma^2}\bigg), \qquad \text{for} \qquad k \neq j.$$

Knowing ϕ_j and σ^2, we know the distributional properties of the AR(1) process of ξ_{jt}.

4.3 Asymptotic Variance

To derive the asymptotic variance, let $g_t(\theta)$ be a vector of functions from stacking $g_{1,(j,k),t}(\theta_1)$, $g_{2,(j,0),t}(\theta_1)$, $g_{3,j,t}(\theta)$, $g_{4,(j,k),t}(\theta)$, and $g_{5,(j,k),t}(\theta)$. Then the estimation problem can be viewed as a GMM problem with moment equation $\mathrm{E}(g_t(\theta)) = 0$. Asymptotically, the above sequential estimator gives the same estimates as minimizing a GMM objective function using moment functions $g_t(\theta)$ and certain weighting matrix W. In applications, we recommend using the sequential approach because it does not involve complicated numerical optimization. Moreover, the sequential approach consists of individual IV regressions, and modern software routines, like `ivreg` in `R`, provide various specification tests that can guide our choice of instrumental variables and observed product characteristics.

By standard GMM results, the asymptotic variance of $\hat{\theta}$ is then

$$(G'WG)^{-1}G'W\Omega W'G(G'WG)^{-1},$$

where W is a weighting matrix such that $G'WG$ is non-singular,

$$G = \mathrm{E}\bigg(\frac{\partial g_t(\theta)}{\partial \theta}\bigg), \qquad \text{and} \qquad \Omega = \mathrm{E}(g_t g_t').$$

We did not write the usual optimal GMM variance matrix $(G'\Omega^{-1}G)^{-1}$ because Ω here might be singular when some moment functions are collinear. Taking moment function $g_{1,(j,k),t}(\theta_1)$ for example, if $z_{(j,k),t} = z_{(k,j),t}$, $g_{1,(j,k),t}(\theta_1) = -g_{1,(k,j),t}(\theta_1)$.

Without loss of generality, let $g_t(\theta)' = (g_{at}(\theta)', g_{bt}(\theta)')$ and

$$\Omega = \begin{pmatrix} \mathrm{E}(g_{at}g_{at}') & \mathrm{E}(g_{at}g_{bt}') \\ \mathrm{E}(g_{bt}g_{at}') & \mathrm{E}(g_{bt}g_{bt}') \end{pmatrix} = \begin{pmatrix} \Omega_a & \Omega_{ab} \\ \Omega_{ba} & \Omega_b \end{pmatrix}.$$

Assume that Ω_a is full-rank, and $\text{rank}(\Omega) = \text{rank}(\Omega_a)$. That is g_{bt} is redundant given the linearly independent moment functions g_{at}. The covariance matrix Ω is singular here. By letting

$$W = \begin{pmatrix} \Omega_a^{-1} & 0 \\ 0 & 0 \end{pmatrix},$$

one can show that

$$(G'WG)^{-1}G'W\Omega W'G(G'WG)^{-1} = (G_a'\Omega_a^{-1}G_a)^{-1},$$

where $G_a = \text{E}(\partial g_{at}(\theta)/\partial\theta)$, which is the optimal covariance matrix using only the linearly independent moment conditions $g_{at}(\theta)$.

The exact form of $\partial g_t(\theta)/\partial\theta$ depends on the selected pairs of products (j,k) used in the estimation. **??** contains the formulas for $\partial g_{1,(j,k),t}(\theta_1)/\partial\theta$, $\partial g_{2,(j,0),t}(\theta_1)/\partial\theta$, $\partial g_{3,j,t}(\theta)/\partial\theta$, $\partial g_{4,(j,k),t}(\theta)/\partial\theta$, and $\partial g_{5,(j,k),t}(\theta)/\partial\theta$, which are useful calculating the exact variance formulas.

5 Counterfactual Implementation

The estimation of consumer preference parameters did not require the direct computation of the value function nor require an assumption about how consumers form future beliefs. However, in order to run any type of counterfactual analysis, the researcher is required to compute the ex-ante value function based on the estimated parameters or the outside option's marketshare, which is a function of the ex-ante value function as we illustrate below.[6] Specifically, we propose a method to simulate counterfactual market shares that does not require value function iteration, the discretization of state variables, nor the use of interpolation to approximate the ex-ante value function.

Our counterfactual simulation method is implemented in two steps. The first step recovers the counterfactual impact on within market shares, relative to a given product. This step thus captures the competitive substitution effects between products and does not depend on consumer beliefs in our model specification. The second step moves beyond the competitive effects and determines the impact on the outside market share. The second step allows the researcher to quantify the impact on overall demand, and evaluates whether the counterfactual change leads to expansion or contraction of overall demand.

[6] We note that the ex-ante value function or equivalent is also generally required for any data generation process that involves obtaining choices.

5.1 Procedure

We consider the counterfactual change of a current product characteristic x_{jt} to counterfactual x^c_{jt} without changing product fixed effect, δ_j, or unobserved product characteristic ξ_{jt}. Other counterfactuals, such as changes to the distribution of state variables, can be addressed similarly. As is standard in structural models, we assume the counterfactual does not affect consumers' preference, product fixed effects and unobserved characteristics. Hence we use the estimated coefficients and unobservable residuals $(\alpha, \beta, \gamma, \delta_j, \xi_{jt})$. In the sequel, we use superscript "c" to denote counterfactual objects, e.g. s^c_{jt} denotes counterfactual market share of product j. We also assume the counterfactual price p^c_{jt} is held constant.

The first step is to generate the counterfactual within or relative market share. By eq. (6), we have counterfactual relative market share as a function of counterfactual (x^c_{jt}, p^c_{jt}),

$$\ln\left(\frac{s^c_{jt}}{s^c_{1t}}\right) = (x^c_{jt} - x^c_{1t})'\tilde{\gamma} - \alpha(p^c_{jt} - p^c_{1t}) + \frac{\delta_j - \delta_1}{1-\beta} + \frac{\xi_{jt} - \xi_{1t}}{1-\beta}.$$

After estimation of $(\alpha, \beta, \tilde{\gamma}, \delta_j, (\xi_{jt} - \xi_{1t}))$, we are able to express s^c_{jt}/s^c_{1t} as a known function of $(x^c_{jt}, p^c_{jt}, \delta_j, \xi_{jt})$. For simplicity of exposition, let

$$\tilde{s}^c_{jt} = s^c_{jt}/s^c_{1t},$$

and let m^c_t denote the vector of all counterfactual state variables.

We can express the counterfactual market share s^c_{1t} as a function of the counterfactual relative market shares and the counterfactual outside market share s^c_{0t}:

$$s^c_{1t} = \frac{1 - s^c_{0t}(m^c_t)}{\sum_{j=1}^J \tilde{s}^c_{jt}(m^c_t)}. \tag{25}$$

We write $s^c_{0t}(m^c_t)$ to emphasize that the counterfactual outside market share s^c_{0t} is a function of counterfactual market state variables.

The second step finds the counterfactual outside market share $s^c_{0t}(m^c_t)$ from the following equation:

$$\ln\left(\frac{1 - s^c_{0t}(m^c_t)}{s^c_{0t}(m^c_t)}\right) = \lambda(m^c_t) + \beta\,\mathrm{E}\Big[\ln(1 - s^c_{0,t+1}(m^c_{t+1}))\;\Big|\; m^c_t\Big], \tag{26}$$

where

$$\lambda(m^c_t) = \ln\left(\sum_{j=1}^J \tilde{s}^c_{j,t}\right) - \beta\,\mathrm{E}\left[\ln\left(\sum_{j=1}^J \tilde{s}^c_{j,t+1}\right)\;\middle|\; m^c_t\right] + v^c_{1t}(m^c_t) - \beta\,\mathrm{E}\big(v^c_{1,t+1}(m^c_{t+1}) \mid m^c_t\big),$$

$$v^c_{1t}(m^c_t) = x^c_{1t}\tilde{\gamma} - \alpha p^c_{1t} + \frac{\delta_1}{1-\beta} + \frac{\xi_{1t}}{1-\beta}. \tag{27}$$

From the first step, we have determined $\ln\left(\sum_{j=1}^{J}\tilde{s}_{j,t}^{c}\right)$. If ξ_{1t} was known, $v_{1t}^{c}(m_t^c)$ and hence $\lambda(m_t^c)$, are known as well. We discuss how to determine ξ_{1t} below.

Equation (26) follows from eq. (11), from which we have

$$\ln\left(\frac{s_{1t}^{c}}{s_{0t}^{c}}\right) - v_{1t}^{c}(m_t^c) = -\beta\,\mathrm{E}\Big(v_{1,t+1}^{c}(m_{t+1}^{c}) - \ln s_{1,t+1}^{c} \,\Big|\, m_t^c\Big).$$

Substituting s_{1t}^{c} above with its formula from eq. (25), we get eq. (26). For a stationary dynamic programming problem, $s_{0t}^{c}(m_t^c)$ is a time invariant function. Equation (26) is then an integral equation of s_{0t}^{c}, from which one solve s_{0t}^{c}.

5.2 Dimension reduction and other details

In many applications, the dimension of the market state variables m_t^c is proportional to the number of states per product with the number of products as an exponential, and could be computationally infeasible to solve. The curse of dimensionality could arise if either the number of products or observed characteristics is large. For example, in our mobile phone application, there are 7 brands and 9 product characteristics (including 7 product features, price and 1 unobservable characteristic), leading to a $9 \times 7 = 63$-dimensional continuous state space. Thus, if we discretize the continuous variables and represent them each with n points, the dimension of the state space m_t^c is n^{63}. Thus, if we choose $n = 10$, we have 10^{63} points in the state space. Solving this problem with value function iteration, for example, becomes computationally infeasible.

Thus, we consider using alternative approaches to computing the value function. Traditionally, researchers assume consumers track all state variables, but as noted above this leads to a curse of dimensionality. One widely known approach that eliminates this problem is to assume consumers track the inclusive value as the relevant state variable (Melnikov, 2013; Gowrisankaran and Rysman, 2012) so that consumers make choices based on the evolution of the inclusive value. An alternative and less restrictive option as it does not rely on the inclusive value sufficiency assumption (which implies that if two different states have the same option value, then they also have the same value function) is to assume consumers track the conditional value function v_{jt} of all products. Thus, the state space in this latter example is of dimension J. This is more general than the inclusive value assumption, since the inclusive value is a deterministic function of the conditional values of all products. Broadly speaking, our counterfactual approach could accommodate any conceivable set of assumptions that can be used to generate the consumer choice data. Depending on the application context, different methods might be more or less suitable.

Below we reduce the dimension by replacing m_t^c with $(v_{1t}^c, \ldots, v_{Jt}^c)$, which is defined by eq. (27). Then eq. (26) reads

$$\ln\left(\frac{1-s_{0t}^c(v_{1t}^c,\ldots,v_{Jt}^c)}{s_{0t}^c(v_{1t}^c,\ldots,v_{Jt}^c)}\right) = \lambda(v_{1t}^c,\ldots,v_{Jt}^c)+\beta\,\mathrm{E}\Big[\ln(1-s_{0,t+1}^c(v_{1,t+1}^c,\ldots,v_{J,t+1}^c))\Big|v_{1t}^c,\ldots,v_{Jt}^c\Big], \tag{28}$$

In practice the conditional expectation terms in the above display and $\lambda(v_{1t}^c, \ldots, v_{Jt}^c)$ can be estimated by nonparametric regression. Because $s_{0t}^c(v_{1t}^c, \ldots, v_{Jt}^c)$ is conditional probability of choice, one can use the series logit method in the treatment effects literature (Hirano, Imbens, and Ridder, 2003) to approximate it:

$$s_0(v_{1t}^c,\ldots,v_{Jt}^c;\rho) = \frac{\exp(\psi(v_{1t}^c,\ldots,v_{Jt}^c)'\rho)}{1+\exp(\psi(v_{1t}^c,\ldots,v_{Jt}^c)'\rho))}, \tag{29}$$

where $\psi(v_{1t}^c, \ldots, v_{Jt}^c)$ is a vector of known approximating functions, e.g. polynomials, of $v_{1t}^c, \ldots, v_{Jt}^c$. We use this functional form for convenience since the market share is bounded, i.e. $s_0 \in [0, 1]$, and other functions that constrains it in such a manner would be applicable as well. We then use eq. (28) to find ρ, e.g. by least squares, to recover the counterfactual outside market share. Once we calcuate the counterfactual outside market share, we can determine s_{jt}^c from eq. (25).

We also need to know ξ_{1t}, which appears in $\lambda(v_{1t}^c, \ldots, v_{Jt}^c)$, in order to implement the above procedure. There are two different ways to implement this, which trade off an additional assumption for computational simplicity.

First, we could simulate ξ_{1t} from the estimated AR(1) process of ξ_{1t} and implement the above procedure for each draw to obtain a range of counterfactual results. This does not require a further assumption. The second is to take an alternative approach that uses the following formula for ξ_{1t}, which follows from eq. (11),

$$\left(\frac{1-\beta\phi_1}{1-\beta}\right)\xi_{1t} = y_{1t} - \delta_1 + \beta\,\mathrm{E}(w_{1,t+1} \mid x_t, p_t, \xi_t). \tag{30}$$

If we assume that

$$\mathrm{E}(w_{1,t+1} \mid x_t, p_t, \xi_t) = \mathrm{E}(w_{1,t+1} \mid x_t, p_t, \xi_{2t} - \xi_{1t}, \ldots, \xi_{Jt} - \xi_{1t}), \tag{31}$$

we can identify and estimate ξ_{jt}, because $\xi_{jt} - \xi_{1t}$ is identified (c.f. eq. (19)). When ξ_{1t} and p_{1t} is highly correlated, the bias (difference between the left-hand side and the right-hand side in the above display) is expected to be small. The extreme case is when (p_{1t}, ξ_{1t}) follow a bivariate normal distribution (as we assumed in estimation), and their correlation coefficient is one. In this extreme case, knowing p_{1t} is equivalent to knowing ξ_{1t}, hence eq. (31) holds.

6 Empirical Application

We now examine an empirical setting in which we use the method previously proposed to obtain estimates of preferences as well as other market or product-level factors including the correlation between price and the unobservable product characteristic, and serial correlation in the unobservable product characteristics. We focus on the market for mobile phone hardware in the period June 2007–May 2008 (12 months). For this setting, we use data from the top 10 states across the United States, with each of the states serving as markets. The top 6 brands overall are chosen as separate products, and all other brands are included in a generic Other brand choice.

6.1 Data

We have a number of product features at the brand level for each of these markets. The features vary both temporally as well as across markets. These product characteristics are averaged at the brand choice level across products within the brand for each market and period.

Table 1 shows the basic summary statistics of the market by showing the mean of product characteristics for each brand in the sample.[7] The top brands in the market include Apple (iPhone), RIM (Blackberry), Samsung, LG, Nokia, Motorola and Others. This market displays differentiation among the brands with the first two brands arguably represent smartphones whereas the rest were primarily focused on feature phones (or dumbphones) during this time frame. The "x" variables are observable product characteristics, and include indicator variables for the presence of Bluetooth support (xblue), GPS capability (xgps), presence of a physical qwerty keyboard (xqwerty), whether music capability was supported (xmusic), and Wi-Fi support (xwifi). The two numeric variables characterized the weight of the device in ounces (xweight), and the talktime in hours (xtalktime). Typical battery life was measured in hours of talktime, which does not seem to be the case at present. Recall also that most phones at the time were feature phones (not smartphones), and typical phones only had a numeric keypad with 10 buttons, rather than the full QWERTY keyboard.

The price shows significant variation, with very low-priced as well as high-priced models over \$400. A majority of the phones at the time did have Bluetooth support, but not GPS support. QWERTY keyboards were not as prevalent, with the exception of Blackberry, which was well known for this feature. About 50% of the phones had some degree of music support,

[7] Due to a non-disclosure agreement, we cannot report brand-level price and market share data.

Table 1: Summary Statistics for Mobile Phone Data: Mean of Characteristics

Brand	price	xblue	xgps	xweight	xqwerty	xmusic	xwifi	xtalktime	share	n of obs
Other		0.34	0.29	3.36	0.16	0.24	0.04	4.41		120
Moto		0.57	0.44	3.42	0.03	0.35	0.00	4.83		120
Samsung		0.61	0.35	2.98	0.10	0.43	0.01	4.31		120
LG		0.68	0.68	3.50	0.18	0.55	0.00	4.05		120
Nokia		0.41	0.09	3.28	0.01	0.35	0.00	4.33		120
Blackberry		0.87	0.43	3.59	0.87	0.73	0.03	4.07		119
Apple		1.00	0.04	4.50	0.00	1.00	1.00	7.87		111
All	131.86	0.64	0.34	3.50	0.19	0.52	0.15	4.80	18.9	

but some of this support was tied in to carrier-based music services (like downloading tones) which were quite expensive, since customers of a carrier like Verizon or AT&T were seen as a captive market. Surprisingly, except for iPhone, the majority of the phones did not support Wi-Fi, a feature which is taken for granted in the present market. Phones weighed an average of 350 ounces (100 grams), and lasted for about 5 hours of talktime before the battery was depleted.

The correlation between the product characteristics is detailed in Table 2. Here, we examine correlation in features for products (brands) across markets and time periods. We find that product characteristics are positively correlated, with the following exceptions. Weight and GPS seem to be negatively correlated, which is somewhat surprising since we might expect them to be positive. However, we observe that some larger phones were already high-priced and left out the GPS feature. Talktime (or battery life) is also negatively correlated with GPS and the presence of a QWERTY keyboard.

6.2 Model

We model J products in each market and time period. Consumers are indexed by i, products by j, markets (States) by ℓ and periods by t. The period utility for a consumer i making a purchase of product j in market ℓ at time t is:

$$u_{ij\ell t} = \delta_j + x'_{j\ell t}\gamma + \alpha p_{j\ell t} + \xi_{j\ell t} + \varepsilon_{ij\ell t}.$$

After purchasing, he receives flow utility $\delta_j + x'_{j\ell t}\gamma + \xi_{j\ell t} + \varepsilon_{ij\ell t}$. The "no purchase" option is modeled as receiving a period utility of 0, with an option to continue in the market as in §2. Consumers who purchase exit the market, and thus can be modeled as receiving

Table 2: Correlation between Product Characteristics for Mobile Phone Data

	xblue	xgps	xweight	xqwerty	xmusic	xwifi	xtalktime
xblue	1.000	0.093	0.661	0.396	0.901	0.563	0.494
xgps	–	1.000	−0.157	0.303	−0.068	−0.437	−0.516
xweight	–	–	1.000	0.080	0.707	0.783	0.738
xqwerty	–	–	–	1.000	0.315	−0.226	−0.336
xmusic	–	–	–	–	1.000	0.680	0.591
xwifi	–	–	–	–	–	1.000	0.947
xtalktime	–	–	–	–	–	–	1.000

the discounted stream of future utilities immediately upon purchase. Thus they obtain in expectation $(\delta_j + \gamma X_{j\ell t} + \xi_{j\ell t})/(1-\beta) + \alpha P_{j\ell t}$. Consumers who do not purchase continue in the market and receive the expected discounted value of waiting or $\beta\,\mathrm{E}(\bar{V}(m_{t+1}) \mid m_t)$.

The estimation follows the multi-step procedure described in §4 above. The standard error (SE) and t values were obtained from the GMM variance formula.

6.3 Results

The results of the estimation are detailed in table 3. There are a few noteworthy observations regarding the first step IV regression results. We exclude price and music as potentially endogenous variables and use the other product characteristics as instruments in the IV regression. We also use additional instruments obtained as the mean product characteristics and price for comparison products in *other* markets. These comparison products are chosen by a clustering process, where Apple and RIM (Blackberry) are grouped in one cluster, which could be interpreted as the smartphone cluster, other well regarded brands of feature phones at the time are grouped in a second cluster (Motorola, Samsung, LG and Nokia), and finally all other brands are grouped in a third cluster. For a product, the products in other clusters serve as comparison products in order to provide a sufficient degree of variation.

First, we observe that the price and all the product characteristics are significant in the regression. The relative sales response to product characteristics is positive for Bluetooth and GPS, but negative for weight and music. Wi-Fi capabilities as well as talk time (which measures effective battery life) are also positive as we might expect. It might seem that the result about music is somewhat counterintuitive; however, there are two contextual reasons that help understand this effect. First, recall that in 2007, music capabilities of most phones

Table 3: Estimation Results of Mobile Phone Market

	Parameters	Estimate	Std. Error	t value
Step 1: preference, $\gamma/(1-\beta)$	price	-0.01	0.001	-19.2
	xblue	5.37	0.387	13.9
	xgps	0.79	0.213	3.7
	xweight	-0.37	0.095	-3.9
	xqwerty	1.03	0.169	6.1
	xmusic	-8.63	0.454	-19.0
	xwifi	3.37	0.379	8.9
	xtalktime	0.29	0.061	4.7
Step 2: discount factor	β	0.79	0.006	122.1
Step 3: fixed effect	δ_{Moto}	-0.36	0.043	-8.4
	$\delta_{Samsung}$	-0.43	0.042	-10.2
	δ_{LG}	-0.29	0.039	-7.5
	δ_{Nokia}	-0.38	0.053	-7.2
	$\delta_{Blackberry}$	-0.55	0.043	-12.9
	δ_{Apple}	-0.03	0.050	-0.5
	δ_{Other}	-0.34	0.038	-8.8
Step 4: correlation between price and unobserved product characteristics	ρ_{Moto}	0.27	0.025	10.6
	$\rho_{Samsung}$	0.21	0.021	10.0
	ρ_{LG}	0.38	0.034	11.0
	ρ_{Nokia}	0.28	0.025	11.2
	$\rho_{Blackberry}$	0.53	0.050	10.6
	ρ_{Apple}	0.89	0.079	11.2
	ρ_{Other}	0.25	0.023	10.8
Step 5: std. error of ξ_{jt}	σ	0.29	0.003	93.2
Step 6: autocorrelation of ξ_{jt}	ϕ_{Moto}	0.63	0.041	15.3
	$\phi_{Samsung}$	0.96	0.041	23.4
	ϕ_{LG}	0.85	0.049	17.2
	ϕ_{Nokia}	0.57	0.040	14.2
	$\phi_{Blackberry}$	-0.44	0.089	-5.0
	ϕ_{Apple}	0.32	0.145	2.2
	ϕ_{Other}	0.46	0.015	30.8

were very rudimentary, and they typically did not support well known MP3 music format, and capabilities of streaming with Spotify or other Internet services were also unavailable. Second, many consumers who cared about music owned iPods or other dedicated music (MP3) players, and phones were really seen as a rather poor substitute for these until the iPhone became popular over the years. We tested for weak instruments and did not find this in our setting.

The coefficients of product characteristics are scaled by $1/(1-\beta)$. Thus, the first step results in table 3 do not directly depend on β. However, obtaining the appropriately scaled coefficients of the product characteristics requires us to either assume or estimate β.

Step 2 of table 3 provides the estimate of β, which is the (negative of) coefficient of w_{t+1} in step 2 detailed in §4. We find that $\hat{\beta} \approx 0.8$, and it is highly significant. For our monthly data, $\beta = 0.8$ implies that after 24 months, which is the typical length of cell phone contract in the US, the cell phone has no additional utility, $\beta^{24} = 0.8^{24} = 0.0047$, for consumers.

Having obtained the discount factor, we proceed with estimating the product fixed effects, which are detailed in §4. The fixed effects are detailed in step 3 of table 3.

We find that the most negative fixed effect is for Blackberry (RIM), followed by Samsung and the other feature phone manufacturers. Apple has the highest fixed effect of all firms.

Finally, we examine the remaining set of all parameter estimates in Steps 4-6 of Table 3. We have previously described the product characteristics, discount factor as well as the fixed effects for the products. We now focus attention on the dynamics of the state transition process, as detailed in §2. The correlation between the product price and the structural error (or unobserved product characteristic) is captured by ρ_j for product j. We find that all these correlations are positive, and Apple has the highest such correlation. One interpretation is that for Apple, there is a stronger connection between its price and unobserved product characteristics, relative to other manufacturers, which is consistent with the recognition it received for designing the iPhone to be unique and highly differentiated. The weakest correlation is observed for Samsung and Other (generic) feature phones.

Next, we find the variance of the unobservable product characteristic ξ_{jt} to be small but significant. This partially explains why our estimates are significant. This unobservable characteristic evolves differently across the products. We note a strong serial correlation for Samsung and LG, indicating their relative stability over time, whereas in the case of Blackberry, we observe a negative value, consistent with new designs being released during this time period.

6.4 Counterfactual

Next, we look to analyze the impact a number of observable product characteristics have on sales.g Specifically, we examine the sales (market share) impact when xwifi, xgps and xblue are individually set to 0 for all products. In order to determine there corresponding impacts, we use the general method proposed in §5. For completeness we discuss two important details associated with the implementation. First, in the series approximation of the outside market share in eq. (29), we use the quadratic polynomial of v^c_{jt} for each $j = 1, \ldots, 7$ and the inclusive value $\ln(\sum_{j=1}^{7} \exp(v^c_{jt}))$. The inclusive value is used to capture the possible interaction between $v^c_{1t}, \ldots, v^c_{Jt}$. Second, we use eq. (30)-(31) to recover $\xi_{Apple,t}$, because of its high correlation (0.89) between its price and $\xi_{Apple,t}$. In particular, we let

$$\hat{\xi}_{1t} = \left(\frac{1 - \hat{\beta}}{1 - \hat{\beta}\hat{\phi}_1} \right) \left[\hat{y}_{1t} - \hat{\delta}_1 + \hat{\beta}\, \mathrm{E}(\hat{w}_{1,t+1} \mid x_t, p_t, \hat{d}_{(2,1),t}, \ldots, \hat{d}_{(J,1),t}) \right],$$

and $\hat{\xi}_{jt} = \hat{\xi}_{1t} + \hat{d}_{(j,1),t}$. For exposition simplicity, we omit the subscript of state/market. Recall $d_{(j,1),t} = \xi_{jt} - \xi_{1t}$. The conditional expectation was estimated nonparametrically.

Figure 1 shows the counterfactual substitution effects among brands. We compute how the log market shares relative to Apple change from the observed data to the counterfactual (e.g. no Wi-Fi). We find that removing the Bluetooth or Wi-Fi dramatically change the within market shares. Without Wi-Fi, iPhone would lose a substantial amount of its market share when compared with other brands. We note that Wi-Fi is almost exclusively available on iPhone (table 1) during the data period. Thus, it could be viewed as providing a competitive advantage to Apple in that it provides full Internet access. Also, removing GPS does not seem to impact the within market share significantly. This might be due to most consumers not using their phones for GPS, since they were very poor substitutes with limited screen size and visibility. Also, the GPS capabilities provided by phones required consumers to pay an additional monthly fee to their cellular service provider.

Table 4 shows the counterfactual outside market share, which can be understood as the impact on overall demand. The average in table 4 is taken over all months for each state (market). Table 4 shows that removing Wi-Fi or GPS has little effect on the overall demand. However, removing Bluetooth has a large effect on the overall demand. Table 5 reports the total effects by showing the market shares in different counterfactual settings.

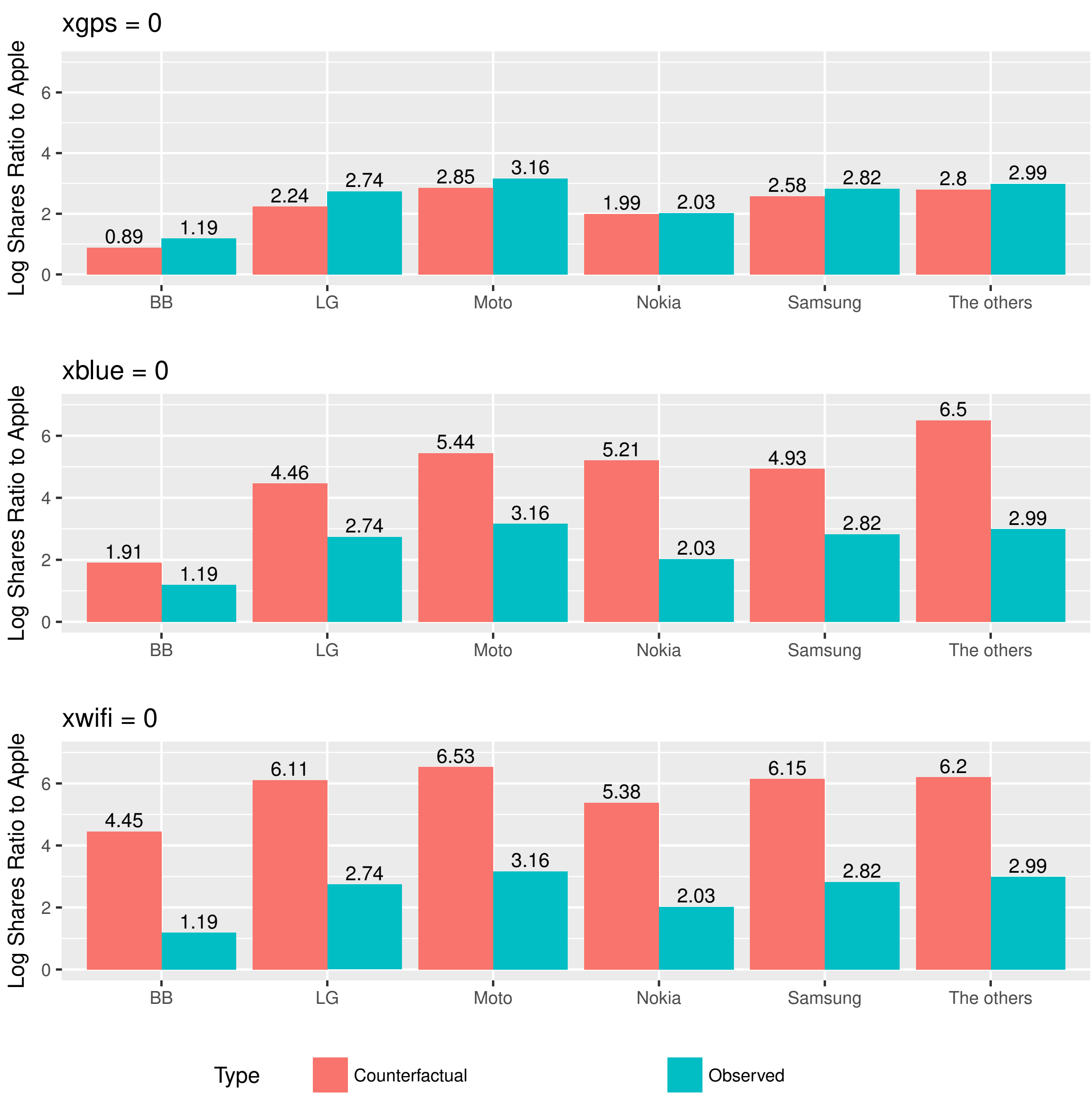

Figure 1: Impact on Within Market Shares

Table 4: Average Counterfactual Outside Market Share (Percentage)

State	No Change	xwifi = 0	xgps = 0	xblue = 0
California	64.4	64.0	68.9	91.5
Florida	71.2	70.9	72.1	91.4
Georgia	66.9	67.7	70.3	91.9
Illinois	67.7	68.6	70.3	90.9
Michigan	68.1	68.7	72.7	91.8
New Jersey	63.1	63.4	70.1	89.8
New York	66.1	66.8	70.5	89.5
Ohio	72.5	72.1	75.1	89.7
Pennsylvania	71.1	70.9	74.1	91.5
Texas	67.0	66.7	68.7	91.4

Table 5: Average Counterfactual Market Shares (Percentage)

Brand	No Change	xwifi = 0	xgps = 0	xblue = 0
Other	6.91	6.19	6.59	4.02
Moto	8.86	9.43	7.66	1.71
Samsung	5.94	6.10	5.44	0.96
LG	5.60	5.93	3.91	0.62
Nokia	2.87	3.02	3.26	1.65
BB	1.40	1.35	1.21	0.08
Apple	0.62	0.02	0.72	0.01

7 Conclusion

We develop a new method to estimate dynamic discrete choice models using only aggregate data. While the extant methods for such estimation are fairly computationally burdensome, our proposed approach has the advantage that it can handle a large number of products, across a number of markets and time periods. The computational complexity is of the order of a linear (or IV) regression to obtain the parameter estimates, making it easily accessible.

We demonstrate the validity through proofs of the asymptotic properties of the estimators, and demonstrate parameter recovery in finite sample simulations. Further, we show the results in a practical application using data from the market for mobile phone handsets.

While the method requires minimal assumptions on the state transition process and other primitives, there are a few limitations worth noting. First, the method allows for product-level differences across both observed and unobserved dimensions, but imposes homogeneity of preferences within a market. However, our method is able to leverage specific properties of a setting where there are 2 or more terminal (or renewal) choices, making the problem similar to a linear model. While the method does not incorporate unobserved consumer heterogeneity in preferences, the approach is suitable for cases where this limitation is offset by the computational simplicity and the fact that no assumptions are needed about the state space or how state variables transition in order to estimate model parameters. We expect that building further on this research to broaden its applicability to be a worthwhile area for further exploration.

A Derivatives for Calculating Asymptotic Variance

We derive the formulas for $\partial g_{1,(j,k),t}(\theta_1)/\partial\theta$, $\partial g_{2,(j,0),t}(\theta_1)/\partial\theta$, $\partial g_{3,j,t}(\theta)/\partial\theta$, $\partial g_{4,(j,k),t}(\theta)/\partial\theta$, and $\partial g_{5,(j,k),t}(\theta)/\partial\theta$. It is easier to calculate the derivatives for $\theta_1 = (\alpha, \beta, \tilde{\gamma}', \delta')'$ and $\theta_2 = (\tilde{\rho}_{'}, \sigma^2, \phi')$. For $\partial g_{1,(j,k),t}(\theta_1)/\partial\theta$, we have

$$
\begin{aligned}
g_{1,(j,k),t,\alpha}(\theta) &= z_{(j,k),t}(p_{jt} - p_{kt}) \\
g_{1,(j,k),t,\beta}(\theta) &= -z_{(j,k),t}(\delta_j - \delta_k)/(1-\beta)^2 \\
g_{1,(j,k),t,\tilde{\gamma}}(\theta) &= -z_{(j,k),t}(x_{jt} - x_{kt})' \\
g_{1,(j,k),t,\delta_i}(\theta) &= \begin{cases} 0 & \text{if } i \neq j, i \neq k \\ -z_{(j,k),t}/(1-\beta) & \text{if } i = j \\ z_{(j,k),t}/(1-\beta) & \text{if } i = k \end{cases} \\
g_{1,(j,k),t,\theta_2}(\theta) &= 0'.
\end{aligned}
$$

For $\partial g_{2,(j,0),t}(\theta_1)/\partial\theta$, we have

$$
\begin{aligned}
g_{2,(j,0),t,\alpha}(\theta) &= x_{j,t,IV,t}(p_{jt} - \beta p_{j,t+1}) \\
g_{2,(j,0),t,\beta}(\theta) &= x_{j,t,IV,t} w_{j,t+1} \\
g_{2,(j,0),t,\tilde{\gamma}}(\theta) &= x_{j,t,IV,t}(-x'_{jt} + \beta x'_{j,t+1}) \\
g_{2,(j,0),t,\delta_i}(\theta) &= \begin{cases} 0 & \text{if } i \neq j \\ -x_{j,t,IV,t} & \text{if } i = j \end{cases} \\
g_{2,(j,0),t,\theta_2}(\theta) &= 0'.
\end{aligned}
$$

For $\partial g_{3,j,t}(\theta)/\partial\theta$, we have

$$
\begin{aligned}
g_{3,j,t,\alpha}(\theta) &= z_{\rho,jt}(1-\beta)(p_{jt} - \beta p_{j,t+1}) \\
g_{3,j,t,\beta}(\theta) &= z_{\rho,jt}[-(y_{jt} + \beta w_{j,t+1}) + (1-\beta) w_{j,t+1} + \tilde{\rho}_j \tilde{p}_{j,t+1}] \\
g_{3,j,t,\tilde{\gamma}}(\theta) &= z_{\rho,jt}(1-\beta)(-x'_{jt} + \beta x'_{j,t+1}) \\
g_{3,j,t,\delta}(\theta) &= 0' \\
g_{3,j,t,\tilde{\rho}_i}(\theta) &= \begin{cases} 0 & \text{if } i \neq j \\ -z_{\rho,jt}(\tilde{p}_{jt} - \beta \tilde{p}_{j,t+1}) & \text{if } i = j \end{cases} \\
g_{3,j,t,\sigma^2}(\theta) &= 0' \\
g_{3,j,t,\phi}(\theta) &= 0'.
\end{aligned}
$$

For $\partial g_{4,(j,k),t}(\theta)/\partial\theta$, we will need $\partial d_{(j,k),t}(\theta)/\partial\theta$:

$$
\begin{aligned}
d_{(j,k),t,\alpha}(\theta) &= (1-\beta)(p_{jt}-p_{kt}) \\
d_{(j,k),t,\beta}(\theta) &= -\left[\ln\left(\frac{s_{jt}}{s_{kt}}\right) - (x_{jt}-x_{kt})'\tilde{\gamma} + \alpha(p_{jt}-p_{kt})\right] \\
d_{(j,k),t,\tilde{\gamma}}(\theta) &= -(1-\beta)(x_{jt}-x_{kt})' \\
d_{(j,k),t,\delta_i}(\theta) &= \begin{cases} 0 & \text{if } i \neq j, i \neq k \\ -1 & \text{if } i = j \\ 1 & \text{if } i = k \end{cases} \\
d_{(j,k),t,\theta_2}(\theta) &= 0'.
\end{aligned}
$$

We have

$$
\begin{aligned}
g_{4,(j,k),t,\theta_1}(\theta) &= d_{(j,k),t} d_{(j,k),t,\theta_1}(\theta) \\
g_{4,(j,k),t,\tilde{\rho}_i}(\theta) &= \begin{cases} 0 & \text{if } i \neq j, i \neq k \\ \tilde{\rho}_k \tilde{p}_{jt} \tilde{p}_{kt} & \text{if } i = j \\ \tilde{\rho}_j \tilde{p}_{jt} \tilde{p}_{kt} & \text{if } i = k \end{cases} \\
g_{4,(j,k),t,\sigma^2}(\theta) &= -1 \\
g_{4,(j,k),t,\phi}(\theta) &= 0'.
\end{aligned}
$$

For $\partial g_{5,(j,k),t}(\theta)/\partial\theta$, we have

$$g_{5,(j,k),t,\alpha}(\theta) = \frac{d_{(j,k),t}}{\beta\sigma^2} d_{(j,k),t,\alpha}(\theta) - \left(\frac{1-\beta}{\beta}\right)(p_{jt} - \beta p_{j,t+1})\frac{d_{(j,k),t}}{\sigma^2} - \left(\frac{1-\beta}{\beta}\right)(y_{jt} + \beta w_{j,t+1})\frac{d_{(j,k),t,\alpha}(\theta)}{\sigma^2}$$

$$g_{5,(j,k),t,\beta}(\theta) = \left(\frac{d_{(j,k),t} d_{(j,k),t,\beta}(\theta)}{\beta\sigma^2} - \frac{d^2_{(j,k),t}}{2\beta^2\sigma^2}\right) + \frac{1}{\beta^2}(y_{jt} + \beta w_{j,t+1})\frac{d_{(j,k),t}}{\sigma^2} - \left(\frac{1-\beta}{\beta}\right)\left[w_{j,t+1}\frac{d_{(j,k),t}}{\sigma^2} + (y_{jt} + \beta w_{j,t+1})\frac{d_{(j,k),t,\beta}(\theta)}{\sigma^2}\right]$$

$$g_{5,(j,k),t,\tilde{\gamma}}(\theta) = \frac{d_{(j,k),t}}{\beta\sigma^2} d_{(j,k),t,\tilde{\gamma}}(\theta) - \left(\frac{1-\beta}{\beta}\right)\left[(-x'_{jt} + \beta x'_{j,t+1})\frac{d_{(j,k),t}}{\sigma^2} + (y_{jt} + \beta w_{j,t+1})\frac{d_{(j,k),t,\tilde{\gamma}}(\theta)}{\sigma^2}\right]$$

$$g_{5,(j,k),t,\delta}(\theta) = \left[\frac{d_{(j,k),t}}{\beta\sigma^2} - \left(\frac{1-\beta}{\beta}\right)(y_{jt} + \beta w_{j,t+1})\frac{1}{\sigma^2}\right] d_{(j,k),t,\delta}(\theta)$$

$$g_{5,(j,k),t,\tilde{\rho}}(\theta) = 0'$$

$$g_{5,(j,k),t,\sigma^2}(\theta) = -\left[\frac{d^2_{(j,k),t}}{2\beta} - \left(\frac{1-\beta}{\beta}\right)(y_{jt} + \beta w_{j,t+1}) d_{(j,k),t}\right]\frac{1}{\sigma^4}$$

$$g_{5,(j,k),t,\phi_i}(\theta) = \begin{cases} 0 & \text{if } i \neq j \\ -1 & \text{if } i = j \end{cases}.$$

B Nested Logit Extension

Multinomial logit specification has the notorious "independent irrelevant alternative" properties. We will consider nested logit here as a remedy. First split the products $\{0, 1 \ldots, J\}$ into $\kappa + 1$ exhaustive and mutually exclusive sets. Denote $\mathcal{G}_A$ the A-th group. The outside good 0 is assumed to be the only member of group 0. When one product, excepting for 0, forms a group by itself, we call it a "*stand-alone*" product. For a product j, let $\bar{s}_{jt}$ be the market share of the group containing j, let $\tilde{s}_{jt} = s_{jt}/\bar{s}_{jt}$ be the within group market share. Of course, if product j is a stand-alone product, $\bar{s}_{jt} = s_{jt}$ and $\tilde{s}_{jt} = 1$.

Assumption 12. *Assume that ε_{it} follows the following generalized EVD*

$$F(\varepsilon_{it}) = \exp\left[-\sum_{A=0}^{\kappa}\Big(\sum_{j\in\mathcal{G}_A} e^{-\varepsilon_{ijt}/\zeta(A)}\Big)^{\zeta(A)}\right]$$

The unknown scale parameter $\zeta(A)$ determines within nest correlation of group $\mathcal{G}_A$. For any group A with one single product, such as $\mathcal{G}_0 = \{0\}$, let $\zeta(A) = 1$.

For any product j, we also use ζ_j to denote the within nest correlation of the group containing j. For example, if $j \in \mathcal{G}_A$, $\zeta_j = \zeta(A)$. It is well know that the within nest correlation coefficient is $1 - \zeta(A)^2$.

B.1 Identification

For any two products j and k on the market in period t, we have the ratio of their market share as follows,

$$\frac{s_{jt}}{s_{kt}} = \frac{\exp(v_{jt}/\zeta_j)}{\exp(v_{kt}/\zeta_k)} \frac{\mu_{jt}^{(\zeta_j-1)/\zeta_j}}{\mu_{kt}^{(\zeta_k-1)/\zeta_k}} \qquad \text{and} \qquad \frac{\bar{s}_{jt}}{\bar{s}_{kt}} = \frac{\mu_{jt}}{\mu_{kt}},$$

by the nested logit model. If $\mathcal{G}_A$ is the group containing j,

$$\mu_{jt} = \Big(\sum_{\ell \in \mathcal{G}_A} \exp(v_{\ell t}/\zeta_j)\Big)^{\zeta_j}.$$

Proposition 6. *If (i) product j and k belongs to the same group, or (ii) product k forms a group by itself, we have*

$$\ln\left(\frac{\tilde{s}_{jt}}{\tilde{s}_{kt}}\right) = \frac{v_{jt} - v_{kt}}{\zeta_j} - \zeta_j^{-1} \ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{kt}}\right).$$

Proof. For any two products j and k (including the outside option 0), we have

$$\ln\left(\frac{s_{jt}}{s_{kt}}\right) = \frac{v_{jt}}{\zeta_j} - \frac{v_{kt}}{\zeta_k} + \left(\frac{\zeta_j - 1}{\zeta_j}\right) \ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{kt}}\right) + \left[\left(\frac{\zeta_j - 1}{\zeta_j}\right) - \left(\frac{\zeta_k - 1}{\zeta_k}\right)\right] \ln \mu_{kt}.$$

The above complicated market shares ratio can be simplified in two specical cases. First, if product j and k are from the same group, then $\mu_{jt} = \mu_{kt}$ and $\zeta_j = \zeta_k$. We have

$$\ln\left(\frac{s_{jt}}{s_{kt}}\right) = \frac{v_{jt} - v_{kt}}{\zeta_j}.$$

When j and k are from the same group, $\bar{s}_{jt} = \bar{s}_{kt}$, $\tilde{s}_{jt}/\tilde{s}_{kt} = s_{jt}/s_{kt}$, and the proposition holds. Second, if product k itself forms a group, $\zeta_k = 1$, $\ln \mu_{kt} = v_{kt}$, and $s_{kt} = \bar{s}_{kt}$. For any product j, we have that

$$\ln\left(\frac{s_{jt}}{s_{kt}}\right) = \frac{v_{jt} - v_{kt}}{\zeta_j} + \left(\frac{\zeta_j - 1}{\zeta_j}\right) \ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{kt}}\right).$$

Subtracting both sides with $\ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{kt}}\right)$, one will get the conclusion. ∎

For each group $A = 1, \ldots, \kappa$, we can identify the within nest correlation $\zeta(A)$ and the preference parameters (α, β, γ) by the following arguments. Without loss of generality, suppose product $2 \in \mathcal{G}_A$, and product 1 either comes from the same group as product 2 or product 1 is a stand-alone product. From proposition 6, we have

$$\ln\left(\frac{\tilde{s}_{2t}}{\tilde{s}_{1t}}\right) = (x_{2t} - x_{1t})'\tilde{\gamma}/\zeta_2 - (p_{2t} - p_{1t})\alpha/\zeta_2 + \frac{\delta_2 - \delta_1}{(1-\beta)\zeta_2} - \zeta_2^{-1}\ln\left(\frac{\bar{s}_{2t}}{\bar{s}_{1t}}\right) + \frac{\xi_{2t} - \xi_{1t}}{(1-\beta)\zeta_2}. \quad \text{(B.1)}$$

Recall that $\tilde{\gamma} = \gamma/(1-\beta)$, and note that $\zeta_2 = \zeta(A)$ since $2 \in \mathcal{G}_A$. Letting $z_{(2,1)t}$ be a vector of IVs that are uncorrelated with $\xi_{2t} - \xi_{1t}$ (assuming the constant term 1 is always included in the IV), we have

$$\mathrm{E}\big(g_{1,(2,1),t}(\theta_o)\big) = 0,$$

where

$$g_{1,(2,1),t}(\theta) = z_{(2,1)t}\left[\ln\left(\frac{\tilde{s}_{2t}}{\tilde{s}_{1t}}\right) - (x_{2t} - x_{1t})'\tilde{\gamma}/\zeta_2 + (p_{2t} - p_{1t})\alpha/\zeta_2 - \frac{\delta_2 - \delta_1}{(1-\beta)\zeta_2} + \zeta_2^{-1}\ln\left(\frac{\bar{s}_{2t}}{\bar{s}_{1t}}\right)\right].$$

We can identify $\tilde{\gamma}/\zeta_2$, α/ζ_2, $(\delta_2 - \delta_1)/[(1-\beta)\zeta_2]$, and the difference $(\xi_{2t} - \xi_{1t})/[(1-\beta)\zeta_2]$. When product 1 is a stand-alone product, $\ln(\bar{s}_{2t}/\bar{s}_{1t}) \neq 0$, hence we can also identify ζ_2^{-1}, hence $\tilde{\gamma}$, α, $(\delta_2 - \delta_1)/(1-\beta)$, and $(\xi_{2t} - \xi_{1t})/(1-\beta)$. If 1 and 2 are from the same group, $\ln(\bar{s}_{2t}/\bar{s}_{1t}) = 0$ and ζ_2 is not identified from the equation.

We next consider the identification of β, δ_2 and ζ_2 (if there was no stand-alone product). The conclusion is

$$\mathrm{E}(g_{2,(2,0),t}(\theta_o)) = 0,$$

where

$$g_{2,(2,0),t}(\theta) = x_{2t,IV}\left[\ln\left(\frac{\bar{s}_{2t}}{\bar{s}_{0t}}\right) + \zeta_2 y_t - \beta\zeta_2 y_{t+1} - \beta\ln\bar{s}_{2,t+1} - \delta_2\right],$$

and

$$y_t = \ln\tilde{s}_{2t} - x_{2t}'\tilde{\gamma}/\zeta_2 + p_{2t}\alpha/\zeta_2, \quad \text{(B.2)}$$

and the IV $x_{2t,IV}$ is a vector of functions of m_t such that $\mathrm{cov}(x_{2t,IV}, \xi_{2t}) = \mathrm{cov}(x_{2t,IV}, \xi_{2,t+1}) = 0$. If ζ_2 has been identified in the first step, the moment condition is linear in (β, δ_2). Otherwise, there are three parameters, β, ζ_2 and δ_2, and this moment condition is nonlinear in them for the presence of $\beta\zeta_2$.

We now derive the above conclusion. Using proposition 6 for the case $j = 2$ and $k = 0$, we have

$$\ln\tilde{s}_{2t} = \frac{v_{2t} - v_{0t}}{\zeta_2} - \zeta_2^{-1}\ln\left(\frac{\bar{s}_{2t}}{\bar{s}_{0t}}\right),$$

because $\tilde{s}_{0t} = 1$. Using the notation of y_t in eq. (B.2), we have

$$y_t = \frac{\delta_2}{(1-\beta)\zeta_2} + \frac{\xi_{2t}}{(1-\beta)\zeta_2} - \zeta_2^{-1} \ln\left(\frac{\bar{s}_{2t}}{\bar{s}_{0t}}\right) - \frac{v_{0t}}{\zeta_2}. \tag{B.3}$$

The next objective is to derive an alternative formula of $v_{0t} = \beta\,\mathrm{E}(\bar{V}_{t+1}(m_{t+1}) \mid m_t)$. It follows the expectation maximization formula for nested logit model (see e.g. Arcidiacono and Miller, 2011, lemma 3) that

$$\begin{aligned}\bar{V}_t &= v_{2t} - [\zeta_2 \ln s_{2t} + (1-\zeta_2)\ln \bar{s}_{2t}] \\ &= -\zeta_2 y_t + \frac{\delta_2}{1-\beta} + \frac{\xi_{2t}}{1-\beta} - \ln \bar{s}_{2t}.\end{aligned}$$

Hence $v_{0t} = \beta\,\mathrm{E}(\bar{V}_{t+1} \mid m_t)$ becomes

$$v_{0t} = \beta\,\mathrm{E}\left(-\zeta_2 y_{t+1} + \frac{\delta_2}{1-\beta} + \frac{\xi_{2,t+1}}{1-\beta} - \ln \bar{s}_{2,t+1} \,\middle|\, m_t\right). \tag{B.4}$$

Substituting v_{0t} in eq. (B.3) with eq. (B.4), we have

$$y_t = \frac{\delta_2}{(1-\beta)\zeta_2} + \frac{\xi_{2t}}{(1-\beta)\zeta_2} - \zeta_2^{-1}\ln\left(\frac{\bar{s}_{2t}}{\bar{s}_{0t}}\right) + \beta\,\mathrm{E}\left(y_{t+1} - \frac{\delta_2}{(1-\beta)\zeta_2} - \frac{\xi_{2,t+1}}{(1-\beta)\zeta_2} + \frac{\ln \bar{s}_{2,t+1}}{\zeta_2} \,\middle|\, m_t\right).$$

This implies the following conditional moment condition,

$$\mathrm{E}\left(y_t + \frac{\ln(\bar{s}_{2t}/\bar{s}_{0t})}{\zeta_2} - \beta y_{t+1} - \beta\frac{\ln \bar{s}_{2,t+1}}{\zeta_2} - \frac{\delta_2}{\zeta_2} - \frac{1}{1-\beta}\frac{\xi_{2t}}{\zeta_2} + \frac{\beta}{1-\beta}\frac{\xi_{2,t+1}}{\zeta_2} \,\middle|\, m_t\right) = 0.$$

Multiplying both sides of the above display by ζ_2, we have the following

$$\mathrm{E}\left(\zeta_2 y_t + \ln\left(\frac{\bar{s}_{2t}}{\bar{s}_{0t}}\right) - \beta\zeta_2 y_{t+1} - \beta\ln \bar{s}_{2,t+1} - \delta_2 - \frac{1}{1-\beta}\xi_{2t} + \frac{\beta}{1-\beta}\xi_{2,t+1} \,\middle|\, m_t\right) = 0.$$

Then we have the stated conclusion by the arguments in the paper.

Letting

$$\tilde{y}_t = \zeta_2 y_t + \ln(\bar{s}_{2t}/\bar{s}_{0t}), \qquad \text{and} \qquad \tilde{w}_{t+1} = -\zeta_2 y_{t+1} - \ln \bar{s}_{2,t+1},$$

we have

$$\mathrm{E}\left(\tilde{y}_t + \beta\tilde{w}_{t+1} - \delta_2 - \frac{1}{1-\beta}\xi_{2t} + \frac{\beta}{1-\beta}\xi_{2,t+1} \,\middle|\, m_t\right) = 0, \tag{B.5}$$

Because we have identified ζ_2, $\tilde{y}_t$ and $\tilde{w}_t$ are identified terms. Equation (B.5) now is identical to the conditional moment equation in the multinomial case excepting for the identified object $\tilde{y}_t$ and $\tilde{w}_t$. Hence we can use the same arguments in the multinomial case to show the identification of the dynamics of state evolution in nested logit case. Also, we denote $\tilde{d}_t = \xi_{2t} - \xi_{1t}$ for the nested logit case, and

$$\tilde{d}_t = (1-\beta)\zeta_2 \ln\left(\frac{\tilde{s}_{2t}}{\tilde{s}_{1t}}\right) - (x_{2t} - x_{1t})'\gamma + (1-\beta)\alpha(p_{2t} - p_{1t}) - (\delta_2 - \delta_1) + (1-\beta)\ln\left(\frac{\bar{s}_{2t}}{\bar{s}_{1t}}\right).$$

B.2 Estimation

We focus on the case that the data are from one single market over T consecutive periods.

B.2.1 Preference

No stand-alone product Suppose excepting for the outside good, every group contains at least two products.

Step 1: For each group $A = 1, \ldots, G$, estimate $(\tilde{\gamma}'/\zeta(A), \alpha/\zeta(A))$ using the following moment equation:

$$\mathrm{E}\big(g_{1,(j,k),t}(\theta_o)\big) = 0, \qquad \text{for } j, k \in \mathcal{G}_A \text{ and } j > k,$$

$$g_{1,(j,k),t}(\theta) = z_{(j,k)t}\left[\ln\left(\frac{\tilde{s}_{jt}}{\tilde{s}_{kt}}\right) - (x_{jt} - x_{kt})'\tilde{\gamma}/\zeta_A + (p_{jt} - p_{kt})\alpha/\zeta_A - \frac{\delta_j - \delta_k}{(1-\beta)\zeta_A}\right].$$

The vector $z_{(j,k)t}$ is a vector of IVs that are uncorrelated with $(\xi_{jt} - \xi_{kt})$.

In practice, one can estimate $(\tilde{\gamma}'/\zeta(A), \alpha/\zeta(A))$ by an IV regression of $\ln(\tilde{s}_{jt}/\tilde{s}_{kt})$ on $x_{jt} - x_{kt}$ and $p_{jt} - p_{kt}$ with IV $z_{(j,k)t}$ using data $t = 1, \ldots, T$. Letting $\widehat{\tilde{\gamma}/\zeta(A)}$ and $\widehat{\alpha/\zeta(A)}$ be the obtained estimates, define

$$y_{jt} = \ln \tilde{s}_{jt} - x_{jt}'\tilde{\gamma}/\zeta_j + p_{jt}\alpha/\zeta_j,$$

and their estimates

$$\hat{y}_{jt} = \ln \tilde{s}_{jt} - x_{jt}'\widehat{\tilde{\gamma}/\zeta_j} + p_{jt}\widehat{\alpha/\zeta_j}.$$

Note that $\zeta_j = \zeta(A)$ when $j \in \mathcal{G}_A$.

Step 2: Estimate β, ζ, and δ. Define a list of group dummy variables $d^G_{A,jt} = 1$ if $j \in \mathcal{G}_A$ and $= 0$ otherwise. Estimate β, ζ, and δ using

$$\mathrm{E}(g_{2,(j,0),t}(\theta_o)) = 0,$$

where

$$g_{2,(j,0),t}(\theta) = x_{jt,IV}\left[\ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{0t}}\right) + \sum_{A=1}^{\kappa}\zeta(A)d^G_{A,jt}y_{jt} - \sum_{A=1}^{\kappa}\beta\zeta(A)d^G_{A,j,t+1}y_{j,t+1} - \beta\ln\bar{s}_{j,t+1} - \delta_j\right].$$

In practice, one can first estimate β and $\zeta(1), \ldots, \zeta(\kappa)$ by solving the nonlinear least square problem,

$$\min_{\beta,\zeta}\sum_{j=1}^{J}\sum_{t=1}^{T-1}\hat{g}_{2,(j,0),t}(\theta)'\hat{g}_{2,(j,0),t}(\theta),$$

where

$$\hat{g}_{2,(j,0),t}(\theta) = (x_{jt,IV} - \bar{x}_{j,IV}) \left[\ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{0t}}\right) + \sum_{A=1}^{\kappa} \zeta(A) d^G_{A,jt} \hat{y}_{jt} - \sum_{A=1}^{\kappa} \beta\zeta(A) d^G_{A,j,t+1} \hat{y}_{j,t+1} - \beta \ln \bar{s}_{j,t+1} \right].$$

Here $\bar{x}_{j,IV} = T^{-1} \sum_{t=1}^{T} x_{jt,IV}$ is the sample average of $x_{jt,IV}$. As for initial values, one can run an IV regression of $\ln(\bar{s}_{jt}/\bar{s}_{0t})$ on $d^G_{A,jt}\hat{y}_{jt}$, $d^G_{A,j,t+1}\hat{y}_{j,t+1}$ and $\ln \bar{s}_{j,t+1}$ with IV $x_{jt,IV}$, and use the coefficients associated with $d^G_{A,jt}\hat{y}_{jt}$ and $\ln \bar{s}_{j,t+1}$ as the initial values for $\zeta(A)$ and β.

After obtaining $\hat{\zeta}$ and $\hat{\beta}$, one can let

$$\hat{\delta}_j = T^{-1} \sum_{t=1}^{T-1} \ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{0t}}\right) + \hat{\zeta}_j \hat{y}_{jt} - \hat{\beta}\hat{\zeta}_j y_{j,t+1} - \hat{\beta} \ln \bar{s}_{j,t+1}$$

be the estimate of δ_j.

With stand-alone product When there are stand-alone products, the above estimation can be simplified. Without loss of generality, assume product $1, \ldots, J_1$ are stand-alone products, and they form group $1, \ldots, J_1$, respectively. If $J_1 = J$, this becomes multinomial logit case.

Step 1: Multiplying both sides of eq. (B.1) by the within nest correlation coefficient, we in general have

$$\ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{kt}}\right) = \frac{\delta_j - \delta_k}{1-\beta} + (x_{jt} - x_{kt})'\tilde{\gamma} - (p_{jt} - p_{kt})\alpha - \zeta_j \ln\left(\frac{\tilde{s}_{jt}}{\tilde{s}_{kt}}\right) + \frac{\xi_{jt} - \xi_{kt}}{1-\beta}.$$

Note that $\ln(\bar{s}_{jt}/\bar{s}_{kt}) = 0$ if j, k are from the same nest, and $\ln(\tilde{s}_{jt}/\tilde{s}_{kt}) = 0$ if j and k are both stand-alone product. When there is at least one stand-alone product, i.e. $J_1 \geq 1$, we can estimate $\zeta(J_1+1), \ldots, \zeta(\kappa)$ ($\zeta(0) = \cdots = \zeta(J_1) = 1$), $\tilde{\gamma}$ and α by the following,

$$g_{1,(j,k),t}(\theta) = (z_{(j,k)t} - \bar{z}_{(j,k)}) \left[\ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{kt}}\right) - (x_{jt} - x_{kt})'\tilde{\gamma} + (p_{jt} - p_{kt})\alpha + \sum_{A=J_1+1}^{\kappa} \zeta(A) d^G_{A,jt} \ln\left(\frac{\tilde{s}_{jt}}{\tilde{s}_{kt}}\right) \right],$$

where $\bar{z}_{(j,k)} = T^{-1} \sum_{t=1}^{T} z_{(j,k)t}$. In practice, we run an IV regression of $\ln(\bar{s}_{jt}/\bar{s}_{kt})$ on $x_{jt} - x_{kt}$, $p_{jt} - p_{kt}$, and $d^G_{A,jt} \ln(\tilde{s}_{jt}/\tilde{s}_{kt})$ with IV $z_{(j,k)t}$.

Step 2: Estimate β. Because ζ_j has been estimated in the first step, $\tilde{y}_t$ and $\tilde{w}_t$ are now known. In general, define

$$\tilde{y}_{jt} = \zeta_j y_{jt} + \ln(\bar{s}_{jt}/\bar{s}_{0t}), \qquad \text{and} \qquad \tilde{w}_{j,t+1} = -\zeta_j y_{j,t+1} - \ln \bar{s}_{j,t+1},$$

We then can estimate β using

$$\mathrm{E}(g_{2,(j,0),t}(\theta_o)) = 0,$$

where

$$g_{2,(j,0),t}(\theta) = x_{jt,IV}(\tilde{y}_{jt} + \beta\tilde{w}_{t+1} - \delta_j).$$

In practice, to estimate β, one simply runs an IV regression of $\hat{\tilde{y}}_{jt}$ on $-\hat{\tilde{w}}_{j,t+1}$ using $x_{jt,IV}$ as the IV.

Step 3: Estimate δ_j using

$$\mathrm{E}(\tilde{y}_{jt} + \beta\tilde{w}_{t+1} - \delta_j) = 0.$$

In practice, one runs a linear regression for each j of $(\hat{\tilde{y}}_{jt} + \hat{\beta}\hat{w}_{j,t+1})$ on a constant of one using data from $t = 1, \ldots, T-1$.

B.2.2 $F(m_t)$ and $F(m_{t+1} \mid m_t)$

We make the same normal distribution asssumption as in the paper. The estimation of the parameters in $F(m_t)$ and $F(m_{t+1} \mid m_t)$ in nested logit case is identical to the multinomial logit case by replacing $d_{(j,k),t}$, y_{jt} and w_{jt} in multinomial logit case with $\tilde{d}_{(j,k),t}$, $\tilde{y}_{jt}$, and $\tilde{w}_{jt}$. Here

$$\tilde{d}_{(j,k)t} = (1-\beta)\zeta_j \ln\left(\frac{\tilde{s}_{jt}}{\tilde{s}_{kt}}\right) - (x_{jt} - x_{kt})'\gamma + (1-\beta)\alpha(p_{jt} - p_{kt}) - (\delta_j - \delta_k) + (1-\beta)\ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{kt}}\right).$$

So we will not repeat the procedures here.

B.3 Mobile Phone Market Application with Nested Logit Specification

Using the nested logit (NL) specification, we re-estimated the cell phone market application. Besides the outside option, there are three nests in the model. Nest 1 consists of Apple and RIM (Blackberry), nest 2 consists of the well regarded brands of feature phones at the time (Motorola, Samsung, LG and Nokia), and nest 3 consists of all other brands. In this specification, “all other brands” is a stand-alone product, hence we used the estimation method outlined in appendix B.2.1. In the estimation, we used the same instrumental variables as we used in the multinomial logit (MNL) specification. The results are detailed in table B.3.

The correlation coefficient for nest 2 (well regarded feature phones) is almost 1.00. This likely because these features phones are very similar. The correlation coefficient for nest 1 (Blackberry and iPhone) is 0.78 due to some important difference between these two phones, e.g. iPhone can access Wi-Fi, though they are both smartphones.

The estimates of many important parameters in the NL case are close to the estimates in the MNL case. The price coefficient, α, in both NL and MNL is -0.01. The estimate of the discount factor, β, in NL is 0.97, which is bigger than 0.8 in the MNL case. The ordering of the estimated fixed effect among different phones from both MNL and NL is similar—iPhone has the highest fixed effect, while Blackberry has the lowest. Also, similar to the estimates in the MNL, iPhone has the highest correlation between price and unobserved product characteristics. The estimates of the serial correlation of ξ_{jt} are somewhat different from the MNL case. The most noticeable difference is the iPhone, whose autocorrelation coefficient is greater than 1. This means $\xi_{iPhone,t}$ is a nonstationary process, which could be due to that iPhone had only entered into the market for a few months.

It is noticeable that the estimated standard error of ξ_{jt} is substantially smaller than the MNL case. This can be understood from the regression formula in the NL case,

$$\ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{kt}}\right) = \frac{\delta_j - \delta_k}{1-\beta} + (x_{jt} - x_{kt})'\tilde{\gamma} - (p_{jt} - p_{kt})\alpha - \zeta_j \ln\left(\frac{\tilde{s}_{jt}}{\tilde{s}_{kt}}\right) + \frac{\xi_{jt} - \xi_{kt}}{1-\beta}.$$

Note that in the MNL case, each product forms a nest by itself, and above equation becomes

$$\ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{kt}}\right) = \frac{\delta_j - \delta_k}{1-\beta} + (x_{jt} - x_{kt})'\tilde{\gamma} - (p_{jt} - p_{kt})\alpha + \frac{\xi_{jt} - \xi_{kt}}{1-\beta}.$$

The "regressor" $\ln(\tilde{s}_{jt}/\tilde{s}_{kt})$ vanishes in the MNL case. The estimated variance of ξ_{jt} essentially depends on the variance of the "error term" $(\xi_{jt} - \xi_{kt})/(1-\beta)$ in the above regression equations. In the NL case, we have one additional regressor $\ln(\tilde{s}_{jt}/\tilde{s}_{kt})$, hence the variance of the residuals will be reduced. The observed reduction of the variance of unobserved product characteristics after controlling for nest or group market share suggests that in empirical research, one might be able to at least reduce the influence of the unobserved product characteristics by using certain observed group characteristics, e.g. the nest or group market share herein.

C ONLINE APPENDIX–Numerical 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 markets, the number of time periods and whether the data generating process originated from a type 1 extreme value distribution or a GEV distribution.

Table 6: Estimation Results of Mobile Phone Market: Nested Logit

	Parameters	Estimate	Std. Error	t value
Step 1: preference, $\gamma/(1-\beta)$, within nest corr	price	-0.01	0.00	-4.60
	xblue	1.00	0.58	1.72
	xgps	0.47	0.09	5.22
	xweight	-0.08	0.05	-1.72
	xqwerty	-1.55	0.45	-3.41
	xmusic	-0.27	1.04	-0.26
	xwifi	0.68	0.91	0.75
	xtalktime	0.12	0.05	2.39
	Corr in nest 1	0.78	0.15	5.35
	Corr in nest 2	1.00	0.00	371.13
Step 2: discount factor	β	0.97	0.10	9.55
Step 3: fixed effect	δ_{Moto}	0.15	0.07	2.07
	$\delta_{Samsung}$	0.16	0.07	2.21
	δ_{LG}	0.12	0.07	1.64
	δ_{Nokia}	0.19	0.07	2.56
	$\delta_{Blackberry}$	0.11	0.08	1.40
	δ_{Apple}	0.28	0.08	3.52
	δ_{Other}	0.16	0.07	2.14
Step 4: correlation between price and unobserved product characteristics	ρ_{Moto}	0.14	0.02	5.54
	$\rho_{Samsung}$	0.17	0.03	6.28
	ρ_{LG}	0.21	0.03	7.14
	ρ_{Nokia}	0.20	0.03	7.89
	$\rho_{Blackberry}$	0.24	0.07	3.28
	ρ_{Apple}	0.69	0.11	6.13
	ρ_{Other}	0.25	0.05	5.00
Step 5: std. error of ξ_{jt}	σ	0.05	0.00	41.22
Step 6: autocorrelation of ξ_{jt}	ϕ_{Moto}	0.08	0.02	4.34
	$\phi_{Samsung}$	0.04	0.01	4.05
	ϕ_{LG}	0.45	0.03	17.73
	ϕ_{Nokia}	0.40	0.03	14.33
	$\phi_{Blackberry}$	0.42	0.07	5.66
	ϕ_{Apple}	4.67	0.23	20.28

[1] The standard error reported here are obtained from sequential estimation steps.

C.1 Logit Model

We first discuss the data generating process associated with the logit model. The consumer's flow utility function follows the specification in §2.1. When consumer i purchases product j in period t, he receives the following flow utility in period t,

$$u_{ijt} = f(x_{jt}, \xi_{jt}) - \alpha p_{jt} + \varepsilon_{ijt} \equiv f(x_{jt}, \xi_{jt}) - 0.5p_{jt} + \varepsilon_{ijt},$$

and receives $f(x_{jt}, \xi_{jt})$ as flow utility in each period post purchase in period t. In the simulation we let

$$f(x_{jt}, \xi_{jt}) = x'_{jt}\gamma + \delta_j + \xi_{jt} = x'_{jt}0 + 0.75 + \xi_{jt},$$

for any product j. So $\alpha = 0.5$, $\gamma = 0$ and $\delta_j = 0.75$ for any product j. Products in the simulation are differentiated by the observed price, p_{jt}, and unobserved characteristics, ξ_{jt}. The discount factor β is set to 0.80. We maintain the independence and logit specification about ε_{ijt}, i.e. Assumption 3 and 4.

We next describe the data generation process of price and the unobserved product characteristics. We specifically account for correlation between ξ_{jt} and p_{jt}. Such a formulation is motivated by the price endogeneity problem researchers face when employing aggregate data, where firms can observe ξ_{jt} and then set prices optimally. We use a reduced form price model to characterize this dependence. Specifically,

$$p_{jt} = c_j + MC_{jt} + \omega_{jt} \qquad \text{and} \qquad \xi_{jt} = \phi_j \xi_{j,t-1} + \nu_{jt},$$

where $(\omega_{jt}, \nu_{jt})'$ is *iid* across products and time periods, and follows normal distribution,

$$\begin{pmatrix} \omega_{jt} \\ \nu_{jt} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma_p^2 & \rho\sigma_\nu\sigma_p \\ & \sigma_\nu^2 \end{pmatrix}\right).$$

Here MC_{jt} denotes the marginal cost of product j at time t. MC_{jt} is independent of $(\omega_{j\tilde{t}}, \nu_{j\tilde{t}})'$ for any period t and $\tilde{t}$. Specifically, MC_{jt} takes the form

$$MC_{jt} = \psi_j MC_{j,t-1}$$

We will use MC_{jt} as the instrumental variable in both estimation steps 1 and 2 outlined in §4.1.

In our simulations the maximum number of products is 5, and we assign the following parameter values. We let $(c_1, \ldots, c_5) = (1, 2.5, 3.5, 4.5, 5.5)$ and $(\psi_1, \ldots, \psi_5) = (0.98, 0.92, 0.88, 0.84, 0.80)$. For the initial state of MC_{j0}, we let $(MC_{10}, \ldots, MC_{50}) = (15, 14.5, 14, 13.5, 13)$.

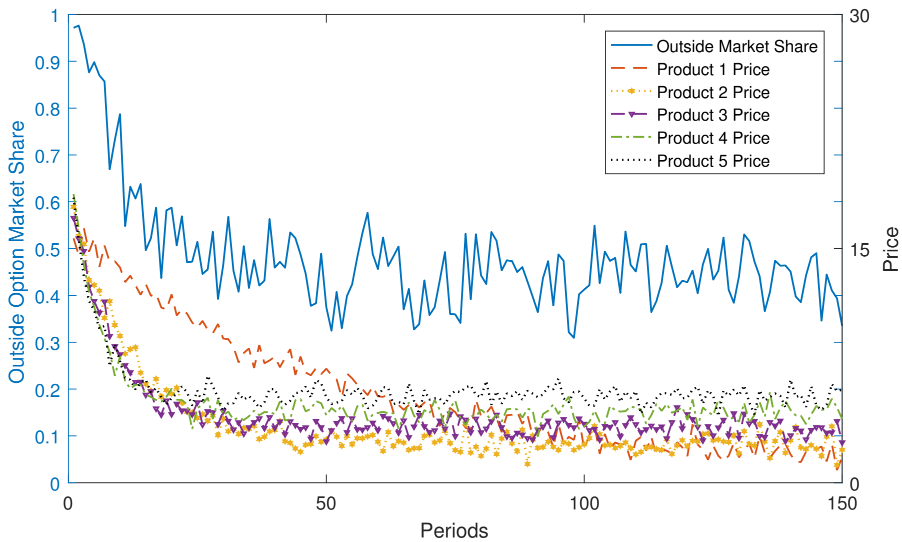

Figure 2: Monte Carlo Prices and Outside Market Share ($J = 5$)

Such specification ensures that product marginal cost, MC_{jt}, has a declining trajectory, which is consistent with durable goods models.

In addition, we let $\phi_j = 0.25$ for any product j.[8] Let $\sigma_p = 0.5$, $\rho = 0.25$, and $\sigma_\nu = 0.1$. Since ξ_{jt} is a stationary AR(1) process, it is easy to see that $\sigma^2 = \mathrm{Var}(\xi_{jt}) = 0.1^2/(1-0.25^2)$, that is $\sigma \approx 0.1033$. Moreover, $\mathrm{corr}(\xi_{jt}, p_{jt}) = \rho$ by serial independence of both ω_{jt} and ν_{jt}.

In fig. 2, we present prices and the outside option's market share in order to illustrate that the DGP is consistent with a durable goods setting. Note the declining prices and decreasing market share of the outside option in Figure 2.

Suppose for J products and one market we have simulated panel data (s_t, p_t, MC_t, ξ_t) for T periods. We first estimate α with an instrumental variable regression. We use the marginal cost variable above as a price instrument. Given the estimates of α we have estimates of y_t and w_t. We then estimate β using two stage least squares as discussed in §5.1.2, using the demeaned price instrument as the instrument. Once β is estimated, we can estimate γ by multiplying the estimate of $\tilde{\gamma}$ with $1-\hat{\beta}$, if we included other observed product characteristics to estimate. Yet, since the DGP only consists of a constant term, we estimate the constant using step 3 in section 5.1.3. The estimation of $\mathrm{Var}(\xi_{jt})$ follows the steps in the previous section. We also estimate $\mathrm{E}(\xi_{jt} \mid p_{jt})$ using step 4 in §5.2.1 to recover ρ and σ.

Each set of simulations we analyze was based on 250 replications. We also analyze sets

[8]We also performed simulations when ξ_{jt} has no serial correlation, i.e. $\phi_j = 0$. Results are available upon request.

with varying number of markets (1 and 10), time periods (150, or 300) and the number of J.

The first set of simulations in Table 7 and 8 illustrate how well and precise our methodology is able to identify the data generating process—including the discount factor. Furthermore, if the discount factor is known (or assumed), the results exhibit less small sample bias and more precision, particularly for the parameters that include the discount factor in estimation, γ, σ, ρ, and ϕ. Specifically, we determine that estimation of ρ is quite challenging in practice and requires a sizeable amount of data and products to precisely estimate when the discount factor is estimated. This again is from the fact that

$$g_{3jt}(\theta) = z_{\rho,j,t}(p_{jt})r_{jt}$$

$$r_{jt} = (1-\beta)(y_{j,t} + \beta w_{j,t+1}) - (1-\beta)\delta_j - \tilde{\rho}_j(\tilde{p}_{j,t} - \beta\tilde{p}_{j,t+1})$$

is impacted by the discount factor. Thus, any bias associated with the discount factor will percolate through and into the estimation of the correlation parameter. Lastly, as is the case in much of the static choice literature where the variance covariance matrix is estimated, it is known that sizeable amounts of data are required to precisely estimate the parameter. This is made more clear with our second set of simulations which increases the time duration to 300 periods from 150. This increase doubles the amount of data and provides improvement in the estimation of ρ and the discount factor.

Finally, table 9 and 10 present the analysis where only 1 market is employed and T equals 150 or 300 periods. The results are similar to the set of simulations which employ 10 markets, but with less precision—most notably for ρ and ϕ.

C.2 Nested Logit Model

Next, we present the result of several Monte Carlo simulations with a nested logit data generating process. Particularly, we analyze the case where there of 3-5 products with product one relegated to one nest and all other products to a second nest. The within nested correlation for product one is normalized to 1 with the second nest taking the value of 0.80. The remaining data generating processes follows exactly as above in the simply MNL case.

We present the same variation of simulations as in the Logit case. The tables below illustrate that our estimator is able to precisely estimate the model primitives associated with the nested logit model.

Table 7: Monte Carlo Simulation Results: 10 Markets and 150 Periods

DGP: 10 Markets, $T = 150$

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7323 (0.0083)	-0.5003 (0.0075)	0.0963 (0.0063)	0.1405 (0.0637)	0.2609 (0.0686)	0.8153 (0.0116)
$J = 3$	0.7381 (0.0082)	-0.5003 (0.0075)	0.0941 (0.0088)	0.1199 (0.0655)	0.2538 (0.0480)	0.8192 (0.0161)
$J = 4$	0.7431 (0.0088)	-0.5002 (0.0077)	0.0911 (0.0102)	0.1403 (0.0600)	0.2473 (0.0435)	0.8253 (0.0190)
$J = 5$	0.7463 (0.0109)	-0.5001 (0.0074)	0.0913 (0.0121)	0.1736 (0.0582)	0.2422 (0.0436)	0.8257 (0.0222)

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7391 (0.0093)	-0.5003 (0.0075)	0.1053 (0.0025)	0.2086 (0.0692)	0.2555 (0.0684)	–
$J = 3$	0.7399 (0.0089)	-0.5003 (0.0075)	0.1050 (0.0017)	0.1902 (0.0560)	0.2496 (0.0472)	–
$J = 4$	0.7411 (0.0089)	-0.5002 (0.0077)	0.1054 (0.0017)	0.2126 (0.0490)	0.2426 (0.0424)	–
$J = 5$	0.7417 (0.0098)	-0.5001 (0.0084)	0.1057 (0.0015)	0.2318 (0.0428)	0.2381 (0.0425)	–

Mean and standard deviation for 250 simulations.

Table 8: Monte Carlo Simulation Results: 10 Markets and 300 Periods

DGP: 10 Markets, $T = 300$

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7494 (0.0058)	-0.5003 (0.0074)	0.1065 (0.0062)	0.2171 (0.0408)	0.2360 (0.0313)	0.7972 (0.0116)
$J = 3$	0.7457 (0.0086)	-0.5003 (0.0076)	0.1041 (0.0080)	0.2382 (0.0371)	0.2388 (0.0289)	0.8028 (0.0145)
$J = 4$	0.7473 (0.0114)	-0.5002 (0.0079)	0.1018 (0.0090)	0.2546 (0.0321)	0.2363 (0.0359)	0.8078 (0.0168)
$J = 5$	0.7483 (0.0134)	-0.5001 (0.0087)	0.1015 (0.0096)	0.2664 (0.0333)	0.2345 (0.0440)	0.8087 (0.0176)

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7496 (0.0057)	-0.5003 (0.0074)	0.1050 (0.0018)	0.2091 (0.0514)	0.2365 (0.0318)	–
$J = 3$	0.7451 (0.0062)	-0.5003 (0.0076)	0.1057 (0.0014)	0.2438 (0.0392)	0.2383 (0.0285)	–
$J = 4$	0.7443 (0.0070)	-0.5002 (0.0079)	0.1061 (0.0013)	0.2653 (0.0340)	0.2354 (0.0354)	–
$J = 5$	0.7443 (0.0083)	-0.5001 (0.0087)	0.1064 (0.0013)	0.2661 (0.0318)	0.2337 (0.0436)	–

Mean and standard deviation for 250 simulations.

Table 9: Monte Carlo Simulation Results: 1 Market and 150 Periods

DGP: 1 Market, $T = 150$

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7339 (0.0265)	-0.5041 (0.0249)	0.0990 (0.0402)	0.1462 (0.2087)	0.2471 (0.1975)	0.8167 (0.0386)
$J = 3$	0.7419 (0.0247)	-0.5048 (0.0234)	0.0971 (0.0314)	0.1301 (0.2272)	0.2380 (0.1293)	0.8190 (0.0503)
$J = 4$	0.7447 (0.0264)	-0.5040 (0.0233)	0.0922 (0.0607)	0.1374 (0.1906)	0.2273 (0.1181)	0.8260 (0.0607)
$J = 5$	0.7487 (0.0331)	-0.5034 (0.0250)	0.0893 (0.0369)	0.1678 (0.1825)	0.2220 (0.1246)	0.8324 (0.0684)

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7439 (0.0300)	-0.5041 (0.0249)	0.1075 (0.0091)	0.2211 (0.2205)	0.2407 (0.1951)	–
$J = 3$	0.7462 (0.0273)	-0.5084 (0.0234)	0.1077 (0.0143)	0.2088 (0.1964)	0.2364 (0.1304)	–
$J = 4$	0.7456 (0.0268)	-0.5040 (0.0233)	0.1069 (0.0068)	0.2178 (0.1774)	0.2227 (0.1154)	–
$J = 5$	0.7462 (0.0290)	-0.5034 (0.0250)	0.1078 (0.0147)	0.2464 (0.1597)	0.2168 (0.1205)	–

Mean and standard deviation for 250 simulations.

Table 10: Monte Carlo Simulation Results: 1 Market and 300 Periods

DGP: 1 Market, $T = 300$

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7510 (0.0197)	-0.5041 (0.0248)	0.1083 (0.0293)	0.2265 (0.1653)	0.2256 (0.0966)	0.7994 (0.0344)
$J = 3$	0.7486 (0.0267)	-0.5048 (0.0238)	0.1047 (0.0246)	0.2424 (0.1462)	0.2354 (0.0960)	0.8040 (0.0449)
$J = 4$	0.7503 (0.0337)	-0.5041 (0.0239)	0.1008 (0.0276)	0.2543 (0.1084)	0.2381 (0.1094)	0.8109 (0.0512)
$J = 5$	0.7515 (0.0416)	-0.5035 (0.0259)	0.1011 (0.0317)	0.2697 (0.0984)	0.2412 (0.1322)	0.8111 (0.0578)

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7527 (0.0191)	-0.5041 (0.0248)	0.1069 (0.0098)	0.2209 (0.1841)	0.2250 (0.0964)	–
$J = 3$	0.7490 (0.0206)	-0.5048 (0.0238)	0.1069 (0.0050)	0.2566 (0.1335)	0.2350 (0.0951)	–
$J = 4$	0.7477 (0.0223)	-0.5041 (0.0239)	0.1070 (0.0040)	0.2705 (0.1097)	0.2364 (0.1068)	–
$J = 5$	0.7483 (0.0253)	-0.5035 (0.0259)	0.1071 (0.0036)	0.2828 (0.0927)	0.2400 (0.1303)	–

Mean and standard deviation for 250 simulations.

Table 11: Monte Carlo Simulation Results: 10 Markets and 150 Periods

DGP: 10 Markets, $T = 150$

	$\delta = 0.75$	$\alpha = -0.5$	$\zeta = 0.8$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 3$	0.7339 (0.0352)	-0.5018 (0.0274)	0.7920 (0.1021)	0.0954 (0.0215)	0.1358 (0.0967)	0.2326 (0.0567)	0.8176 (0.0319)
$J = 4$	0.7367 (0.0381)	-0.5015 (0.0266)	0.7967 (0.0674)	0.0917 (0.0183)	0.1503 (0.1086)	0.2447 (0.0516)	0.8245 (0.0288)
$J = 5$	0.7405 (0.0243)	-0.5005 (0.0156)	0.8005 (0.0432)	0.0902 (0.0135)	0.1730 (0.0899)	0.2411 (0.0439)	0.8261 (0.0262)

	$\delta = 0.75$	$\alpha = -0.5$	$\zeta = 0.8$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 3$	0.7416 (0.0387)	-0.5018 (0.0274)	0.7920 (0.1021)	0.1039 (0.0057)	0.2029 (0.1625)	0.2261 (0.0527)	–
$J = 4$	0.7415 (0.0403)	-0.5015 (0.0266)	0.7967 (0.0674)	0.1041 (0.0051)	0.2186 (0.1476)	0.2384 (0.0477)	–
$J = 5$	0.7401 (0.0263)	-0.5005 (0.0156)	0.8005 (0.0431)	0.1038 (0.0036)	0.2327 (0.0810)	0.2412 (0.0423)	–

Mean and standard deviation for 250 simulations.

Table 12: Monte Carlo Simulation Results: 10 Markets and 300 Periods

DGP: 10 Markets, $T = 300$

	$\delta = 0.75$	$\alpha = -0.5$	$\zeta = 0.8$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 3$	0.7449 (0.0192)	-0.5008 (0.0156)	0.7974 (0.0426)	0.1024 (0.0115)	0.2390 (0.0515)	0.2300 (0.0415)	0.8019 (0.0204)
$J = 4$	0.7435 (0.0351)	-0.5003 (0.0241)	0.7991 (0.0477)	0.1011 (0.0151)	0.2507 (0.0769)	0.2386 (0.0322)	0.8058 (0.0247)
$J = 5$	0.7460 (0.0284)	-0.5003 (0.0197)	0.7998 (0.0442)	0.1002 (0.0112)	0.2609 (0.0773)	0.2420 (0.0298)	0.8071 (0.0223)

	$\delta = 0.75$	$\alpha = -0.5$	$\zeta = 0.8$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 3$	0.7455 (0.0157)	-0.5008 (0.0156)	0.7974 (0.0426)	0.1032 (0.0024)	0.2461 (0.0886)	0.2288 (0.0441)	–
$J = 4$	0.7440 (0.0238)	-0.5003 (0.0241)	0.7991 (0.0477)	0.1039 (0.0037)	0.2584 (0.0978)	0.2378 (0.0327)	–
$J = 5$	0.7437 (0.0273)	-0.5003 (0.097)	0.7998 (0.0442)	0.1040 (0.0039)	0.2697 (0.0764)	0.2413 (0.0292)	–

Mean and standard deviation for 250 simulations.

References

Abbring, J. H. and Ø. Daljord (2017). Identifying the discount factor in dynamic discrete choice models. Memo, University of Chicago.

Albuquerque, P. and B. J. Bronnenberg (2009). Estimating demand heterogeneity using aggregated data: An application to the frozen pizza category. *Marketing Science 28*(2), 356–372.

Allman, E. S., C. Matias, and J. A. Rhodes (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics 37*, 3099–3132.

Arcidiacono, P. and R. Miller (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica 79*(6), 1823–1867.

Bajari, P., C. S. Chu, D. Nekipelov, and M. 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.

Belomestny, D. (2003). Rates of convergence for constrained deconvolution problem. *arXiv preprint math/0306237*.

Belomestnyi, D. (2002). On the problem of characterizing the distribution of random variables by the distribution of their sum. *Journal of Mathematical Sciences 111*(3), 3498–3504.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S. T. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica 82*(5), 1749–1797.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016). Estimating multivariate latent-structure models. *The Annals of Statistics 44*(2), 540–563.

Chou, C. and G. Ridder (2017). Three-step ccp estimation of dynamic programming discrete choice models with large state space. Memo, University of Leicester.

Daljord, Ø., D. Nekipelov, and M. Park (2018). A simple and robust estimator for discount factors in optimal stopping dynamic discrete choice models. Memo, University of Chicago.

Derdenger, T. and V. Kumar (2018). Estimating dynamic discrete choice models with aggregate data: Small sample properties of the inclusive value approximation. Memo, Carnegie Mellon Univeristy.

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica 80*(5), 2231–2267.

Fang, H. and Y. Wang (2015). Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. *International Economic Review 56*(2), 565–596.

Goettler, R. L. and B. R. Gordon (2011). Does amd spur intel to innovate more? *Journal of Political Economy 119*(6), 1141–1200.

Gowrisankaran, G. and M. Rysman (2012). Dynamics of consumer demand for new durable goods. *Journal of political Economy 120*(6), 1173–1219.

Helmberg, G. (2008). *Introduction to Spectral Theory in Hilbert Space*. Mineola, New York: Dover Publications, Inc.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Hotz, V. and R. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies 60*(3), 497–529.

Hu, Y. and M. Shum (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics 171*(1), 32–44.

Hu, Y., M. Shum, W. Tan, and R. Xiao (2017). A simple estimator for dynamic models with serially correlated unobservables. *Journal of Econometric Methods 6*(1).

Kalouptsidi, M., P. T. Scott, and E. Souza-Rodrigues (2015). Identification of counterfactuals in dynamic discrete choice models. Working paper, NBER.

Kasahara, H. and K. Shimotsu (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica 77*(1), 135–175.

Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete decision processes. *Econometrica 70*(2), 801–816.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record* (673).

Melnikov, O. (2013). Demand for differentiated durable products: The case of the us computer printer market. *Economic Inquiry 51*(2), 1277–1298.

Miller, R. and S. Sanders (1997). Human capital development and welfare participation. *Carnegie-Rochester Conference Series on Public Policy* (46), 1–43.

Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions. In *Handbook of Statistics*, Volume 11, pp. 419–454. Elsevier.

Norets, A. (2009). Inference in dynamic discrete choice models with serially orrelated unobserved state variables. *Econometrica 77*(5), 1665–1682.

Norets, A. and X. Tang (2014). Semiparametric inference in dynamic binary choice models. *Review of Economic Studies 81*(3), 1229–1262.

Song, I. and P. K. Chintagunta (2003). A micromodel of new product adoption with heterogeneous and forward-looking consumers: Application to the digital camera category. *Quantitative Marketing and Economics 1*(4), 371–407.

Sudhir, K. (2013). Demand estimation using aggregate data: Static discrete choice models. Slides, Yale School of Management.

Sun, Y. and M. Ishihara (2018). A computationally efficient fixed point approach to dynamic structural demand estimation. *Journal of Econometrics Forthcoming*.