informs
http://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Articles in Advance, pp. 1–22
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Linear Estimation of Aggregate Dynamic Discrete Demand for Durable Goods: Overcoming the Curse of Dimensionality

Cheng Chou,[a] Tim Derdenger,[b] Vineet Kumar[c]

[a] University of Leicester, Leicester LE1 7RH, United Kingdom; [b] Carnegie Mellon University, Pittsburgh, Pennsylvania 15289; [c] Yale University, New Haven, Connecticut 06511

Contact: cchou@le.ac.uk (CC); derdenge@andrew.cmu.edu (TD); vineet.kumar@yale.edu, https://orcid.org/0000-0001-8784-6858 (VK)

Received: August 27, 2018
Revised: February 24, 2019
Accepted: April 22, 2019
Published Online in Articles in Advance: September 5, 2019

https://doi.org/10.1287/mksc.2019.1182

Abstract. We develop a new approach using market-level data to model, identify, and estimate a dynamic discrete choice demand model for durable goods with continuous unobserved product-specific state variables. They are specified as serially correlated and correlated with the observed product characteristics, particularly price. We provide a method to estimate all model primitives, including the consumer's discount factor and the state transition distributions of unobserved product characteristics without the need to reduce the dimension of the state space or by other approximation techniques, such as discretizing state variables. We prove the identification of model primitives and provide an estimation algorithm in which the most computationally demanding step is a linear regression. Finally, we show how it can be implemented in an application in which we estimate the demand for smartphones.

History: Yuxin Chen served as the editor-in-chief and Paul Ellickson served as associate editor for this article.
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mksc.2019.1182.

Keywords: dynamic discrete choice • curse of dimensionality • durable goods • demand estimation

1. Introduction

In recent years, dynamic discrete choice (DDC) models have become more prevalent in marketing and economics because of their ability to analyze the dynamic aspects of firms' and consumers' preferences and the consequent intertemporal trade-offs 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.[1] Estimation is further complicated when the unobserved product characteristics are serially correlated and correlated with observed state variables given that computing the ex ante expected value function involves 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 main contribution is to develop a novel approach using market-level data to model, identify, and estimate a dynamic discrete choice demand model for durable goods with *continuous unobserved product-specific state variables* in addition to the commonly included individual product idiosyncratic errors. The unobserved states or product characteristics are specified as serially correlated and correlated with the observed product characteristics, particularly price. We provide a method to estimate all model primitives, including the consumer's discount factor, without the need to reduce the dimension of the state space or by other approximation techniques, such as discretizing state variables. In this sense, our method avoids the curse of dimensionality—a large practical problem when implementing DDC models.

We provide rigorous proof of identification and an algorithm for estimation in which the most computationally demanding step is a linear regression. Following the sequence of linear regressions, applied researchers will have estimated all primitives of the dynamic structural model. The estimation simplicity and the absence of the curse of dimensionality aid model specification because the researcher no longer faces the trade-off between including more state variables with the feasibility of estimation or the dilemma of reducing the dimension of state variables at the cost of incurring omitted variable bias. Thus, researchers are able to estimate multiple model specifications at little computational cost. The major *limitations* of the method are that (a) there need to be two or more terminal choices in the DDC model (e.g., purchasing a product, then leaving the market permanently) and (b) the DDC model can only accommodate the multinomial logit or generalized extreme value (GEV) nested logit structure, not unobservable heterogeneity.

Our identification results are novel relative to the literature on identifying DDC models. Our model for durable goods can be understood as a general DDC model in which a subset of unobserved state variables (unobserved product characteristics herein) are continuous, serially correlated, and correlated with other observed state variables. The existing identification results (Magnac and Thesmar 2002; Kasahara and Shimotsu 2009; Norets 2009; Arcidiacono and Miller 2011, 2018; Hu and Shum 2012; Hu et al. 2017) in the literature cannot be applied here.

Most of the research focusing on individual-level data do not include persistent unobservable state variables (e.g., Bajari et al. 2016, Daljord et al. 2018).[2] The following exceptions involving persistent unobservables are worth noting. Hu and Shum (2012) study dynamic binary choice models with continuous unobserved state variables, but their identification result is limited to the conditional choice probabilities and state transition distribution functions, not to model primitives such as flow utility functions and discount factors. Norets (2009) does include a serially correlated unobservable *idiosyncratic error*, which is individual-specific rather than an aggregate product shock as in our case. Arcidiacono and Miller (2011) model persistent unobservables but limit them to a discrete set of values.

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

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

Extending the BLP models to a dynamic setting with forward-looking agents is challenging. Formal identification in these papers is not specified. Some researchers ignore persistent unobserved shocks (Song and Chintagunta 2003) or make them time-invariant (Goettler and Gordon 2011). Others focus on improving the computational speed of fixed point estimators by approximation. Melnikov (2013) and Gowrisankaran and Rysman (2012) approximate based on inclusive value sufficiency, allowing the researcher to collapse the multidimensional state into one dimension, making it more tractable. Derdenger and Kumar (2019) study the approximation properties of this approach and show that it is a biased and an inconsistent estimator. Su and Judd (2012) and Dubé et al. (2012) propose a constrained optimization approach to estimate static and dynamic structural models based on aggregate data. Also noteworthy are Sun and Ishihara (2019), who present a simple Monte Carlo–based approach to significantly diminish the burden of dynamic structural models.

Although the literature has made advances in computational tools that eliminate the costly nested fixed point algorithm used in dynamic models, our approach is different in that we have focused on proving identification of model primitives and also in that our approach avoids any computation of the value function in the estimation process.

The simplicity of our estimator is quite powerful, but it does come at a cost. In addition to the two previously mentioned limitations, we discuss two others. First, consumers in our model 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 (IV). That said, the model and estimator do allow for nonterminal choices when an individual is faced with a choice of, say, "lease one car" as long as the consumer choice does not affect the future transition of state variables. We do not track individual product inventory holdings. Thus, environments in which choices exhibit state dependence with repeat purchases would not be appropriately characterized.

Second, our computationally simple approach applies to a class of models similar to Berry (1994), that is, type 1 and GEV distribution for idiosyncratic errors. This limitation eliminates any possibility 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 are quite difficult in practice. Albuquerque and Bronnenberg (2009, p. 356) illustrate that, "in isolation neither variable ((market share or brand penetration)) may lead to precise estimates of heterogeneity." Sudhir (2013, p. 53) 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.

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 two approaches. The first is to simply draw from the identified distribution of only one unobserved state variable a large number of times to provide an identified set on the policy experiment. In practice, there is little to no added cost to this method as compared with what a researcher does in order to generate a confidence interval (draws from all parameters). Next, 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 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 is not identified. The intuition is similar to the identification of a linear panel data model in which 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. Thus, it is typically not a limitation in applications. 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 subsamples and estimate the preferences and/or the dynamic evolution of state variables for each subsample as long as there are enough periods in each sample.

There are a number of institutional features of an empirical context that make our model more suitable. Our approach is likely to prove useful in settings in which the dynamics and intertemporal trade-offs are of first-order importance to researchers and in which the state space is large, which reflect a number of empirical settings. Durable goods with a long replacement fit best, for example, solar panels. However, even products with a smaller replacement cycle would work if the discount factor is low. Because our method allows the researcher to recover the discount factor easily, one could simply run the model to determine suitability even when the researcher is not sure about the discount factor. The data required for the model is aggregate market-level data but allows significant flexibility in the nature of variation. Although our identification results only require $T = 2$ periods of data (with multiple markets), in practice, for estimation, a longer panel is helpful. Thus, the researcher can deploy this method even with data from only one market (e.g., national) or a smaller panel with data from multiple markets (e.g., states or metropolitan areas). The Monte Carlo studies in the Online Appendix Section O.1 demonstrate recovery for different combinations of markets and time periods.

We apply our estimator using data from the cell phone device market. Using monthly data from 10 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 with 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 with its absence 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 Section 2, we present the basic modeling approach. In Section 3, we detail the assumptions and show the identification for the model parameters. In Section 4, we obtain the estimators of preference parameters and state transition distribution. In Section 5, we discuss counterfactual implementation. In Section 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. In the appendix, we detail the GEV model along with the cellphone demand estimates using a nested logit model.

This paper comes with an online appendix, which contains (i) numerical studies about our estimators under various scenarios, (ii) discussion about the interpretation of the assumptions in an empirical context, (iii) implementation details about the counterfactual procedure, (iv) proofs of some identification theorems, (v) various formulas that are helpful for calculating the asymptotic variance of our estimators, and (vi) detailed discussion about the identification of nested logit specification.

2. Model

Our model follows the previous literature on dynamic discrete choice models of demand, particularly those that employ market-level data. 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 and 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 to purchase a product from the available products $\mathcal{J}_t \setminus \{0\}$. If consumer i decides to purchase, consumer i then chooses which to buy. Once a consumer has purchased a product, the consumer 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 greatly simplifies the identification and estimation because its expected lifetime utility 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 an 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). Conditional Probability $\Pr(\Omega_{i,t+1} \mid \Omega_{it}, \ \Omega_{i,t-1}, \ldots) = \Pr(\Omega_{i,t+1} \mid \Omega_{it})$.

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, that is, $j \notin \mathcal{J}_t$.

We normalize the expected period of utility of the outside good to be zero. Hence, if consumer i does not purchase in period t, consumer i 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 et al. 2017).

When consumer i purchases product j at time t, consumer i's flow utility *during the purchase period t* is

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

Consumer i then receives the identical flow utility $f(x_{jt}, \xi_{jt})$ in each period $\tau > t$ following consumer i's purchase. 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 to ξ_{jt} as the unobserved characteristics of product j at time t, which may be serially correlated. Possible interpretations of unobservable product–period-specific shocks ξ_{jt} are not limited to the following:

i. Product quality: If the firm has a quality control in the production process, then there is likely some degree of randomness or stochasticity in the manufacturing process. This would vary by product period and fit the assumptions about ξ in the paper. Note that, depending on the production process, this could also be serially correlated, which we accommodate in our model. In our application using cell phone data, ξ_{jt} can be thought as the quality of software on the phone, battery life, durability, etc.

ii. Advertising: We might have product-period unobservable advertising levels by both manufacturers or network carriers and retailers such as Best Buy in our application. Because advertising expenditure decisions are set well in advance, it is quite likely for such expenditures to be serially correlated.

The inclusion of these unobserved product characteristics (states) are important. The data the researcher collects to estimate demand models is almost always incomplete as it does not contain all the state variables that consumers use to make their decisions. Ignoring unobserved product characteristics could bias the estimation of price coefficients was first discussed in the work of Berry (1994) about the endogeneity of price.

Another econometric problem when one only uses idiosyncratic errors as in Bajari et al. (2016) is that, if the data-generating process had product-period unobservables (e.g., advertising or quality control variations over time) but were ignored, then the idiosyncratic errors would pick up those factors as in Song and Chintagunta (2003). In such a case, we would have correlation of idiosyncratic errors across individuals and time if the unobservable product characteristics were serially correlated. Because almost all papers effectively specify such idiosyncratic errors to be independent across agents, this would lead to a misspecification and biased parameter estimates.

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 product characteristics (or price) are expected to improve over time, the consumer has incentive to wait.

Consumer i in period t chooses from the set of choices $\mathcal{J}_t$, which includes the option zero to wait without purchasing any product. However, if the consumer purchases, recall that the consumer 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:

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

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:

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

We also let

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

The value function $V_t(\Omega_{it})$ involves consumer i's flow utility shock ε_{it}. Assumption 2(i) ensures the following in which the expectation in the preceding display is taken over $\varepsilon_{i,t+1}$:

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

where

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

Assumption 2 (Conditional Independence). *For all t, we have* (i) $\Omega_{i,t+1} \perp\!\!\!\perp \varepsilon_{it} \,|\, (x_t, p_t, \xi_t)$; (ii) $\varepsilon_{i,t+1} \perp\!\!\!\perp \Omega_{it} \,|\, (x_{t+1}, p_{t+1}, \xi_{t+1})$.

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 $F(\cdot \,|\, m_t)$ of ε_{it}, we have

$$s_{jt}(m_t) = \int 1\big(v_{jt}(m_t) + \varepsilon_{ijt} \geq v_{kt}(m_t) + \varepsilon_{ikt}, k \in \mathcal{J}_t\big) \cdot F(\mathrm{d}\varepsilon_{it} \,|\, m_t). \tag{4}$$

Our results 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 $F(\Omega_{i,t+1} \,|\, \Omega_{it})$, and the initial distribution function $F(\Omega_{it})$ for some period t. In general, we need to know θ_1, $F(\Omega_{i,t+1} \,|\, \Omega_{it})$ 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

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

Moreover, we assume that $\varepsilon_{it} \perp\!\!\!\perp m_t$ and $F(\varepsilon_{it})$ are known for all t. We can write $F(m_t) = F(x_t, p_t) \cdot F(\xi_t \mid x_t, p_t)$. Thus, the cumulative distribution function $F(x_t, p_t)$ 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 IV that are uncorrelated with unobserved characteristics 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 characteristics $\xi_{1t}, \ldots, \xi_{Jt}$ 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 unobserved characteristics 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 (AR) process and $x_{t+1} \perp\!\!\!\perp (\xi_t, \xi_{t+1}) \mid (x_t, p_t)$ or $(x_{t+1}, p_{t+1}) \perp\!\!\!\perp \xi_t \mid (x_t, p_t)$. Most identification results are constructive; hence, they can be used as formulas for estimation.

It is well known that, without assuming that $F(\varepsilon_{it} \mid m_t)$ is known, the flow utility functions and discount factor are not separately identified (e.g., Magnac and Thesmar 2002). 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 is a type I extreme value distribution. In the online appendix, we present identification for a GEV distribution.

Assumption 3. *Assume that consumer i's utility shocks $\varepsilon_{it} = (\varepsilon_{i0t}, \ldots, \varepsilon_{iJt})'$ are independent of $m_t = (x_t, p_t, \xi_t)$. Let $\varepsilon_{i0t} + \omega, \ldots, \varepsilon_{iJt} + \omega$ be an independent identically distributed type I extreme value with density $f(\varepsilon_{ijt} + \omega = \varepsilon) = \exp[-(\varepsilon + e^{-\varepsilon})]$, where $\omega \approx 0.5772$ is Euler's constant.*

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 because ξ_t in m_t is unobservable.

It should be remarked that the assumption of independent idiosyncratic shocks is required, but the assumption on type I distribution or any specific distribution is not essential for our identification arguments because our arguments start from expressing the difference between the payoffs of purchasing different products as a function of market shares, which holds for more general distribution of ε_{it} (Hotz and Miller 1993). However, it greatly simplifies the exposition and estimation.

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. Our arguments can also be applied to show identification when the choice set varies over time. 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

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

To show identification, we only use $\ln(s_{2t}/s_{1t})$ and $\ln(s_{2t}/s_{0t})$. Equation (5) is similar to Berry (1994). The key difference is that $v_{0t}(m_t)$ in Berry (1994) or BLP equals zero, and $v_{0t}(m_t)$ here depends on an unknown value function.

In Equation (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 (6) explains the relative market share by the difference of product characteristics. Equation (6) resembles a linear regression because 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

$$\begin{aligned} &\mathrm{E}(g_{1,(2,1),t}(\theta_{1o})) = 0, \\ &g_{1,(2,1),t}(\theta_1) = z_{(2,1),t}\left[\ln\left(\frac{s_{2t}}{s_{1t}}\right) - (x_{2t} - x_{1t})'\tilde{\gamma} \right. \\ &\qquad \left. + \alpha(p_{2t} - p_{1t}) - \frac{\delta_2 - \delta_1}{1-\beta}\right]. \end{aligned} \tag{7}$$

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

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

Define the already identified term $y_t \equiv \ln(s_{2t}/s_{0t}) - x_{2t}'\tilde{\gamma} + \alpha p_{2t}$: Note that y_t is a function of m_t only. We rewrite Equation (8) with $y_t(m_t)$,

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

By the expectation maximization formula for multinomial logit (e.g., Arcidiacono and Miller 2011):

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

Define another identified term w_t, a function of m_t only, and the corresponding value function

$$\begin{aligned} w_t &\equiv x_{2t}'\tilde{\gamma} - \alpha p_{2t} - \ln s_{2t}(m_t) \\ \bar{V}_t(m_t) &= w_t(m_t) + \frac{\delta_2}{1-\beta} + \frac{\xi_{2t}}{1-\beta}, \qquad \text{for all } t. \end{aligned}$$

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

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

Because $y_t(m_t)$ 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 preceding 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 Equation (12) is 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

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

We use $\mathrm{E}(\xi_{2t}) = \mathrm{E}(\xi_{2,t+1}) = 0$. Because $(\delta_2 - \delta_1)/(1-\beta)$ is identified, we identify δ_1 given δ_2 and β.

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

$$\begin{aligned} &\mathrm{E}(g_{2,(2,0),t}(\theta_{1o})) = 0, \\ &g_{2,(2,0),t}(\theta_1) = (y_t + \beta w_{t+1} - \delta_2,\ (y_t + \beta w_{t+1} - \delta_2)x_{2t,IV}')'. \end{aligned} \tag{14}$$

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

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

provided that $\mathrm{cov}(w_{t+1}, x_{2t,IV}) \neq 0$ (corresponding to rank condition in IV regression). From the definition of w_{t+1}, the rank condition requires that $x_{2t,IV}$ 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 Assumptions 1–3 hold. Let $d_x = \dim x_{jt}$. If there is a vector of IV $z_{(2,1),t}$ 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 $x_{2t,IV}$ of m_t such that $\mathrm{cov}(x_{2t,IV}, \xi_{2t}) = \mathrm{cov}(x_{2t,IV}, \xi_{2,t+1}) = 0$ and $\mathrm{cov}(w_{t+1}, x_{2t,IV}) \neq 0$, we can identify consumer preference parameters α, β, and γ and product fixed effect δ with two periods of data.*

Moreover, these constructive identification arguments suggest a simple estimation method for $\theta_1 = (\alpha, \beta, \delta', \gamma')'$. An IV regression can estimate $\tilde{\gamma}$, $(\delta_2 - \delta_1)/(1-\beta)$, and α. Another IV regression of y_t on $-w_{t+1}$ with IV $x_{2t,IV}$ can be used to estimate the discount factor β. 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 literature, 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, by $\ln(s_{2t}/s_{0t}) = v_2(x_{2t}, p_{2t}) - v_0(m_t)$, and $v_0(m_t) = \beta \mathrm{E}(\bar{V}_{t+1}(m_{t+1}) \,|\, m_t)$, we have

$$\ln(s_{2t}/s_{0t}) = v_2(x_{2t}, p_{2t}) - \beta \mathrm{E}\big[v_2(x_{2,t+1}, p_{2,t+1}) - \ln s_{2,t+1} \,|\, m_t\big]. \tag{15}$$

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 need $x_{2t,IV}$, a nonrandom function of m_t. From Equation (15), we have

$$\begin{aligned}&\mathrm{E}(\ln(s_{2t}/s_{0t}) \,|\, x_{2t,IV})\\ &\quad = \mathrm{E}(v_{2t} \,|\, x_{2t,IV}) - \beta \mathrm{E}(v_{2,t+1} - \ln s_{2,t+1} \,|\, x_{2t,IV}).\end{aligned}$$

Here we use v_{2t} to denote $v_2(x_{2t}, p_{2t}, \xi_{2t})$. 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 display by $\mathrm{E}(\xi_{2t} \,|\, x_{2t,IV}) = \mathrm{E}(\xi_{2,t+1} \,|\, x_{2t,IV}) = 0$.

In the online appendix, we show that, 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.

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. Although the identification of consumer preferences did not require us to assume stationarity of the state evolution process, stationarity is necessary for us to identify the state transition distribution.

Assumption 4 (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 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 $\mathrm{E}(\xi_{jt} \,|\, x_t, p_t)$ with the stationary Assumption 5 about ξ_t. Then we show nonparametric identification of $F(\xi_t \,|\, x_t, p_t)$ with additional restrictions.

Assumption 5. (i) *The marginal distribution function $F(\xi_t)$ and the conditional distribution function $F(\xi_t \,|\, x_t, p_t)$ are both time invariant.* (ii) $\xi_{t+1} \perp\!\!\!\perp (x_t, p_t) \,|\, (x_{t+1}, p_{t+1})$.

Though Assumption 5(i) is implied by Assumption 4, we state it separately because it involves unobserved characteristics ξ_t whose interpretation depends on empirical applications. It is more informative to applied researchers to state the restriction about ξ_t separately.

By Equation (6) and the identification of β, we can identify $d_t \equiv \xi_{2t} - \xi_{1t}$. It follows from Equation (6) that

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

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 a conditioning variable in the preceding 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 preceding 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 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}. In practice, after the estimation of model primitives, one can try to solve ξ_{2t} from the equation numerically by trying random starting guesses of the solution. If the equation has multiple solutions, the numerical solution is likely to depend on the choice of starting values. We tried to solve ξ_{2t} for our model used in the Monte Carlo studies reported in the online appendix and found that the solution of ξ_{2t} does not depend on the starting values, which suggests that ξ_{2t} is identifiable for that model.

One sufficient yet uninteresting condition is that $(x_{2,t+1}, p_{2,t+1}, s_{2,t+1}, \xi_{2,t+1}) \perp\!\!\!\perp \xi_t \mid (x_t, p_t)$. 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, the 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 the next period.

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

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

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

$$\mathrm{E}\left[(1-\beta)y_t + \beta(1-\beta)w_{t+1} - (1-\beta)\delta_2 - \xi_{2t} + \beta\xi_{2,t+1} \mid x_t, p_t\right] = 0. \tag{16}$$

Now define

$$\begin{aligned} h(x,p) &\equiv \mathrm{E}\big[(1-\beta)y_t + \beta(1-\beta)w_{t+1} \\ &\qquad - (1-\beta)\delta_2 \mid x_t = x, p_t = p\big], \\ \pi(x,p) &\equiv \mathrm{E}(\xi_{2t} \mid x_t = x, p_t = p). \end{aligned}$$

The function $h(x,p)$ is nonparametrically identified because we observe y_t, w_{t+1}, x_t and p_t. The unknown function $\pi(x,p)$ is the parameter of interest. It can be shown (see the online appendix for the proof) that Equation (16) can written as a Fredholm integral equation of type 2,

$$\pi(x,p) - \beta \int \pi(x', p') F(\mathrm{d}x', \mathrm{d}p' \mid x, p) = h(x,p). \tag{17}$$

We know that there would be a unique solution of $\pi(x,p)$ (the proof is similar to lemma 2 of Chou and Ridder (2017)).

Proposition 2. *In addition to the conditions of Proposition 1, suppose Assumptions 4 and 5 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 6. (i) *The unobserved characteristics* $\xi_{1t}, \ldots, \xi_{Jt}$ *are independent conditional on* (x_t, p_t); (ii) *assume that* $\mathrm{Var}(\xi_{1t} \mid x_t, p_t) = \cdots = \mathrm{Var}(\xi_{Jt} \mid x_t, p_t) = \sigma^2(x_t, p_t)$.

Using $d_t = \xi_{2t} - \xi_{1t}$, it can be shown that

$$\mathrm{E}(d_t^2 \mid x_t, p_t) = 2\sigma^2(x_t, p_t) + \left[\mathrm{E}(\xi_{2t} \mid x_t, p_t) - \mathrm{E}(\xi_{1t} \mid x_t, p_t)\right]^2.$$

Because 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 preceding display.

As for the unconditional variance, we use

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

Moreover, $\mathrm{E}(\xi_{jt}^2 \mid x_t, p_t) = \sigma^2(x_t, p_t) + \mathrm{E}(\xi_{jt} \mid x_t, p_t)^2$. Because 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 6 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 the unobserved product characteristics (Assumption 6(i)), we can write $F(\xi_{1t}, \ldots, \xi_{Jt} \mid x_t, p_t) = \prod_{j=1}^{J} F(\xi_{1t} \mid x_t, p_t)$. If the conditional distribution of ξ_{jt} given x_t, p_t belongs to the location scale family, the conditional mean and variance 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 7. *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.*

Proposition 4. *In addition to the conditions of Proposition 3, suppose Assumption 7 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.*

Proof. See the online appendix.

Remark 2 (Heteroskedasticity)**.** When we have three or more products, we only need to assume that there are at least two products whose conditional variance $\mathrm{Var}(\xi_{jt} \mid x_t, p_t)$ is the same. To see this, suppose there are three 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 Equation (6), we have

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

which is identified. By the same arguments, we have

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

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

3.2.2. $F(m_{t+1} \,|\, m_t)$**.** Note that $m_t = (x_t, p_t, \xi_t)$ and ξ_t is a $J \times 1$ vector. We are going to show the semiparametric identification of $F(m_{t+1} \,|\, m_t)$ by restricting the relationship between ξ_{t+1} and m_t to have a certain linear functional form.

We present two versions of identification results under two different Assumptions 8 and 8′. Under either assumption, the conclusion $F(m_{t+1} \,|\, m_t)$ is identified. Depending on the context of one's empirical research, one may find one assumption is more appropriate than the other. Roughly speaking, Assumption 8 is more appropriate if ξ_t can be understood as design or product quality, which can affect the price. Assumption 8′, however, is more appropriate if ξ_t can be understood as the spending of advertisement that is determined based on the product price.

Assumption 8. *Assume that* (i) $\xi_{t+1} \perp\!\!\!\perp (x_t, p_t) \,|\, \xi_t$, (ii) $x_{t+1} \perp\!\!\!\perp (\xi_t, \xi_{t+1}) \,|\, (x_t, p_t)$, (iii) *and* $p_{t+1} \perp\!\!\!\perp (x_t, p_t, \xi_t) \,|\, (x_{t+1}, \xi_{t+1})$. *This implies that the following decomposition*

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

(iv) Assume that $F(\xi_{t+1} \,|\, \xi_t) = F(\xi_{1,t+1} \,|\, \xi_{1t}) \cdots F(\xi_{J,t+1} \,|\, \xi_{Jt})$, *and* $\xi_{j,t+1}$ *and* ξ_{kt} *are uncorrelated for any two distinct products j and k, and (v)* $\xi_{j,t+1} = \phi_j \xi_{jt} + \nu_{j,t+1}$, *where* $\nu_{j,t+1}$ *has mean zero and is independent of* ξ_{jt}.

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

Proposition 5. *In addition to the conditions of Proposition 4, suppose Assumption 8 holds. We can identify* $F(\xi_{t+1} \,|\, \xi_t)$, *henceforth,* $F(m_{t+1} \,|\, m_t)$.

Proof. See the online appendix.

Assumption 8′. *Assume that* (i) $(x_{t+1}, p_{t+1}) \perp\!\!\!\perp \xi_t \,|\, (x_t, p_t)$, (ii) $\xi_{t+1} \perp\!\!\!\perp (x_t, p_t) \,|\, (x_{t+1}, p_{t+1}, \xi_t)$. *This implies the following decomposition:*

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

(iii) *Assume that* $F(\xi_{t+1} \,|\, x_{t+1}, p_{t+1}, \xi_t) = \prod_{j=1}^{J} F(\xi_{j,t+1} \,|\, x_{j,t+1}, p_{j,t+1}, \xi_{jt})$, and *(iv)* $\xi_{j,t+1} = \phi_{0j} + \phi_{1j}\xi_{jt} + \phi_{2j} p_{j,t+1} + \phi_{3j}' x_{j,t+1} + \nu_{j,t+1}$, *where* $\nu_{j,t+1}$ *has mean zero and is independent of* $(\xi_{jt}, p_{j,t+1}, x_{j,t+1})$.

Proposition 5′. *In addition to the conditions of Proposition 4, suppose Assumption 8′ holds. We can identify* $F(\xi_{t+1} \,|\, x_{t+1}, p_{t+1}, \xi_t)$, *henceforth* $F(m_{t+1} \,|\, m_t)$.

Proof. See the online appendix.

Remark 3 (Nonterminal Choices)**.** Up to now, we have assumed that a consumer's choice is terminal, for example, "buy one car then exit the market." We can extend the analysis to allow for nonterminal choice, for example, "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:

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

Recall that $v_{0t} = \beta \mathrm{E}(\bar{V}_{t+1}(m_{t+1}) \,|\, m_t)$. 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 nonterminal choice can be accommodated.

It should be remarked that the main reason why this works is that the current nonterminal choice, "lease one car" today, does not affect the future market state m_{t+1}. Hence, the expected future payoff $\mathrm{E}(\bar{V}_{t+1}(m_{t+1}) \,|\, m_t)$ 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.

Remark 4 (Changing Choice Set)**.** It is important to remark that our identification (preceding) and estimation arguments (following) rely only on the need for the same products to exist in two consecutive periods. Of course for estimation, one needs multiple markets when there are only two periods of data available. Consider one example: in periods 1 and 2, there are two products a and b, and in periods 3 and 4, there are two products b and c. This example has both entry (product c) and exit (product a). In each period, we assume we have enough markets. To identify/estimate the preference parameters, we can use all four periods by taking account of the log market share ratio between b and a for periods 1 and 2, b and c for periods 3 and 4. To identify/estimate the product-specific correlation between price and unobserved product characteristic and the serial correlation of unobserved product characteristic, we can use periods 1 and 2 for product a, periods 1–4 for product b, and periods 3 and 4 for product c.

4. Estimation

For simplicity of exposition, we focus on the case that the data are from one single market, for example, the United States, over T consecutive periods. Both numerical studies and empirical application show the case with multiple markets. We first describe the estimation routine for parameters and then describe the variance estimation. The estimation routine involves only IV and linear regressions. We then discuss the assumptions made in the paper, followed by comments about data requirements in practice. However, for applied marketers who are only interested in understanding how to implement our procedure, we include Table 1, which concisely presents the six ordinary least square (OLS) or IV regressions that are required to estimate the model primitives.

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:

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

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

The vector $z_{(j,k),t}$ is a vector of IV that is uncorrelated with $(\xi_{jt} - \xi_{kt})$. The moment equation follows from Equation (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, for example, run a k-means clustering by price and consider only intercluster 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, for example, they come from the same cluster, the instrument $z_{(j,k),t}$ is likely to be weak.

Table 1. Estimation Recipe

Step	Dependent variable	Independent variables[a]		IV
1	$\ln(s_{jt}/s_{kt})$	$(x_{jt} - x_{kt})$	$-(p_{jt} - p_{kt})$	$z_{(j,k),t}$[b]
		$\tilde{\gamma}$	α	
2	$\hat{y}_{jt}$‡	$-\hat{w}_{j,t+1}$‡		$x_{jt,IV}$[c]
		β		
3	$(\hat{y}_{jt} + \hat{\beta}\hat{w}_{j,t+1})$	1[d]		OLS
		δ_j		
4	$(1-\hat{\beta})(\hat{y}_{jt} + \hat{\beta}\hat{w}_{j,t+1})$	$(\tilde{p}_{jt} - \hat{\beta}\tilde{p}_{j,t+1})$		$z_{\rho,jt}$[e]
		$\tilde{\rho}_j$		
5	$\hat{d}^2_{(j,k),t}/2 + \hat{\tilde{\rho}}_j\hat{\tilde{\rho}}_k\tilde{p}_{jt}\tilde{p}_{kt}$ ‡	1		OLS
		σ^2		
6	$\hat{d}^2_{(j,k),t}/(2\hat{\beta}\hat{\sigma}^2) - [(1-\hat{\beta})/\hat{\beta}](\hat{y}_{jt} + \hat{\beta}\hat{w}_{j,t+1})\hat{d}_{(j,k),t}/\hat{\sigma}^2$	1		OLS
		ϕ_j		

‡Variable definitions:

1. $\hat{y}_{jt} = \ln\left(\frac{s_{jt}}{s_{0t}}\right) - x'_{jt}\hat{\tilde{\gamma}} + \hat{\alpha}p_{jt}$;
2. $\hat{w}_{jt} = x'_{jt}\hat{\tilde{\gamma}} - \hat{\alpha}p_{jt} - \ln s_{jt}$;
3. $\hat{d}_{(j,k),t} = (1-\hat{\beta})\left[\ln\left(\frac{s_{jt}}{s_{kt}}\right) - (x_{jt} - x_{kt})'\hat{\tilde{\gamma}} + \hat{\alpha}(p_{jt} - p_{kt}) - \frac{\hat{\delta}_j - \hat{\delta}_k}{1-\hat{\beta}}\right]$.

[a]The (IV) regression coefficient estimates associated with independent variables are estimates of the parameters underneath the independent variables.
[b]The IV $z_{(j,k),t}$ is uncorrelated with $(\xi_{jt} - \xi_{kt})$.
[c]The IV $x_{jt,IV}$ is uncorrelated with ξ_{jt} and $\xi_{j,t+1}$.
[d]This indicates regression with intercept term only.
[e]$z_{\rho,jt}(p_{jt})$ is a vector of functions of p_{jt}, for example, $z_{\rho,jt}(p_{jt}) = (p_{jt}, p^2_{jt}, \ldots, p^K_{jt})'$ for some integer $K \ge 1$ or the optimal IV formula, Equation (O.11) in the online appendix.

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

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

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

4.1.2. Discount Factor.

Step 2. Estimate β using

$$\mathrm{E}(g_{2,(j,0),t}(\theta_{1o})) = \mathrm{E}(x_{jt,IV}(y_{jt} + \beta w_{j,t+1} - \delta_j)) = 0, \qquad \text{for} \quad 0 < j < J,$$

This moment equation follows from Equation (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 preceding moment equation when $x_{jt,IV} = 1$. 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 is used in the estimation of the other parameters,

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

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, we assume that the distribution $F(x_t, p_t)$ and $F(x_{t+1} \mid x_t, p_t)$ are known.

Assumption 9.

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* $(p_{jt}, \xi_{jt})'$ *follows a bivariate normal distribution with mean* $(\mu_{pjt}, 0)'$, $\mathrm{Var}(p_{jt}) = \sigma^2_{pjt}$, $\mathrm{Var}(\xi_{jt}) = \sigma^2$, *and* $\mathrm{cov}(p_{jt}, \xi_{jt}) = \rho_j\sigma\sigma_{pjt}$. *Let* $\tilde{p}_{jt} = (p_{jt} - \mu_{pjt})/\sigma_{pjt}$ *be the standardized 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 $\tilde{\rho}_j = \rho_j\sigma$. 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 Characteristics.

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

$$g_{3,j,t}(\theta) = z_{\rho,jt}(p_{jt})r_{jt}, \tag{18}$$
$$r_{jt} = (1-\beta)(y_{jt} + \beta w_{j,t+1}) - (1-\beta)\delta_j - \tilde{\rho}_j(\tilde{p}_{jt} - \beta\tilde{p}_{j,t+1}).$$

Here $z_{\rho,jt}(p_{jt})$ is a vector of functions of p_{jt}, for example, $z_{\rho,jt}(p_{jt}) = (p_{jt}, p^2_{jt}, \ldots, p^K_{jt})'$ for some integer $K \geq 1$. We discuss the optimal choice of $z_{\rho,jt}(p_{jt})$. 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 $z_{\rho,jt}$. The online appendix collects the detailed steps for obtaining the moment equations of Steps 4–6.

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.

4.2.2. Variance of Unobserved Product Characteristics.

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 using data $t = 1, \ldots, T$ and all selected pairs of products. Knowing $\tilde{\rho}_j = \rho_j\sigma$ and σ, we know the joint distribution of (ξ_{jt}, p_{jt}).

4.2.3. Serial Correlation of Unobserved Product Characteristics.

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.

4.3. Asymptotic Variance and the Advantages of Sequential Estimation

There are two major reasons why we favor sequential estimation. First, sequential estimation is numerically reliable. Regardless of the dimension of the state variables, each step in the sequential estimation involves only linear or IV regression; hence, there is no numerical optimization. This also implies that the output of the estimation does not suffer from any randomness, always results in the same estimates, and does not require us to specify any starting point.

Second, sequential estimation is empirically appealing. Most steps in our sequential estimation process involve either running linear IV regression or straightforward computations. The identification and estimation involves finding relevant IV. One can then easily test whether one's chosen IVs are weak or not by standard statistical tests in linear IV regression. Using sequential estimation, one can immediately see the significance of the selected observed product characteristics by software routines.

Remark 5 (Estimating Standard Errors in Practice)**.** Even though we present a sequential estimator, one should not use the standard errors reported along these sequential steps. This is because these standard errors do not take into account the sampling error from the estimation in the earlier steps. One should instead use the asymptotic variance formulas that are derived assuming a joint parametric generalized method of moments (GMM) estimation process. See the online appendix for details. To evaluate those formulas, one plugs in the parameter estimates from either the sequential or joint procedure because they are both consistent. Thus, to summarize, the estimation of parameters is carried out sequentially, but the estimation of standard errors is done jointly.

4.4. Summary of Assumptions

Given that we have presented numerous assumptions for identification and estimation of parameters, we provide a summary of them in Table 13 of the online appendix highlighting the settings in which the assumptions are consistent and inconsistent. There are a few points that deserve further explanation.

First, we note that Assumptions 1–4 and similar ones are standard in the dynamic structural models literature (Hotz and Miller 1993, Rust 1994, Bajari et al. 2007). For the estimation of preference parameters, only Assumptions 1–3 are required.

Assumptions 5–9 on unobservable state variables (product-period specific) are mostly new because most prior research does not consider a persistent unobservable state.[6] These are required for identification of the observable and unobservable state evolution process. Third, Assumption 9 is only required to make parametric estimation possible because of data limitations encountered in practice. There, we specify a bivariate normal distribution, which characterizes the contemporaneous dependence between price and the unobservable product characteristics. We also specify that an unobservable characteristic for a product j only depends on its price and not the price of other products $k \neq j$.

Finally, many of the assumptions are made on conditional distributions or moments of conditional distributions. Typically, the conditioning variables are some combination of observables x_{jt} and p_{jt}. Thus, the dimension of observable product characteristics x_{jt} and their variation play a significant role in the assumption. In situations in which we have many product characteristics or when they span a greater support, the assumptions that restrict conditional moments could be viewed as less restrictive. Because our method does not have to worry about the curse of dimensionality from increasing the dimension of x_{jt}, one can make these assumptions less restrictive by adding more observable product characteristics if data permitted. It is easy to understand from the following example. Suppose there are two products, and they are dishes served by two restaurants. Let ξ_{1t} and ξ_{2t} be the unobserved taste of the two dishes. Assumption 6(ii) in the paper says that $\mathrm{Var}(\xi_{1t} \mid x_t, p_t) = \mathrm{Var}(\xi_{2t} \mid x_t, p_t)$. Without any conditioning variables, Assumption 6(ii) requires $\mathrm{Var}(\xi_{1t}) = \mathrm{Var}(\xi_{2t})$, which is strong. However, if the conditioning variables include food ingredients, recipes, tenure of chef, etc., it is reasonable to assume Assumption 6(ii).

Another point is worth noting on the conditional moment restrictions. The larger the support of the conditioning variables, (x_t, p_t), the less restrictive the assumptions are. If, on the other hand, all product characteristics are binary and prices show no variation, then the restrictions become stronger. For example, if we have an X variable that indicates whether the smartphone supports music or not (binary), that would be more restrictive. If, on the other hand, the music variable actually indicated support for different formats (e.g., MP3, WAV), we can view it as less restrictive.

4.5. Data Requirements for Estimation

We now discuss the data requirements for employing our estimator. For consumer parameters α, β, γ, it uses the data on all products J across markets M and time periods T. For product-specific parameters (δ_j fixed effect and evolution of state-space parameters), the length of the panel T and markets M is relevant. With Assumption 9 (normal distribution), estimation reduces to a linear regression. Thus, the realistic sample size would be comparable to the sample size that the researcher would use for a linear regression.

With regard to the nonparametric estimation within certain steps, there are only two instances. They are $\mathrm{E}(\tilde{p}_{j,t+1} \mid p_{jt})$ and $\mathrm{E}(r_{jt}^2 \mid p_{jt})$ in the construction of the optimal IV for estimation Step 4.

First, one does not have to use the optimal IV. Instead, one can use a sequence of polynomials of price as the IV, and the estimator is still consistent though inefficient. Similarly, even if the nonparametric estimates $\hat{\mathrm{E}}(\tilde{p}_{j,t+1} \mid p_{jt})$ and $\hat{\mathrm{E}}(r_{jt}^2 \mid p_{jt})$ of the conditional expectations have large estimation error, the constructed optimal IV is still a valid IV (because the nonparametric estimates are still a function of price p_{jt}), and hence, the estimation is valid though it's no longer efficient.

Second, the nonparametric regressions for estimating $\mathrm{E}(\tilde{p}_{j,t+1} \mid p_{jt})$ and $\mathrm{E}(r_{jt}^2 \mid p_{jt})$ involve only one single regressor, p_{jt}. Hence, it does not require much more data than the linear regression of $\tilde{p}_{j,t+1}$ on p_{jt} or the linear regression of r_{jt}^2 on p_{jt}. It is known that if one used a linear regression to estimate the conditional expectation, the mean-squared-error (MSE) is of order $O(n^{-1})$, and the MSE of nonparametric regression has the order of $O(n^{-4/5})$. In other words, if one believes that 20 observations is sufficient to estimate the linear regression of $\tilde{p}_{j,t+1}$ on p_{jt}, then 25 observations are enough for its nonparametric regression counterpart.

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.

In addition to the stationary Assumption 4, any counterfactual analysis requires the following:

i. Consumer preference: Preference parameters do not change under the counterfactual.

ii. Consumer expectations: Expectations are specified (e.g., rational expectations or perfect foresight).

iii. State evolution: Determine how the state variables evolve. The typical assumption is that observable states evolve in the same manner as the evolution process present in data although the researcher is free to specify a different evolution process and then compute counterfactual outcomes for that case.

As long as we have these, we can perform a counterfactual analysis by simulating individual consumer choices under the counterfactual setting because primitives for the agent and the state evolution parameters are identified. All that is required are assumptions on consumer expectations and what beliefs consumers have about the evolution of the state space (e.g., consumers track the evolution of each individual product's characteristics, the conditional value function of each good, or a market statistic such as the inclusive value). As a result, we are able to employ our model primitives to examine the impact of a change of any of the observed characteristics in the flow utilities, a price change, exit of a product, early entry of an observed product, and policies that change consumer expectations.[7]

In order to implement any counterfactual exercises and recover the impact of market share or revenues, we must specify ξ_{1t}. With this, all other ξ_{jt} are identified because $\xi_{jt} - \xi_{1t}$ is identified. One such approach is to simulate ξ_{1t} from its estimated AR(1) process and determine the ex ante value function for each draw to obtain a range of counterfactual results.[8]

6. Empirical Application

We now examine an empirical setting in which we use our method 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 United States during 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 six 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. More specifically, variables are generated using a weighted average based on sales in each period.

Table 2 shows the basic summary statistics of the market by showing the mean of product characteristics for each brand in the sample.[9] 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 representing smartphones, whereas the rest were primarily focused on feature phones (or dumb phones) during this time frame. The "x" variables are observable

Table 2. 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	

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 talk time in hours (*xtalktime*). Typical battery life was measured in hours of talk time, 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, but some of this support was tied in to carrier-based music services (such as downloading tones), which were quite expensive because customers of a carrier such as 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 five hours of talk time before the battery was depleted.

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, consumer i receives flow utility $\delta_j + x'_{j\ell t}\gamma + \xi_{j\ell t} + \varepsilon_{ij\ell\tau}$ in each following period $\tau > t$. The no-purchase option is modeled as receiving a period utility of zero with an option to continue in the market as in Section 2. Consumers who purchase exit the market and, thus, can be modeled as receiving 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 $\beta \mathrm{E}(\bar{V}(m_{t+1}) \mid m_t)$.

Estimates are obtained based on Section 4 and standard errors 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, in which 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, observe that the price and all the product characteristics are significant. 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 battery life) are also positive as we might expect. Our result about music appears counterintuitive, but two contextual reasons help understand this effect. First, in 2007, music capabilities of most phones were very rudimentary, and they typically did not support the 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

Table 3. Estimation Results of Mobile Phone Market

	Parameter	Estimate	Standard error	t-value	F-value[a]
Step 1: preference, $\gamma/(1-\beta)$	price	–0.01	0.001	–19.2	13.5
	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	9.3
	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	9.5
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	75.6
	$\rho_{Samsung}$	0.21	0.021	10.0	75.5
	ρ_{LG}	0.38	0.034	11.0	71.5
	ρ_{Nokia}	0.28	0.025	11.2	88.7
	$\rho_{Blackberry}$	0.53	0.050	10.6	72.1
	ρ_{Apple}	0.89	0.079	11.2	49.6
	ρ_{Other}	0.25	0.023	10.8	61.4
Step 5: standard 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	

[a]"F-value" is the first stage F test statistic on excluded IV.

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. Finally, 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 Section 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 a cell phone contract in the United States, the cell phone has no additional utility, $\beta^{24} = 0.8^{24} = 0.0047$, for consumers.

We proceed with estimating the product fixed effects, detailed in Section 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.

Finally, we examine the remaining set of all parameter estimates in Steps 4–6 of Table 3. We have previously described the product characteristics and 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 Section 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.

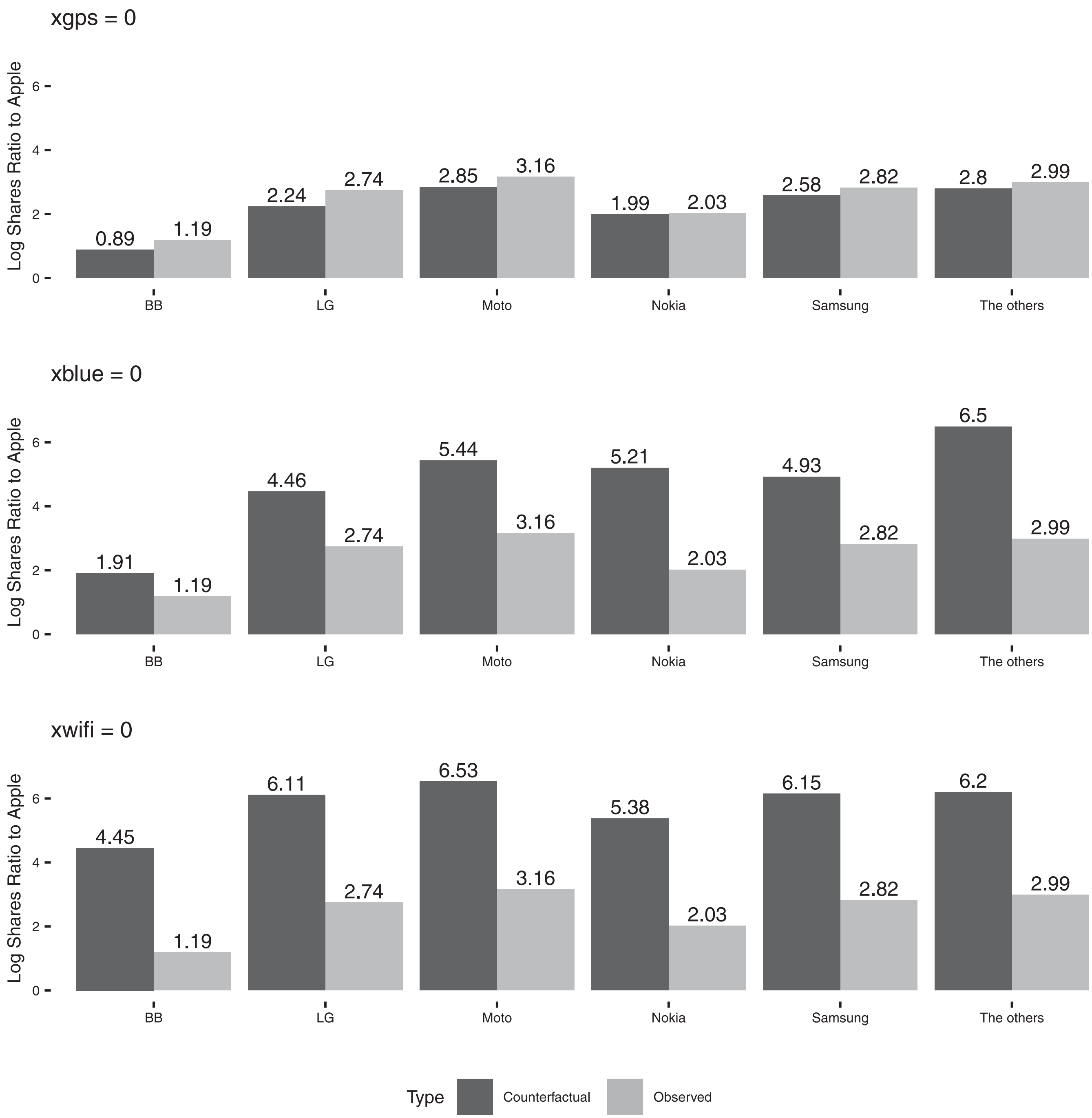
Next, we find the variance of the unobservable product characteristic ξ_{jt} to be small but significant. This partially explains why our estimates are significant.

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. Empirical results for nested logit are in Section A.2.

6.4. Counterfactual

Next, we look to analyze the impact a number of observable product characteristics have on sale. Specifically, we examine the sales (market share) impact when *xwifi*, *xgps*, and *xblue* are individually set to zero for all products. In order to determine the corresponding impact for each phone, we use the method proposed in Section 5 of the online appendix. For completeness, we discuss two important details associated with the implementation. First, in the series approximation of the outside market share in Equation (O.5), 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

Figure 1. Impact on Within-Market Shares



the possible interaction between $v^c_{1t}, \ldots, v^c_{Jt}$. Second, we use Equations (O.6) and (O.7) from the online appendix to recover $\xi_{Apple,t}$ because of its high correlation (0.89) between its price and $\xi_{Apple,t}$. In particular, we let

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

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 changes the within-market shares. Without Wi-Fi, the 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 2) 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. Consumers were likely not using their phones for GPS because they were very poor substitutes with limited screen size and visibility during the data period. Also, GPS capabilities provided by phones required consumers to pay an additional monthly fee to their network.

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.

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. Although 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 as well as characteristics 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.

Although 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 is only applicable for logit or GEV distributions. However, our method is able to leverage specific properties of a setting in which there are two or more terminal (or renewal) choices, making the problem similar to a linear model. Although the method does not incorporate unobserved consumer heterogeneity in preferences, the approach is suitable for cases in which 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 preference parameters. We expect that building further on this research to broaden its applicability to be a worthwhile area for further exploration. Future work would benefit from recognizing Albuquerque and Bronnenberg (2009), Sudhir (2013), and others and include additional microdata and moment conditions to precisely pin down the distribution of unobserved consumer heterogeneity.

Acknowledgments
The authors benefited from discussions with Robert Miller, Geert Ridder, Xun Tang, and participants in the econometrics seminar at UC Riverside. They also thank two referees, one associate editor, and senior editor Yuxin Chen for very useful comments. All errors remain their own.

Appendix. Nested Logit Extension
The multinomial logit specification has the notorious "independent irrelevant alternative" properties. We consider a nested logit model as a remedy. First, split the products $\{0,1\ldots,J\}$ into $\kappa+1$ exhaustive and mutually exclusive sets. Denote $\mathcal{G}_A$ the Ath 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 A.1. *Assume that ε_{it} follows the following GEV distribution: $F(\varepsilon_{it})=\exp\Big[-\sum_{A=0}^{\kappa}\Big(\sum_{j\in\mathcal{G}_A}e^{-\varepsilon_{ijt}/\zeta(A)}\Big)^{\zeta(A)}\Big]$. The unknown scale parameter $\zeta(A)$ determines the 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 known that the within-nest correlation coefficient is $1-\zeta(A)^2$. Because of space limitation, we only describe the estimation steps. The online appendix shows the identification and the other details of obtaining our estimation recipe.

A.1. Estimation
We focus on the case in which the data are from one single market over T consecutive periods.

A.1.1. Preference: No Stand-Alone Product. Apart from 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:

$$\mathrm{E}(g_{1,(j,k),t}(\theta_o))=0,\qquad \text{for } j,\ k\ \in\ \mathcal{G}_A \text{ and } j>k,$$
$$g_{1,(j,k),t}(\theta)=z_{(j,k),t}\Bigg[\ln\left(\frac{\tilde{s}_{jt}}{\tilde{s}_{kt}}\right)-(x_{jt}-x_{kt})'\tilde{\gamma}/\zeta_A+(p_{jt}-p_{kt})\alpha/\zeta_A-\frac{\delta_j-\delta_k}{(1-\beta)\zeta_A}\Bigg].$$

The vector $z_{(j,k),t}$ is a vector of IV 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

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

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

$$\min_{\beta,\zeta}\sum_{j=1}^{J}\sum_{t=1}^{T-1}\hat{g}_{2,(j,0),t}(\theta)'\hat{g}_{2,(j,0),t}(\theta),$$
$$\hat{g}_{2,(j,0),t}(\theta)=(x_{jt,IV}-\bar{x}_{j,IV})\Bigg[\ln\left(\frac{\bar{s}_{jt}}{\bar{s}_{0t}}\right)+\sum_{A=1}^{\kappa}\zeta(A)d^G_{A,jt}\hat{y}_{jt}-\sum_{A=1}^{\kappa}\beta\zeta(A)d^G_{A,j,t+1}\hat{y}_{j,t+1}-\beta\ln\bar{s}_{j,t+1}\Bigg].$$

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 $\hat{\delta}_j$ be the estimate of δ_j:

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

A.1.2. Preference: With Stand-Alone Product. When there are stand-alone products, the 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. It can be shown that 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 products. When there is at least one stand-alone product, that is, $J_1\ge 1$, we can estimate $\zeta(J_1+1),\ldots,\zeta(\kappa)$ $(\zeta(0)=\cdots=\zeta(J_1)=1)$, $\tilde{\gamma}$, and α by the following:

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

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. Define

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

Because ζ_j has been estimated in the first step, $\tilde{y}_t$ and $\tilde{w}_t$ are known. 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, run 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

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

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

A.1.3. $F(m_t)$ and $F(m_{t+1} \mid m_t)$.

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

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

with or without stand-alone products. So we do not repeat the procedures.

A.2. Mobile Phone Market Application with Nested Logit Specification

Using the nested logit (NL) specification, we reestimated 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

Table A.1. Estimation Results of Mobile Phone Market: Nested Logit

	Parameter	Estimate	Standard error[a]	tvalue	F-value[b]
Step 1: preference, $\gamma/(1-\beta)$, and within nest corr	price	–0.01	0.00	–4.60	6.81
	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	13.17
	xwifi	0.68	0.91	0.75	
	xtalktime	0.12	0.05	2.39	
	Corr in nest 1	0.78	0.15	5.35	27.70
	Corr in nest 2	1.00	0.00	371.13	29.6
Step 2: discount factor	β	0.97	0.10	9.55	29.6
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	50.11
	$\rho_{Samsung}$	0.17	0.03	6.28	46.84
	ρ_{LG}	0.21	0.03	7.14	45.39
	ρ_{Nokia}	0.20	0.03	7.89	59.89
	$\rho_{Blackberry}$	0.24	0.07	3.28	45.61
	ρ_{Apple}	0.69	0.11	6.13	31.62
	ρ_{Other}	0.25	0.05	5.00	35.85
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	

Note. xblue, Bluetooth indicator; xgps, GPS indicator; xweight, weight in ounces; xwerty, QWERTY keyboard indicator; xmusic, music playing capability indicator; xwifi, Wi-Fi support indicator; xtalktime, talk time (battery life) in minutes; std., standard; Corr, correlation.

[a]The standard errors reported here are obtained from sequential estimation steps.

[b]"F-value" is the first stage F test statistic on excluded IV.

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 use the estimation method outlined for the case with stand-alone product. In estimation, we use the same IV as we use in the multinomial logit (MNL) specification. The results are detailed in Table A.1.

The correlation coefficient for nest 2 (feature phones) is almost 1.00, likely because these phones are very similar. The correlation coefficient for nest 1 (Blackberry and iPhone) is 0.78 because of some important differences between these two phones, for example, 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, and 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 one. This means $\xi_{iPhone,t}$ is a nonstationary process, which could be because the iPhone had only been on the market for a few months. Note 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 preceding 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 regression equations. In the NL case, we have one additional regressor $\ln(\tilde{s}_{jt}/\tilde{s}_{kt})$; hence, the variance of the residuals is 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, for example, the nest or group market share.

Endnotes

[1] The inclusion of the latter unobserved state is necessary to account for the endogeneity of price.

[2] We note that Daljord et al. (2018) presents an innovative way to identify the discount factor in DDC models with individual data. The primary difference is that our setting involves persistent unobservable state variables, whereas those are not present in the aforementioned paper.

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

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

[5] In the solution, one needs $\text{E}(\xi_{2,t+1} \mid x_t, p_t)$. We identify $\text{E}(\xi_{2,t+1} \mid x_{t+1}, p_{t+1})$. Then we can identify $\text{E}(\xi_{2,t+1} \mid x_t, p_t) = \text{E}[\text{E}(\xi_{2,t+1} \mid x_{t+1}, p_{t+1}) \mid x_t, p_t]$ by Assumption 5.

[6] The closest paper is that of Norets (2009), who models a serially correlated *idiosyncratic* shock.

[7] In stationary models such as ours, Arcidiacono and Miller (2018) determine that a counterfactual policy change induced by an innovation to the state transition is identified as long as the true utility value associated with the Arcidiacono and Miller (2011) representation of the value function is known. One computationally light method that allows for the recovering of counterfactual outcomes in which state transitions change is with the use of the inclusive value sufficiency assumption. In this method, both the change in flow utilities and state transitions are accounted for with the latter by simply reestimating the AR(1) process for the counterfactual inclusive value. For exit models, we are able to simulate forward the unobserved state variables because again we identify and estimate its transition process.

[8] In the online appendix Section O.3, we present a solution for solving the value function that does not require value function iteration, the discretization of state variables, nor the use of interpolation and an alternative way to determine ξ_{1t}.

[9] Because of a nondisclosure agreement, we cannot report brand-level price and market share data.

References

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

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

Arcidiacono P, Miller R (2018) Identifying dynamic discrete choice models off short panels. Working paper, Duke University, Durham, NC.

Bajari P, Benkard CL, Levin J (2007) Estimating dynamic models of imperfect competition. *Econometrica* 75(5):1331–1370.

Bajari P, Chu CS, Nekipelov D, Park M (2016) Identification and semiparametric estimation of a finite horizon dynamic discrete choice model with a terminating action. *Quant. Marketing Econom.* 14(4):271–323.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Berry ST (1994) Estimating discrete-choice models of product differentiation. *RAND J. Econom.* 25(2):242–262.

Berry ST, Haile PA (2014) Identification in differentiated products markets using market level data. *Econometrica* 82(5):1749–1797.

Chou C, Ridder G (2017) Three-step CCP estimation of dynamic programming discrete choice models with large state space. Working paper, University of Leicester, Leicester, UK.

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

Derdenger T, Kumar V (2019) Estimating dynamic discrete choice models with aggregate data: Small sample properties of the inclusive value approximation. *Quant. Marketing Econom.* Forthcoming.

Dubé JP, Fox JT, Su CL (2012) Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica* 80(5):2231–2267.

Fang H, Wang Y (2015) Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. *Internat. Econom. Rev.* 56(2):565–596.

Goettler RL, Gordon BR (2011) Does AMD spur Intel to innovate more? *J. Political Econom.* 119(6):1141–1200.

Gowrisankaran G, Rysman M (2012) Dynamics of consumer demand for new durable goods. *J. Political Econom.* 120(6):1173–1219.

Hotz VJ, Miller RA (1993) Conditional choice probabilities and the estimation of dynamic models. *Rev. Econom. Stud.* 60(3):497–529.

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

Hu Y, Shum M, Tan W, Xiao R (2017) A simple estimator for dynamic models with serially correlated unobservables. *J. Econom. Methods* 6(1):1–16.

Kalouptsidi M, Scott P, Souza-Rodrigues E (2017) Identification of counterfactuals in dynamic discrete choice models. Center for Economic Policy Research Discussion Paper No. DP12470, London.

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

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

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

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

Norets A, Tang X (2014) Semiparametric inference in dynamic binary choice models. *Rev. Econom. Stud.* 81(3):1229–1262.

Rust J (1994) Structural estimation of Markov decision processes. Heckman J, Leamer EE, eds. *Handbook of Econometrics,* vol. 4 (Elsevier, Amsterdam), 3081–3143.

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

Su CL, Judd KL (2012) Constrained optimization approaches to estimation of structural models. *Econometrica* 80(5):2213–2230.

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

Sun Y, Ishihara M (2019) A computationally efficient fixed point approach to dynamic structural demand estimation. *J. Econometrics* 208(2):563–584.

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

Cheng Chou, Tim Derdenger and Vineet Kumar

February 2019

1 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 or a finite mixture model.

1.1 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,

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

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.

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 a 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 $(MC_{10}, \ldots, MC_{50}) = (15, 14.5, 14, 13.5, 13)$. 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.[1] 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. 1, we present prices and the outside option's market share in order to illustrate that the data generation process (DGP) is consistent with a durable goods setting. Note the declining prices and decreasing market share of the outside option in Fig. 1.

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

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

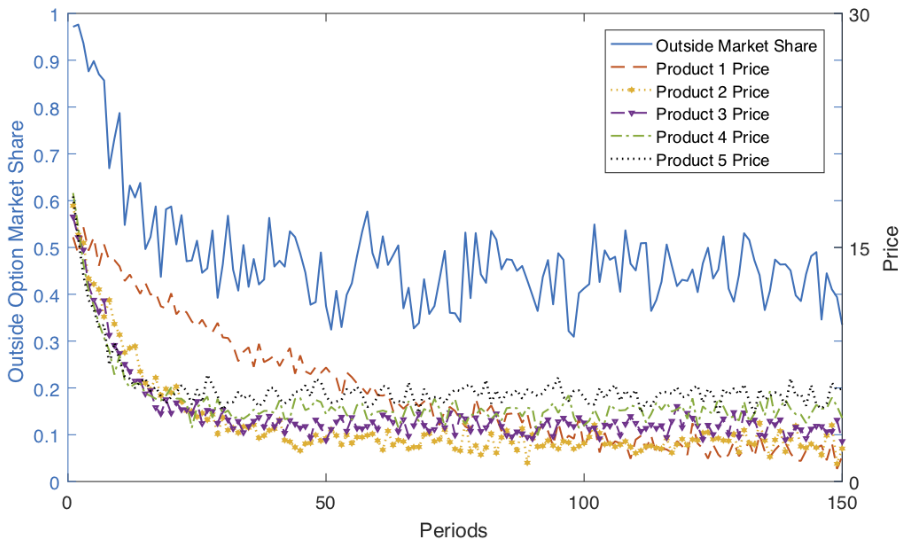


Figure 1: Monte Carlo Prices and Outside Market Share ($J = 5$)

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 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 1 and 2 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

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

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

Table 1: 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 2: 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.

is impacted by the discount factor. Thus, any bias associated with the discount factor will propagates 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 3 and 4 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.

Table 3: 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 4: 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 5: Nested Logit 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 6: Nested Logit 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 (0.0197)	0.7998 (0.0442)	0.1040 (0.0039)	0.2697 (0.0764)	0.2413 (0.0292)	–

Mean and standard deviation for 250 simulations.

1.2 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. Finally, the presence of multimarkets aids in the recovery of model parameters.

1.3 Heterogeneous Logit Model

Lastly, we present the result of several Monte Carlo simulations where the DGP includes consumer heterogeneity in price, but we estimate a multinomial logit model. Doing so

Table 7: Nested Logit Monte Carlo Simulation Results: 1 Markets and 150 Periods

DGP: 1 Market, $T = 150$							
	$\delta = 0.75$	$\alpha = -0.5$	$\zeta = 0.8$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 3$	1.0176 (4.7648)	-0.5113 (0.0926)	0.7635 (0.3300)	0.3646 (3.4888)	0.2148 (0.3081)	0.2253 (0.1907)	1.0357 (3.7963)
$J = 4$	0.6528 (0.5058)	-0.5057 (0.0838)	0.7943 (0.2161)	0.1424 (0.2823)	0.1982 (0.2974)	0.2718 (0.1648)	0.7696 (0.3559)
$J = 5$	0.7338 (0.1032)	-0.5020 (0.0533)	0.8038 (0.1406)	0.1011 (0.0577)	0.1907 (0.2524)	0.2700 (0.1343)	0.8133 (0.1044)
	$\delta = 0.75$	$\alpha = -0.5$	$\zeta = 0.8$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 3$	0.7561 (0.1276)	-0.5113 (0.0926)	0.7635 (0.3300)	0.1113 (0.0247)	0.2533 (0.4121)	0.2134 (0.1427)	–
$J = 4$	0.7469 (0.1241)	-0.5057 (0.0838)	0.7943 (0.2161)	0.1117 (0.0219)	0.2374 (0.3609)	0.2576 (0.1211)	–
$J = 5$	0.7420 (0.0852)	-0.5020 (0.0533)	0.8038 (0.1406)	0.1089 (0.0150)	0.2416 (0.2417)	0.2694 (0.1220)	–

Mean and standard deviation for 250 simulations.

Table 8: Nested Logit Monte Carlo Simulation Results: 1 Markets and 300 Periods

DGP: 1 Market, $T = 300$							
	$\delta = 0.75$	$\alpha = -0.5$	$\zeta = 0.8$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 3$	0.7393 (0.0703)	-0.5073 (0.0523)	0.7828 (0.1284)	0.1092 (0.0437)	0.2667 (0.2042)	0.2199 (0.1165)	0.7931 (0.0751)
$J = 4$	0.7200 (0.1497)	-0.5038 (0.0803)	0.7933 (0.1558)	0.1170 (0.0700)	0.2545 (0.2325)	0.2460 (0.1100)	0.7917 (0.1058)
$J = 5$	0.7410 (0.1155)	-0.5023 (0.0666)	0.8030 (0.1465)	0.1129 (0.0495)	0.2542 (0.2112)	0.3188 (0.1143)	0.7976 (0.0827)
	$\delta = 0.75$	$\alpha = -0.5$	$\zeta = 0.8$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 3$	0.7526 (0.0510)	-0.5073 (0.0523)	0.7828 (0.1284)	0.1050 (0.0083)	0.2475 (0.2563)	0.2227 (0.1171)	–
$J = 4$	0.7481 (0.0927)	-0.5038 (0.0803)	0.7933 (0.1558)	0.1108 (0.0168)	0.2481 (0.2835)	0.2757 (0.1040)	–
$J = 5$	0.7456 (0.0913)	-0.5023 (0.0666)	0.8030 (0.1456)	0.1122 (0.0177)	0.2612 (0.2126)	0.3180 (0.1091)	–

Mean and standard deviation for 250 simulations.

allows us to determine how model primitives are impacted from this model misspecification. These sets of Monte Carlo studies differ from the above in that the number of simulations run is 100 vs 250 and the number of markets is equal to 1. This change is due to the computational complexity and the time it takes to generate the data. That said, the process does follow the above multinomial logit data generating process with the exception that there are three different consumer types rather than one. The three consumers have price preference parameters equal to $\alpha_1 = -0.4$, $\alpha_2 = -0.5$, $\alpha_3 = -0.6$. The initial weights for each of these consumers in period 0 takes four different parameterizations in order to capture varying degrees of consumer heterogeneity, with case (1) the most heterogeneous and (4) being no heterogeneity.

(1) $\omega_{1,0} = 0.33$, $\omega_{2,0} = 0.34$, $\omega_{3,0} = 0.33$

(2) $\omega_{1,0} = 0.20$, $\omega_{2,0} = 0.60$, $\omega_{3,0} = 0.20$

(3) $\omega_{1,0} = 0.10$, $\omega_{2,0} = 0.80$, $\omega_{3,0} = 0.10$

(4) $\omega_{1,0} = 0.00$, $\omega_{2,0} = 1.00$, $\omega_{3,0} = 0.00$

Below we present four different tables, one for each of the above cases along with varying the number of product from 2 to 5. Within each table, we present the results for all the

Table 9: Heterogeneous Monte Carlo Simulation Results Case (1): 1 Market and 300 Periods

DGP: 1 Market, $T = 300$

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7375 (0.0176)	-0.5573 (0.0278)	0.0850 (0.0149)	0.2482 (0.2398)	0.1741 (0.0838	0.8344 (0.0291)
$J = 3$	0.7402 (0.0200)	-0.5604 (0.0255)	0.0794 (0.0205)	0.2890 (0.1878)	0.1425 (0.0907)	0.8459 (0.0389)
$J = 4$	0.7449 (0.0233)	-0.5602 (0.0245)	0.0727 (0.0214)	0.2277 (0.2052)	0.0936 (0.0971)	0.8592 (0.0411)
$J = 5$	0.7462 (0.0290)	-0.5583 (0.0273)	0.0715 (0.0232)	0.2710 (0.1674)	0.0496 (0.1135)	0.8619 (0.0439)

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7423 (0.0213)	-0.5573 (0.0278)	0.1028 (0.0050)	0.3534 (0.1845)	0.1681 (0.0822)	–
$J = 3$	0.7342 (0.0211)	-0.5604 (0.0255)	0.1029 (0.0033)	0.3670 (0.1976)	0.1327 (0.0870)	–
$J = 4$	0.7290 (0.0223)	-0.5602 (0.0245)	0.1033 (0.0025)	0.3715 (0.2036)	0.0828 (0.0905)	–
$J = 5$	0.7239 (0.0268)	-0.5583 (0.0273)	0.1034 (0.0024)	0.3678 (0.0931)	0.0435 (0.1083)	–

Mean and standard deviation for 100 simulations.

Table 10: Heterogeneous Monte Carlo Simulation Results Case (2): 1 Market and 300 Periods

DGP: 1 Market, $T = 300$

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7337 (0.0182)	-0.5464 (0.0276)	0.0872 (0.0152)	0.1913 (0.2566)	0.1547 (0.0881)	0.8298 (0.0300)
$J = 3$	0.7349 (0.0211)	-0.5490 (0.0255)	0.0810 (0.0208)	0.2278 (0.2159)	0.1161 (0.0911)	0.8425 (0.0395)
$J = 4$	0.7394 (0.0246)	-0.5486 (0.0242)	0.0737 (0.0217)	0.2045 (0.1532)	0.0583 (0.0964)	0.8573 (0.0417)
$J = 5$	0.7401 (0.0305)	-0.5464 (0.0271)	0.0721 (0.0234)	0.2364 (0.1349)	0.0042 (0.1066)	0.8610 (0.0441)

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7373 (0.0211)	-0.5464 (0.0276)	0.1026 (0.0049)	0.3218 (0.1837)	0.1493 (0.0829)	–
$J = 3$	0.7280 (0.0210)	-0.5490 (0.0255)	0.1028 (0.0033)	0.3262 (0.1736)	0.1082 (0.0889)	–
$J = 4$	0.7218 (0.0221)	-0.5486 (0.0242)	0.1033 (0.0025)	0.3278 (0.1908)	0.0507 (0.0916)	–
$J = 5$	0.7154 (0.0267)	-0.5464 (0.0271)	0.1036 (0.0025)	0.3331 (0.0945)	0.0023 (0.1043)	–

Mean and standard deviation for 100 simulations.

model parameters. First, our method does a fair job at recovering consumer preferences for varying degrees of heterogeneity. Naturally, as the degree of heterogeneity decreases the precision and lack of bias increases. However, recovering parameters associated with the unobservables is quite difficult, particularly when J increases and even with modest amounts of consumer heterogeneity.

Table 11: Heterogeneous Monte Carlo Simulation Results Case (3): 1 Market and 300 Periods

DGP: 1 Market, $T = 300$

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7294 (0.1900)	-0.5333 (0.0272)	0.0897 (0.0157)	0.1342 (0.2514)	0.1264 (0.0838)	0.8245 (0.0311)
$J = 3$	0.7289 (0.0224)	-0.5352 (0.0253)	0.0828 (0.0213)	0.2085 (0.1975)	0.0790 (0.0916)	0.8459 (0.0403)
$J = 4$	0.7332 (0.0262)	-0.5345 (0.0240)	0.0747 (0.0221)	0.1588 (0.1936)	0.0150 (0.0941)	0.8556 (0.0424)
$J = 5$	0.7336 (0.0323)	-0.5319 (0.0269)	0.0726 (0.0236)	0.1909 (0.1207)	-0.0429 (0.1031)	0.8607 (0.0444)

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7318 (0.0209)	-0.5333 (0.0272)	0.1024 (0.0049)	0.2569 (0.2332)	0.1219 (0.0839)	–
$J = 3$	0.7211 (0.0210)	-0.5353 (0.0253)	0.1027 (0.0032)	0.2776 (0.1979)	0.0748 (0.0893)	–
$J = 4$	0.7136 (0.0219)	-0.5345 (0.0240)	0.1034 (0.0025)	0.2691 (0.1708)	0.0126 (0.0900)	–
$J = 5$	0.7058 (0.0265)	-0.5319 (0.0269)	0.1040 (0.0025)	0.2931 (0.0977)	-0.0438 (0.0988)	–

Mean and standard deviation for 100 simulations.

Table 12: Heterogeneous Monte Carlo Simulation Results Case (4): 1 Market and 300 Periods

DGP: 1 Market, $T = 300$

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7509 (0.0216)	-0.5034 (0.0260)	0.1055 (0.0179)	0.2241 (0.1685)	0.2262 (0.0888)	0.7943 (0.0361)
$J = 3$	0.7491 (0.0271)	-0.5057 (0.0246)	0.1025 (0.0240)	0.2542 (0.1299)	0.2449 (0.0975)	0.8012 (0.0459)
$J = 4$	0.7543 (0.0322)	-0.5065 (0.0233)	0.0964 (0.0253)	0.2547 (0.1141)	0.2429 (0.1069)	0.8134 (0.0494)
$J = 5$	0.7562 (0.0402)	-0.5068 (0.0263)	0.0969 (0.0288)	0.2762 (0.1009)	0.2519 (0.1312)	0.8128 (0.0548)

	$\delta = 0.75$	$\alpha = -0.5$	$\sigma = 0.1033$	$\rho = 0.25$	$\phi = 0.25$	$\beta = 0.8$
$J = 2$	0.7528 (0.0203)	-0.5034 (0.0260)	0.1028 (0.0051)	0.2138 (0.1941)	0.2261 (0.0889)	–
$J = 3$	0.7503 (0.0206)	-0.5057 (0.0246)	0.1031 (0.0034)	0.2543 (0.1377)	0.2441 (0.0957)	–
$J = 4$	0.7508 (0.0216)	-0.5065 (0.0233)	0.1034 (0.0026)	0.2719 (0.1217)	0.2411 (0.1045)	–
$J = 5$	0.7520 (0.0261)	-0.5068 (0.0263)	0.1034 (0.0024)	0.2896 (0.0964)	0.2512 (0.1302)	–

Mean and standard deviation for 100 simulations.

2 Assumption Table

Here we detail the nature of the assumptions we have made in the paper, noting situations that are consistent with our assumptions as well as those that are inconsistent with the assumption. We expect this might help the reader understand and evaluate the suitability of the method to their application.

Table 13: Summary of Assumptions

A#	**Interpretation**	**Consistent**	**Inconsistent**
A5(i)	Time invariance of marginal and conditional (on x, p) unobservable state distribution	Unobserved quality control process is constant over time, or changes in distribution of unobservable quality are accompanied with changes in observed product characteristics or prices. If we interpret ξ_{jt} as advertising (or quality control) of the product (either hardware or software in our empirical application), and if the observables (price and characteristics) for the products don't change, then the conditional distribution of unobservable quality remains the same.	If we interpret ξ as advertising, then the advertising expenditures becomes less (or more) volatile over time, while product characteristics and price remain constant. Similarly, quality control process improves while x and p remain the same. *Note: Since this is a conditional expectation (on p and x), it does not restrict advertising from increasing in volatility when prices decrease, for example. Thus, in practice it is fairly flexible.*
A5(ii)	Future unobservable state $\xi_{j,t+1}$ is independent of current observable state (x_{jt}, p_{jt}), conditional on future observed state $(x_{j,t+1}, p_{j,t+1})$.	If $\xi_{j,t+1}$ is set based on $x_{j,t+1}$ and $p_{j,t+1}$, then we are ok. Similarly, advertising expenditures are made after the product is manufactured and price is set. Also, if product quality control process is independent of past period features and prices.	In period $t+1$, firm observes only the prior period's x_{jt} and p_{jt} and sets advertising (or quality control) levels $\xi_{j,t+1}$ based on that, and *before* current period's $x_{j,t+1}$ and $p_{j,t+1}$.
A6(i)	(Conditional on x_t, p_t) independence of contemporaneous unobservable states (ξ_{jt}) across products j	Each firm makes its quality control or advertising choices independently based on x and p. Note that these choices can depend on the observable characteristics of the prices and characteristics of products made by competitors. *Note: Even in the strategic case, if the strategy only depends on observable characteristics of all products, this assumption will be valid.*	Firms set advertising expenditures based on expected strategic responses of competitors *and* they have some information about competitor's advertising choice.
A6(ii)	Two or more products with same variance of unobservable (ξ) conditional on (x_t, p_t)	We have some subset of products that have same variance, conditional on observables (i.e. when their observable characteristics and prices are the same). If we have at least one firm with multiple products, and we expect that the (conditional) variance for these multiple products is identical, then the condition is satisfied. This might happen when the firm has a single quality control process across all products.	Each product has different conditional variance. This maybe possible if each firm has only one product, and each firm has very different advertising policy or quality control policy even when the observable product characteristics and prices of these products are same.

A#	Interpretation	Consistent	Inconsistent
A7	(Conditional on x_t, p_t) unobservable states (ξ_{jt}) across products j have same distribution (except mean)	At least two products made by similar manufacturer with same quality control process, or similar advertising policies. Note that only the (conditional) distribution is required to be identical, not the actual realizations. Also the conditional mean can be different, so only higher-order moments, that is the shape, not the location, of their probability density functions, need to be same.	There are no two products with similar advertising or quality control policies, implying all products have materially different processes that vary in higher-order moments, conditional on observable state.
A8	State Evolution Dependence Structure	Observable characteristics evolve based on previous period observables (characteristics and prices). Observables characteristics do not depend on current or prior unobservables, except price. Unobservables can be quality control process that impact fit and finish of product which do not impact the features developed in future. *Note: price and quality control may be contemporaneously related, as might be expected, since firm can set price based on realization of unobservable quality.*	Firm invests more in observable product characteristics (e.g. better camera) because its unobservable quality control (or finish) was poor.
A8'	State Evolution Dependence Structure	Current advertising or quality control process does not impact future prices or product characteristics. Current advertising only depends on current prices and product characteristics, but not on past observable characteristics or prices	Firm sets higher advertising level to compensate because its past observable product characteristics did not drive demand.
A9	Unobservable State Evolution	Unobservable characteristics do not depend on competitor's price levels. In the quality control interpretation of ξ_{jt}, this is very likely since the quality control processes are long-term and are unlikely to be changed in response to a competitor's contemporaneous price level. In the advertising interpretation, it implies that advertising or promotional budgets are set independent of current competitor prices. They can depend on own prices.	A firm (Apple) sets advertising budget to be higher to respond to a competitor (Samsung) slashing its price levels.

3 Alternative Counterfactual Procedure

Here we present an alternative to solving the ex-ante value function that does not require value function iteration, the discretization of state variables, nor the use of interpolation. Our counterfactual 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.

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}:

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

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:

$$\ln\left(\frac{1-s^c_{0t}(m^c_t)}{s^c_{0t}(m^c_t)}\right) = \lambda(m^c_t) + \beta\,\mathrm{E}\Big[\ln(1-s^c_{0,t+1}(m^c_{t+1}))\;\Big|\;m^c_t\Big], \tag{2}$$

where

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

$$v^c_{1t}(m^c_t) = x^c_{1t}\tilde{\gamma} - \alpha p^c_{1t} + \frac{\delta_1}{1-\beta} + \frac{\xi_{1t}}{1-\beta}. \tag{3}$$

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

Eq. (2) follows from eq. (11), from which we have

$$\ln\left(\frac{s^c_{1t}}{s^c_{0t}}\right) - v^c_{1t}(m^c_t) = -\beta\,\mathrm{E}\big(v^c_{1,t+1}(m^c_{t+1}) - \ln s^c_{1,t+1}\;\big|\;m^c_t\big).$$

Substituting s^c_{1t} above with its formula from eq. (1), we get eq. (2). For a stationary dynamic programming problem, $s^c_{0t}(m^c_t)$ is a time invariant function. Eq. (2) is then an integral equation of s^c_{0t}, from which one solve s^c_{0t}.

3.1 Dimension reduction and other details

In many applications, the dimension of the market state variables m^c_t 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^c_t 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. (3). Then eq. (2) reads

$$\ln\left(\frac{1 - s_{0t}^c(v_{1t}^c, \ldots, v_{Jt}^c)}{s_{0t}^c(v_{1t}^c, \ldots, v_{Jt}^c)}\right) = \lambda(v_{1t}^c, \ldots, v_{Jt}^c) + \beta\,\mathrm{E}\Big[\ln(1 - s_{0,t+1}^c(v_{1,t+1}^c, \ldots, v_{J,t+1}^c))\Big| v_{1t}^c, \ldots, v_{Jt}^c\Big], \tag{4}$$

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

$$s_0(v_{1t}^c, \ldots, v_{Jt}^c; \rho) = \frac{\exp(\psi(v_{1t}^c, \ldots, v_{Jt}^c)'\rho)}{1 + \exp(\psi(v_{1t}^c, \ldots, v_{Jt}^c)'\rho))}, \tag{5}$$

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 constrained it in such a manner would be applicable as well. We then use eq. (4) to find ρ, e.g. by least squares, to recover the counterfactual outside market share. Once we calculate the counterfactual outside market share, we can determine s_{jt}^c from eq. (1).

As we discussed in §5 of the main text, we also need to know ξ_{1t}, which appears in $\lambda(v_{1t}^c, \ldots, v_{Jt}^c)$ above, but more generally in order to solve for the ex-ante value function. There are two different ways to implement this, which trades off an additional assumption for computational simplicity. The first is discussed in the text and consists of drawing from the estimated distribution of ξ_{jt}. The second is to take an alternative approach that uses the following formula for ξ_{1t}, which follows from eq. (11),

$$\left(\frac{1-\beta\phi_1}{1-\beta}\right)\xi_{1t} = y_{1t} - \delta_1 + \beta\,\mathrm{E}(w_{1,t+1} \mid x_t, p_t, \xi_t). \tag{6}$$

If we assume that

$$\mathrm{E}(w_{1,t+1} \mid x_t, p_t, \xi_t) = \mathrm{E}(w_{1,t+1} \mid x_t, p_t, \xi_{2t} - \xi_{1t}, \ldots, \xi_{Jt} - \xi_{1t}), \tag{7}$$

we can identify and estimate ξ_{jt}, because $\xi_{jt} - \xi_{1t}$ is identified (c.f. eq. (18)). 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. (7) holds.

4 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 $\theta_2 = (\tilde{\rho}_{\prime}, \sigma^2, \phi')$. 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

$$
\begin{aligned}
g_{5,(j,k),t,\alpha}(\theta) &= \frac{d_{(j,k),t}}{\beta\sigma^2} d_{(j,k),t,\alpha}(\theta) - \left(\frac{1-\beta}{\beta}\right)(p_{jt} - \beta p_{j,t+1}) \frac{d_{(j,k),t}}{\sigma^2} - \\
&\quad \left(\frac{1-\beta}{\beta}\right)(y_{jt} + \beta w_{j,t+1}) \frac{d_{(j,k),t,\alpha}(\theta)}{\sigma^2} \\
g_{5,(j,k),t,\beta}(\theta) &= \left(\frac{d_{(j,k),t} d_{(j,k),t,\beta}(\theta)}{\beta\sigma^2} - \frac{d^2_{(j,k),t}}{2\beta^2\sigma^2}\right) + \\
&\quad \frac{1}{\beta^2}(y_{jt} + \beta w_{j,t+1}) \frac{d_{(j,k),t}}{\sigma^2} - \\
&\quad \left(\frac{1-\beta}{\beta}\right)\left[w_{j,t+1} \frac{d_{(j,k),t}}{\sigma^2} + (y_{jt} + \beta w_{j,t+1}) \frac{d_{(j,k),t,\beta}(\theta)}{\sigma^2}\right] \\
g_{5,(j,k),t,\tilde{\gamma}}(\theta) &= \frac{d_{(j,k),t}}{\beta\sigma^2} d_{(j,k),t,\tilde{\gamma}}(\theta) - \\
&\quad \left(\frac{1-\beta}{\beta}\right)\left[(-x'_{jt} + \beta x'_{j,t+1}) \frac{d_{(j,k),t}}{\sigma^2} + (y_{jt} + \beta w_{j,t+1}) \frac{d_{(j,k),t,\tilde{\gamma}}(\theta)}{\sigma^2}\right] \\
g_{5,(j,k),t,\delta}(\theta) &= \left[\frac{d_{(j,k),t}}{\beta\sigma^2} - \left(\frac{1-\beta}{\beta}\right)(y_{jt} + \beta w_{j,t+1}) \frac{1}{\sigma^2}\right] d_{(j,k),t,\delta}(\theta) \\
g_{5,(j,k),t,\tilde{\rho}}(\theta) &= 0' \\
g_{5,(j,k),t,\sigma^2}(\theta) &= -\left[\frac{d^2_{(j,k),t}}{2\beta} - \left(\frac{1-\beta}{\beta}\right)(y_{jt} + \beta w_{j,t+1}) d_{(j,k),t}\right] \frac{1}{\sigma^4} \\
g_{5,(j,k),t,\phi_i}(\theta) &= \begin{cases} 0 & \text{if } i \neq j \\ -1 & \text{if } i = j \end{cases}.
\end{aligned}
$$

References

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

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.

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