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

Cheng Chou, Tim Derdenger and Vineet Kumar[*]

February 2019

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 where the most computationally demanding step is a linear regression. Lastly, we show how it can be implemented in an application where we estimate the demand for smartphones.

1 Introduction

In recent years, dynamic discrete choice (DDC) models have become more prevalent in marketing and economics due to their ability to analyze the dynamic aspects of firms' and consumers' preferences, and the consequent intertemporal tradeoffs across a wide range of substantive contexts. As researchers recognize and seek to incorporate these factors into their modeling, the complexity of estimating such models remains a challenging barrier for research. Specifically, defining a tractable state space for such models is often a difficult task, leading some to adopt ad hoc approximation approaches. The task becomes even more challenging in the absence of an approximation method and when the researcher incorporates multiple dimensions of unobserved state variables, individual and choice specific.

In the demand estimation literature these unobservables relate to traditional individual-product specific idiosyncratic errors and unobserved product characteristics.[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 involve high-dimensional integration over all unobserved state variables (idiosyncratic and product characteristics). This is especially problematic when there are many available products, each with their own unobserved characteristic.

Our 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 where 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 will aid model specification because the researcher no longer faces the trade off between including

---

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

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 (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 to 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; Norets, 2009; Kasahara and Shimotsu, 2009; Arcidiacono and Miller, 2011, 2018; Hu and Shum, 2012; Hu, Shum, Tan, and Xiao, 2017) in the literature of DDC models cannot be applied here.

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

Our 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 like interpolation or other value function approximations, or behavioral assumptions that consumers only consider some function of the state space and not the entire state (Melnikov, 2013; Gowrisankaran and Rysman, 2012). Second, our estimator does not exhibit a curse of dimensionality, because it does not require the estimation or approximation of the ex-ante expected value function, as is almost always the case with prior papers (e.g. Rust, 1994; Bajari et al., 2016). Third, we estimate more

---

[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 differences is that our setting involves persistent unobservable state variables, whereas those are not present in the aforementioned paper.

model primitives than the current literature since 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 BLP model (Berry, 1994; Berry, Levinsohn, and Pakes, 1995; Berry and Haile, 2014) on demand estimation with market level data including unobservable product characteristics have been extensively used. This is similar to our setting, but focused on a static environment.

Extending the BLP models to a dynamic setting with forward-looking agents is challenging. Some researchers either don't model persistent unobserved shocks (Song and Chintagunta, 2003), or make them time-invariant (Goettler and Gordon, 2011). Others have focused on improving the computational speed of fixed point estimators with a variety of approaches. Melnikov (2013) and Gowrisankaran and Rysman (2012) develop an approximation based on inclusive value sufficiency that allows the researcher to collapse the multi-dimensional state into one dimension, making the problem much more computationally tractable. Moreover, the formal identification in the paper is not specified. Derdenger and Kumar (2018) have studied the approximation properties of this approach, and have shown that in general it is a biased and an inconsistent estimator. Dubé, Fox, and Su (2012) propose a constrained optimization approach (Su and Judd, 2012) to estimate static and dynamic structural models base on aggregate data. Also noteworthy is Sun and Ishihara (2018) who present a simple Monte Carlo based approach to significantly diminish the burden of dynamic structural models.

While 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 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 does allow for non-terminal choices where 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 where 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), i.e. 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 is quite difficult in practice. Albuquerque and Bronnenberg (2009) illustrate that, "in isolation neither variable [(market share or brand penetration)] may lead to precise estimates of heterogeneity". Sudhir (2013) also states that "identification of heterogeneity is tough with aggregate data." As a result, we attempt to mitigate the lack of unobserved heterogeneity through the estimation of a GEV model. Future work would benefit from recognizing Albuquerque and Bronnenberg (2009), Sudhir (2013) and others and include additional micro-data and moment conditions to precisely pin down the distribution of unobserved consumer heterogeneity.

Third, in our model we generally can only identify the difference between two unobserved product characteristics, a challenge for counterfactual analysis. We attempt to address this concern with 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 to 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 will not be identified. The intuition is similar to the identification of a linear panel data model where regression coefficients and the unobserved fixed effect are assumed time

invariant for at least two periods in order to use a fixed effect or first difference estimator to identify/estimate the model. Thus, it is 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 sub-samples, and estimate the preferences and/or the dynamic evolution of state variables for each sub-sample, as long as there are enough number of 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 where the dynamics and intertemporal tradeoffs are of first-order importance to researchers, and where the state space is large, which reflect a number of empirical settings. Durable goods with a long replacement fit best, e.g. solar panel. However, even products with a smaller replacement cycle would work if the discount factor is low. Since 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. While 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 §1 demonstrate recovery for different combinations of markets and time periods.

After presenting the identification and estimation of our estimator, we illustrate its use with data from the cell phone market. Using monthly data from ten different states we estimate consumer preferences for phone hardware including smartphones. We determine Apple had the largest fixed effect and Blackberry had the smallest out of all brands. Additionally, we find the unobserved product characteristics were positively serially correlated for Apple, yet were negatively for Blackberry. After the recovery of consumer preferences, we run several counterfactuals to identify the feature that most impacts consumer adoption. Counterfactual analysis finds that removing Bluetooth or Wi-Fi from phones dramatically changes the within market shares. Without Wi-Fi, Apple's iPhone would lose substantial market share compared to other brands. This is due to Wi-Fi almost exclusively being available only on the iPhone. Moreover, Bluetooth was found to have the largest overall demand on the market, 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 §2, we present the basic modeling

approach. In §3, we detail the assumptions, and show the identification for the model parameters. In §4, we obtain the estimators of preference parameters and state transition distribution. In §5, we discuss counterfactual implementation. In §6, we provide an empirical application of the model in the smartphone hardware market around the introduction of the iPhone. In the counterfactual analysis, we evaluate market outcomes when product characteristics exogenously change. In the Appendix A we present the extension to the GEV model along with the cellphone demand estimates using a nested logit model structure.

This paper comes with an online appendix, which contains (a) numerical studies about our estimators under various scenarios, (b) discussion about the interpretation of the assumptions in empirical marketing research context, (c) implementation details about the counterfactual procedure, and (d) various formulas that will be helpful for calculating the asymptotic variance of our estimators.

2 Model

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

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

2.1 Consumer Utility

Consumers consider numerous product and market characteristics that may affect their current and future purchase utilities, such as price, age of product and quality. The state can be described as $\Omega_{it} \equiv (x_t, p_t, \xi_t, \varepsilon_{it})$, where p_t denotes the vector of product prices, x_t de-

notes the vector of the other observable product characteristics, ξ_t denotes the unobserved (to econometrician) product characteristics, and ε_{it} is the vector of individual choice-specific idiosyncratic shocks, which are unobservable to researchers. Denote $m_t \equiv (x_t, p_t, \xi_t)$ the market level state.

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

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

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

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

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

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

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

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

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

Let $\delta = (\delta_1, \dots, \delta_J)'$. The term δ_j is the unobserved product fixed effect. The vector $\xi_t = (\xi_{1t}, \dots, \xi_{Jt})'$ is unobservable to researchers, and ξ_{jt} is a scalar with $\mathrm{E}(\xi_{jt}) = 0$. One typically views $\delta_j + \xi_{jt}$ as a measure of functional or design quality. Hereafter, we refer ξ_{jt} as the 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 below 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, both by manufacturers, or network carriers and retailers like Best Buy in our application. Since advertising expenditures 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. A consequence of this fact was first discussed in the work of Berry (1994) about the endogeneity of price. Berry (1994) makes the important point that the earlier study of Trajtenberg (1989), which included idiosyncratic errors but did *not* include unobservable product characteristics ($\xi_{j,t}$) concluded that there was a positive price coefficient, implying that higher prices led to increased demand.

Another econometric problem when one only uses idiosyncratic errors like 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. Since almost all papers effectively specify such idiosyncratic errors to be independent across agents, this would lead to a mis-specification 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 the product characteristics (or price) are expected to improve over time, then the consumer is incentivized to wait.

Consumer i in period t chooses from the set of choices $\mathcal{J}_t$, which includes the option 0

to wait without purchasing any product. However, if the consumer purchases, recall that he exits the market immediately upon purchase.

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

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

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

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

We also let

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

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

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

where

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

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

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

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

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

10

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

$$s_{jt}(m_t) = \Pr(v_{jt}(m_t) + \varepsilon_{ijt} \geq v_{kt}(m_t) + \varepsilon_{ikt}, k \in \mathcal{J}_t \mid m_t)$$

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

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

3   Identification

We start by clarifying the data and the structural parameters of the model. With the data, we observe market shares s_{jt}, observable product characteristics x_{jt} and prices p_{jt} for $j \in \mathcal{J}_t$. Structural parameters include consumer preference parameters $\theta_1 = (\alpha, \beta, \gamma', \delta')'$, the state transition distribution function $F(\Omega_{i,t+1} \mid \Omega_{it})$, and the initial distribution function $F(\Omega_{it})$ for some period t. In general, we need to know θ_1, $F(\Omega_{i,t+1} \mid \Omega_{it})$ and $F(\Omega_{it})$ in order to simulate the consumer's dynamic decisions starting from period t and market shares under various counterfactual experiments.

Using conditional independence (Assumption 2), we have

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

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

We give a brief summary of our results in this section. To identify preference θ_1, one only needs to know $F(\varepsilon_{it} \mid m_t)$ and to have 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 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 $F(\varepsilon_{it}|m_t)$ 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 will be Type I extreme value distribution. In Appendix A we present identification and estimation 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 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 since 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. Remark 6 shows that our arguments can also be applied to show the 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

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

Hence for any two products $j, k \in \mathcal{J}_t$, we have $s_{jt}/s_{kt} = \exp(v_{jt}(m_t))/\exp(v_{kt}(m_t))$, or

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

The moment conditions used in the identification arguments as well as in the estimation below are from the log shares ratio between two products. To show identification, we will only use the two ratios $\ln(s_{2t}/s_{1t})$ and $\ln(s_{2t}/s_{0t})$. Eq. (5) is similar to Berry (1994). The key difference is that $v_{0t}(m_t)$ in Berry or BLP equals zero, while $v_{0t}(m_t)$ here depends on an unknown value function.

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

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

with

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

Eq. (6) explains the relative market share by the difference of product characteristics. Eq. (6) resembles a linear regression since we observe $\ln(s_{2t}/s_{1t})$, $(x_{2t} - x_{1t})$ and $(p_{2t} - p_{1t})$. Let $z_{(2,1),t}$ denote a vector of instruments that are uncorrelated with $\xi_{2t} - \xi_{1t}$. We can identify $\tilde{\gamma}$, α, and $(\delta_2 - \delta_1)/(1 - \beta)$ with *one period of data* from the moment equation

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

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

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

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

Define the already identified term y_t:

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

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

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

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

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

Define another identified term w_t:

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

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

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

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

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

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

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

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

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

The conditional moment eq. (12) will be very useful. As its first application, we show

14

the identification of product fixed effect δ when the discount factor β is known. Given β, letting $\eta(m_t) = 1$, we have

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

Here we used $\mathrm{E}(\xi_{2t}) = \mathrm{E}(\xi_{2,t+1}) = 0$. Since we have identified $(\delta_2 - \delta_1)/(1 - \beta)$, we identify δ_1 given δ_2 and β.

As the second application of eq. (12), we show the identification of β. The market level state m_t includes $(x_{1t}', x_{2t}', p_{1t}, p_{2t})$. Let $x_{2t,IV}$ 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

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

$$g_{2,(2,0),t}(\theta_1) = \left(y_t + \beta w_{t+1} - \delta_2, \; (y_t + \beta w_{t+1} - \delta_2)x_{2t,IV}' \right)'. \tag{14}$$

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 to 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, the above 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.

---

[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 $\mathrm{cov}(x_{jt}, \xi_{2t}) = \mathrm{cov}(x_{jt}, \xi_{2,t+1}) = 0$ for both $j = 1$ and 2.

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

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

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

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

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

Next,

$$
\begin{aligned}
\ln(s_{2t}/s_{0t}) &= v_2(x_{2t}, p_{2t}) - v_0(m_t) \\
&= v_2(x_{2t}, p_{2t}) - \beta \, \mathrm{E}(\bar{V}_{t+1}(m_{t+1}) \mid m_t) \\
&= v_2(x_{2t}, p_{2t}) - \beta \, \mathrm{E}[v_2(x_{2,t+1}, p_{2,t+1}) - \ln s_{2,t+1} \mid m_t].
\end{aligned}
\tag{15}
$$

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 are required to use $x_{2t,IV}$, a nonrandom function of m_t. Taking the conditional expectation of both sides of eq. (15) given $x_{2t,IV}$, we have

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

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

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

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

$$\mathrm{E}(\text{eq. (16)} \mid m_t) = 0,$$

with

$$\ln\left(\frac{s_{2t}}{s_{0t}}\right) - x_{2t}'\tilde{\gamma} + \alpha p_{2t} + \beta\left(x_{2,t+1}'\tilde{\gamma} - \alpha p_{2,t+1} - \ln s_{2,t+1}\right) - \delta_2 -$$
$$\frac{1}{1-\beta}\xi_{2t} + \frac{\beta}{1-\beta}\xi_{2,t+1}. \quad (16)$$

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

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

which can be rearranged as follows,

$$\mathrm{E}\left(\ln\left(\frac{s_{2t}}{s_{0t}}\right) \,\middle|\, x_{2t}\right) - \beta\,\mathrm{E}(\ln s_{2,t+1} \mid x_{2t}) +$$
$$[\beta\,\mathrm{E}(x_{2,t+1} \mid x_{2t}) - x_{2t}]'\tilde{\gamma} + [\mathrm{E}(p_{2t} \mid x_{2t}) - \beta\,\mathrm{E}(p_{2,t+1} \mid x_{2t})]\alpha - \delta_2 = 0. \quad (17)$$

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

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

of estimating $\tilde{\gamma}$ can be very large. The same issue applies to the variance of estimating α when price p_{2t} is also persistent.

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

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

Let $\sigma_x^2 = \mathrm{Var}(x_{2t})$. Then the variance of estimating γ from the above static model is proportional to σ_x^{-2}. Considering eq. (17), suppose β is known, and $\mathrm{E}(x_{2,t+1} \mid x_{2t}) = \rho x_{2t}$. The variance of the regressor $(\beta \, \mathrm{E}(x_{2,t+1} \mid x_{2t}) - x_{2t}) = (\beta\rho - 1)x_{2t}$ is then $(1 - \beta\rho)^2 \sigma_x^2$, hence the variance of estimating γ from the dynamic model is proportional to $(1-\beta\rho)^{-2}\sigma_x^{-2}$. Clearly, if β and ρ are close to one, $(1-\beta\rho)^{-2}$ will be very large. Given this observation, it is not recommended to run a dynamic model with only one product. ∎

3.2 Dynamics of State Evolution

We now focus on identification of the firm side variables, m_t, which in turn impact the state space for the consumer. While the identification of consumer preferences above did not require us to assume stationarity of the state evolution process, stationarity will be 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 $F(m_{t+1} \mid m_t)$ is time invariant, and $F(m_t)$ is the stationary distribution of m_t.*

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

3.2.1 Identification of $F(m_t)$

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

Assumption 5. *(i) The marginal distribution function $F(\xi_t)$ and the conditional distribution function $F(\xi_t \mid x_t, p_t)$ are both time invariant.*

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

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

$$d_t = \xi_{2t} - \xi_{1t}. \tag{18}$$

It follows from eq. (6) that

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

Variable d_t is an identified object.

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

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

The unknown ξ_{2t} appears both linearly and nonlinearly as conditioning variable in the above display. Recall that $w_{t+1} = x_{2,t+1}'\tilde\gamma - \alpha p_{2,t+1} - \ln s_{2,t+1}$ and $y_t = \ln(s_{2t}/s_{0t}) - x_{2t}'\tilde\gamma + \alpha p_{2t}$. In general, in order to show identification of ξ_{2t}, one needs to prove that the left-hand-side (LHS) of the above display is globally monotone in ξ_{2t}, whose primitive condition is unclear to us because y_t and w_{t+1} depend on market shares, hence value function. It is expected that $\partial y_t / \partial \xi_{2t} > 0$ and $-\partial w_{t+1} / \partial \xi_{2,t+1} > 0$, because the market share is expected to be increasing in ξ_{2t}. As a result, the sign of the derivative of the LHS of the above display with respect to ξ_{2t} is indeterminate, when $\xi_{2,t+1}$ is positively correlated with ξ_{2t}. Intuitively, the increase in ξ_{2t} can make both purchasing now and waiting to purchase in the future more desirable, hence the market share is not necessarily monotone in ξ_{2t}. In practice, after the estimation of model primitives, one can try to solve ξ_{2t} from the above equation numerically by trying random starting guess 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 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 $(x_{2,t+1}, p_{2,t+1}, s_{2,t+1}, \xi_{2,t+1}) \perp\!\!\!\perp \xi_t \,|\, (x_t, p_t)$. In this condition one can drop ξ_t from the conditioning variables from the conditional expec-

tation and solve ξ_{2t}.[5] However, in practice, unobserved product characteristics ξ_t are serially correlated as seen in our empirical application. Moreover, next period's market share is typically correlated with the current period's unobserved product characteristics despite their price and x_t. For example, if the current ξ_t is high, consumers tend to buy now rather than waiting until next period.

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

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

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

$$E\big[(1 - \beta)y_t + \beta(1 - \beta)w_{t+1} - (1 - \beta)\delta_2 - \xi_{2t} + \beta\xi_{2,t+1} \,\big|\, x_t, p_t\big] = 0. \tag{19}$$

Now define

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

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

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

Eq. (19) implies an integral equation of $\pi(x, p)$, from which $\pi(x, p)$ is identified. We have from eq. (19) that

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

The third line follows from $\xi_{t+1} \perp\!\!\!\perp (x_t, p_t) \mid (x_{t+1}, p_{t+1})$ in Assumption 5. The fourth line used the stationary assumption about $F(\xi_t \mid x_t, p_t)$. The conditional CDF $F(x_{t+1}, p_{t+1} \mid x_t, p_t)$

---

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

in the last line is identifiable from data about (x_t, p_t). We then have a Fredholm integral equation of type 2,

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

We know that there would be a unique solution of $\pi(x, p)$ (the proof similar to Lemma 2 of Chou and Ridder, 2017). The proof is to view the left-hand-side of the above equation as a linear operator of $\pi(x, p)$. Of course, if one is concerned only about $\pi(p_{2t}) \equiv \mathrm{E}(\xi_{2t} \mid p_{2t})$, one may consider the conditional moment equation

$$\mathrm{E}\big[(1 - \beta)y_t + \beta(1 - \beta)w_{t+1} - (1 - \beta)\delta_2 - \xi_{2t} + \beta\xi_{2,t+1} \mid p_{2t}\big] = 0, \qquad (20)$$

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

Proposition 2. *In addition to the conditions of Proposition 1, suppose Assumption 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 $\mathrm{Var}(\xi_{1t} \mid x_t, p_t) = \cdots = \mathrm{Var}(\xi_{Jt} \mid x_t, p_t) = \sigma^2(x_t, p_t)$.

The homoskedasticity assumption is not essential. Remark 3 below discusses the extension with heteroskedasticity.

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

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

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

As for the unconditional variance, we use

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

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

Proposition 3. *In addition to the conditions of Proposition 2, suppose Assumption 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 will determine the distribution of $F(\xi_t \mid x_t, p_t)$.

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

Assumption 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.*

Let $\tilde{\xi}_{jt} = \xi_{jt} - \mathrm{E}(\xi_{jt} \mid x_t, p_t)$. From eq. (18), we have

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

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

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

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

By eq. (6), we have

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

which is identified. By the same arguments, we have

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

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

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

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

Below we present two versions of identification results under two different assumptions 8 and 8′. Under either assumption, the conclusion will be $F(m_{t+1} \mid m_t)$ 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) $\xi_{t+1} \perp\!\!\!\perp (x_t, p_t) \mid \xi_t,$

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

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

This implies that the following decomposition

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

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

(v) and

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

23

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 $t+1$.

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

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

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

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

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

By Assumption 8.(iv), $\mathrm{E}(\xi_{2,t+1}\xi_{1t}) = 0$. So we have the following formula for $\mathrm{E}(\xi_{2,t+1}\xi_{2t})$:

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

By the conditional independence between ξ_{1t} and ξ_{2t} given x_t, p_t (Assumption 6(i)), we have

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

Hence $\mathrm{E}(\xi_{2,t+1}\xi_{2t})$ is identified. We can then identify $\mathrm{E}(\xi_{1t}\xi_{1,t+1})$ from $\mathrm{E}[(\xi_{2t} - \xi_{1t})(\xi_{2,t+1} - \xi_{1,t+1})] = \mathrm{E}(d_t d_{t+1})$. Previously, we have identified the distribution $F(\xi_{jt} \mid x_t, p_t)$. Hence the marginal distribution $F(\xi_{jt})$ is identified. Under the assumption that $\nu_t \perp\!\!\!\perp \xi_{2t}$, we can identify the distribution $F(\nu_{jt})$ by deconvolution.

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

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

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

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

Assumption 8′. *Assume that*

(i) $(x_{t+1}, p_{t+1}) \perp\!\!\!\perp \xi_t \mid (x_t, p_t)$,

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

This implies the following decomposition:

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

(iii) *Assume that*

$$F(\xi_{t+1} \mid x_{t+1}, p_{t+1}, \xi_t) = \prod_{j=1}^{J} F(\xi_{j,t+1} \mid x_{j,t+1}, p_{j,t+1}, \xi_{jt}),$$

(iv) *and*

$$\xi_{j,t+1} = \phi_{0j} + \phi_{1j}\xi_{jt} + \phi_{2j}p_{j,t+1} + \phi_{3j}'x_{j,t+1} + \nu_{j,t+1},$$

where $\nu_{j,t+1}$ has mean zero and is independent of of $(\xi_{jt}, p_{j,t+1}, x_{j,t+1})$.

These assumptions can be interpreted as follows. At the beginning of period $t+1$, each manufacturer j produced their products for period $t+1$ and determined the price p_{t+1} based on (x_t, p_t) in the previous year. Then they decide the spending of advertisement ξ_{t+1} for this year based on the new price and product features, and the spending last year.

View the equation of Assumption 8′.(iv) as a linear regression, we then have that $\phi_j = (\phi_{0j}, \phi_{1j}, \phi_{2j}, \phi_{3j}')'$ equals

$$\phi_j = \left[\mathrm{E} \begin{pmatrix} 1 & \xi_{jt} & p_{j,t+1} & x_{j,t+1} \\ \xi_{jt} & \xi_{jt}^2 & \xi_{jt}p_{j,t+1} & \xi_{jt}x_{j,t+1} \\ p_{j,t+1} & p_{j,t+1}\xi_{jt} & p_{j,t+1}^2 & p_{j,t+1}x_{j,t+1} \\ x_{j,t+1} & x_{j,t+1}\xi_{jt} & x_{j,t+1}p_{j,t+1} & x_{j,t+1}^2 \end{pmatrix} \right]^{-1} \mathrm{E} \begin{pmatrix} \xi_{j,t+1} \\ \xi_{j,t+1}\xi_{jt} \\ \xi_{j,t+1}p_{j,t+1} \\ \xi_{j,t+1}x_{j,t+1} \end{pmatrix}$$

Because we have identified $\mathrm{E}(\xi_{jt} \mid x_t, p_t)$, we first identify $\mathrm{E}(\xi_{jt} \mid p_{jt}) = \mathrm{E}[\mathrm{E}(\xi_{jt} \mid p_{jt}, x_{jt}) \mid p_{jt}]$, then identify $\mathrm{E}(\xi_{jt}p_{jt}) = \mathrm{E}[\mathrm{E}(\xi_{jt} \mid p_{jt})p_{jt}]$. For the same reason, we can identify $\mathrm{E}(\xi_{jt}x_{jt})$. We have also shown the identification of $\mathrm{E}(\xi_{j,t+1}\xi_{jt})$. We only need to show the identification of $\mathrm{E}(\xi_{jt}p_{j,t+1})$ and $\mathrm{E}(\xi_{jt}x_{j,t+1})$.

Using eq. (13) with $\eta(m_t) = p_{j,t+1}$, we have

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

Hence

$$\mathrm{E}(\xi_{2t}p_{j,t+1}) = (1-\beta)\,\mathrm{E}((y_t + \beta w_{t+1})p_{j,t+1}) - (1-\beta)\delta_2\,\mathrm{E}(p_{j,t+1}) + \beta\,\mathrm{E}(\xi_{2,t+1}p_{j,t+1}),$$

is identified. Letting $\eta(m_t) = x_{j,t+1}$, we can also identify $\mathrm{E}(\xi_{2t}x_{j,t+1})$. Hence ϕ_j is identified. We can also nonparametrically identify the distribution of $F(\nu_t)$ by deconvolution, because we can identify the distribution of $\phi_{0j} + \phi_{1j}\xi_{jt} + \phi_{2j}p_{j,t+1} + \phi'_{3j}x_{j,t+1}$. To see this, we have

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

by part (i) of Assumption 8'. Now multiply both sides by $F(x_t, p_t)$, we have

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

Both distribution functions in the right-hand-side have been identified, hence we can identify $F(x_{t+1}, p_{t+1}, x_t, p_t, \xi_t)$. The marginal distribution $F(x_{t+1}, p_{t+1}, \xi_t)$ can be obtained by integrating out x_t, p_t. As a known function of $(x_{t+1}, p_{t+1}, \xi_t)$, the distribution of $\phi_{0j} + \phi_{1j}\xi_{jt} + \phi_{2j}p_{j,t+1} + \phi'_{3j}x_{j,t+1}$ is identified. Moreover, $v_{j,t+1} \perp\!\!\!\perp (\xi_{jt}, x_{t+1}, p_{t+1})$ and we know the distribution of $\xi_{j,t+1}$, we can identify $F(\nu_{jt})$.

Proposition 5'. *In addition to the conditions of Proposition 4, suppose Assumption 8' holds. We can identify $F(\xi_{t+1} \mid x_{t+1}, p_{t+1}, \xi_t)$, henceforth $F(m_{t+1} \mid m_t)$.*

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

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

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

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

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

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

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

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

Remark 6 (Changing choice set). It is important to remark that our identification (above) and estimation arguments (below) rely only on the need for the *same products to exist in two consecutive periods*. Of course for estimation, one will need multiple markets when there are only two periods of data available. Consider one example: in period 1 and 2, there are two products a and b, and in period 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 number of 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 to 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, e.g. the US, over T consecutive periods. Both numerical studies and empirical application show the case with multiple markets. Below, we first describe the estimation routine for parameters and then describe the variance estimation. The estimation routine involves only IV and linear regressions. We will 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:

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

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

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

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

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

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

and their estimates

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

4.1.2 Discount factor

Step 2: Estimate β using

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

28

Table 1: Estimation Recipe

Step	Dependent Variable	Independent Variables[1]		IV
1	$\ln(s_{jt}/s_{kt})$	$(x_{jt} - x_{kt})$ $\qquad$ $-(p_{jt} - p_{kt})$		$z_{(j,k),t}$ [2]
		$\tilde{\gamma}$ $\qquad\qquad\qquad$ α		
2	$\hat{y}_{jt}$ ‡	$-\hat{w}_{j,t+1}$ ‡		$x_{jt,IV}$ [3]
		β		
3	$(\hat{y}_{jt} + \hat{\beta}\hat{w}_{j,t+1})$	1 [4]		OLS
		$\delta_{\mathbf{j}}$		
4	$(1 - \hat{\beta})(\hat{y}_{jt} + \hat{\beta}\hat{w}_{j,t+1})$	$(\tilde{p}_{jt} - \hat{\beta}\tilde{p}_{j,t+1})$		$z_{\rho,jt}$ [5]
		$\tilde{\rho}_{\mathbf{j}}$		
5	$\hat{d}^2_{(j,k),t}/2 + \hat{\tilde{\rho}}_j \hat{\tilde{\rho}}_k \tilde{p}_{jt}\tilde{p}_{kt}$ ‡	1		OLS
		$\sigma^{\mathbf{2}}$		
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
		$\phi_{\mathbf{j}}$		

‡ Variable Definitions:

$\quad$ 1. $\hat{y}_{jt} = \ln\left(\frac{s_{jt}}{s_{0t}}\right) - x'_{jt}\hat{\tilde{\gamma}} + \hat{\alpha}p_{jt}$

$\quad$ 2. $\hat{w}_{jt} = x'_{jt}\hat{\tilde{\gamma}} - \hat{\alpha}p_{jt} - \ln s_{jt}$

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

[1] The (IV) regression coefficient estimates associated with independent variables are estimates of the parameters underneath the independent variables.

[2] The IV $z_{(j,k),t}$ is uncorrelated with $(\xi_{jt} - \xi_{kt})$.

[3] The IV $x_{jt,IV}$ is uncorrelated with ξ_{jt} and $\xi_{j,t+1}$.

[4] This indicates regression with intercept term only.

[5] $z_{\rho,jt}(p_{jt})$ is a vector of functions of p_{jt}, e.g. $z_{\rho,jt}(p_{jt}) = (p_{jt}, p_{jt}^2, \ldots, p_{jt}^K)'$ for some integer $K \geq 1$, or eq. (22).

where

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

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

4.1.3 Expected unobserved product fixed effect

Step 3: Estimate δ_j using

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

which corresponds to the above moment equation when $x_{jt,IV} = 1$. 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 will be used in the estimation of the other parameters,

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

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

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

Assumption 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}, ξ_{jt}) follows bivariate normal distribution:

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

Let

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

be the standardised price. The bivariate normal distribution implies that

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

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

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

4.2.1 Correlation between product price and unobserved product characteristic

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

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

where

$$g_{3,j,t}(\theta) = z_{\rho,jt}(p_{jt})r_{jt}, \tag{21}$$

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

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

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

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

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

from eq. (20) in identification. By the bivariate normal assumption and the assumption

$\xi_{j,t+1} \perp\!\!\!\perp p_{jt} \mid p_{j,t+1}$, we have

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

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

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

$$z(p_{jt}) = \frac{\tilde{p}_{jt} - \beta \,\mathrm{E}(\tilde{p}_{j,t+1} \mid p_{jt})}{\Sigma_j(p_{jt})}, \tag{22}$$

where

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

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

4.2.2 Variance of unobserved product characteristic

Step 5: estimate σ using

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

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

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

The above moment condition holds because

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

From above, we also have

$$\frac{\mathrm{E}\left(d^2_{(j,k),t}\right)}{2\sigma^2} = 1 - \frac{\mathrm{E}(\xi_{jt}\xi_{kt})}{\sigma^2}. \tag{23}$$

4.2.3 Serial correlation of unobserved product characteristic

Step 6: estimate ϕ_j using

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

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

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

The above moment equation follows from

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

We have

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

By eq. (23), we have

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

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

4.3 Asymptotic Variance

To derive the asymptotic variance, let $g_t(\theta)$ be a vector of functions from stacking $g_{1,(j,k),t}(\theta_1)$, $g_{2,(j,0),t}(\theta_1)$, $g_{3,j,t}(\theta)$, $g_{4,(j,k),t}(\theta)$, and $g_{5,(j,k),t}(\theta)$. Then the estimation problem can be viewed as a generalized method of moments (GMM) problem with moment equation $\mathrm{E}(g_t(\theta)) = 0$. Asymptotically, the above sequential estimator gives the same estimates as minimizing a GMM objective function using moment functions $g_t(\theta)$ and certain weighting matrix W.

Below we present the asymptotic variance of $\hat{\theta}$. By standard GMM results, the asymptotic variance is

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

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

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

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

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

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

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

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

one can show that

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

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

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

Remark 7 (Estimating Standard Errors in Practice). Even though we present a sequential estimator above, 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 GMM estimation process. To evaluate these formulas, one plugs in the parameters 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 where the assumptions are consistent and inconsistent. There are a few points that deserve further explanation.

First, we note that Assumptions 1 to 4 are standard in the dynamic structural models literature, similar assumptions are detailed in a number of papers (Rust, 1994; Hotz and Miller, 1993; Bajari, Benkard, and Levin, 2007). For the estimation of preference parameters, only Assumptions 1 to 3 are required.

Assumptions 5 to 9 on unobservable state variables (product-period specific) are mostly new since 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 due to 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 unobservable characteristic for a product j only depends on its price and not the price of competing 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 where 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 $\mathrm{Var}(\xi_{1t}|x_t, p_t) = \mathrm{Var}(\xi_{2t}|x_t, p_t)$. Without any conditioning variables, Assumption 6.(ii) will require $\mathrm{Var}(\xi_{1t}) = \mathrm{Var}(\xi_{2t})$, which is strong. However, if the conditioning variables include food ingredients,

---

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

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 different support for types of formats (e.g. MP3, WAV, OGG etc.), then we can view it as being 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 non-parametric 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 will still be 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 20 number of observations is sufficient to estimate the linear regression of $\tilde{p}_{j,t+1}$ on p_{jt}, then 25 number of observations will be 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 that:

(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 the above, we can perform a counterfactual analysis by simulating individual consumer choices under the counterfactual setting due to the fact that 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. consumer 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, as well as 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 (c.f. eq. (18)). One such approach is to simulate ξ_{1t} from its estimated

---

[7]In stationary models like 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 which allows for the recovering of counterfactual outcomes where 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 re-estimating 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.

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 US 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 6 brands overall are chosen as separate products, and all other brands are included in a generic Other brand choice.

6.1   Data

We have a number of product features at the brand level for each of these markets. The features vary both temporally as well as across markets. These product characteristics are averaged at the brand choice level across products within the brand for each market and period. 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 represent smartphones whereas the rest were primarily focused on feature phones (or dumbphones) during this time frame. The "x" variables are observable product characteristics, and include indicator variables for the presence of Bluetooth support (xblue), GPS capability (xgps), presence of a physical qwerty keyboard (xqwerty), whether music capability was supported (xmusic), and Wi-Fi support (xwifi). The two numeric variables characterized the weight of the device in ounces (xweight), and the talktime in hours (xtalktime). Typical battery life was measured in hours of talktime, which does not seem to be the case at present. Recall also

---

[8]In the online appendix §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]Due to a non-disclosure agreement, we cannot report brand-level price and market share data.

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	

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

The correlation between the product characteristics is detailed in Table 3. Here, we examine correlation in features for products (brands) across markets and time periods. We find that product characteristics are positively correlated, with the following exceptions. Weight and GPS seem to be negatively correlated, which is somewhat surprising since we might expect them to be positive. However, we observe that some larger phones were already high-priced and left out the GPS feature. Talktime (or battery life) is also negatively correlated with GPS and the presence of a QWERTY keyboard.

6.2 Model

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

Table 3: Correlation between Product Characteristics for Mobile Phone Data

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

purchase of product j in market ℓ at time t is:

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

After purchasing, he receives flow utility $\delta_j + x'_{j\ell t}\gamma + \xi_{j\ell t} + \varepsilon_{ij\ell\tau}$ in each following period $\tau > t$. The "no purchase" option is modeled as receiving a period utility of 0, with an option to continue in the market as in §2. Consumers who purchase exit the market, and thus can be modeled as receiving the discounted stream of future utilities immediately upon purchase. Thus they obtain in expectation $(\delta_j + \gamma X_{j\ell t} + \xi_{j\ell t})/(1 - \beta) + \alpha P_{j\ell t}$. Consumers who do not purchase continue in the market and receive the expected discounted value of waiting or $\beta \, \mathrm{E}(\bar{V}(m_{t+1}) \mid m_t)$.

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

6.3 Results

The results of the estimation are detailed in Table 4. There are a few noteworthy observations regarding the first step IV regression results. We exclude price and music as potentially endogenous variables and use the other product characteristics as instruments in the IV regression. We also use additional instruments obtained as the mean product characteristics and price for comparison products in *other* markets. These comparison products are chosen by a clustering process, where Apple and RIM (Blackberry) are grouped in one cluster, which could be interpreted as the smartphone cluster, other well regarded brands of feature phones at the time are grouped in a second cluster (Motorola, Samsung, LG and Nokia), and finally all other brands are grouped in a third cluster. For a product, the products in other clusters serve as comparison products in order to provide a sufficient degree of variation.

Table 4: Estimation Results of Mobile Phone Market

	Parameter	Estimate	Std. Error	t Value	F Value[1]
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: std. error of ξ_{jt}	σ	0.29	0.003	93.2	
Step 6: autocorrelation of ξ_{jt}	ϕ_{Moto}	0.63	0.041	15.3	
	$\phi_{Samsung}$	0.96	0.041	23.4	
	ϕ_{LG}	0.85	0.049	17.2	
	ϕ_{Nokia}	0.57	0.040	14.2	
	$\phi_{Blackberry}$	-0.44	0.089	-5.0	
	ϕ_{Apple}	0.32	0.145	2.2	
	ϕ_{Other}	0.46	0.015	30.8	

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

First, we observe that the price and all the product characteristics are significant in the regression. The relative sales response to product characteristics is positive for Bluetooth and GPS, but negative for weight and music. Wi-Fi capabilities as well as talk time (which measures effective battery life) are also positive as we might expect. It might seem that the result about music is somewhat counter-intuitive; however, there are two contextual reasons that help understand this effect. First, recall that in 2007, music capabilities of most phones were very rudimentary, and they typically did not support well known MP3 music format, and capabilities of streaming with Spotify or other Internet services were also unavailable. Second, many consumers who cared about music owned iPods or other dedicated music (MP3) players, and phones were really seen as a rather poor substitute for these until the iPhone became popular over the years. We tested for weak instruments and did not find this in our setting.

The coefficients of product characteristics are scaled by $1/(1 - \beta)$. Thus, the first step results in Table 4 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 4 provides the estimate of β, which is the (negative of) coefficient of w_{t+1} in step 2 detailed in §4. We find that $\hat{\beta} \approx 0.8$, and it is highly significant. For our monthly data, $\beta = 0.8$ implies that after 24 months, which is the typical length of cell phone contract in the US, the cell phone has no additional utility, $\beta^{24} = 0.8^{24} = 0.0047$, for consumers.

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

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

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

Next, we find the variance of the unobservable product characteristic ξ_{jt} to be small but

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

We also note that in Appendix A.3 are the empirical results pertaining to a nested logit structure.

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 0 for all products. In order to determine the corresponding impact for each phone, we use the method proposed in §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 eq. (5), we use the quadratic polynomial of v_{jt}^c for each $j = 1, \ldots, 7$ and the inclusive value $\ln(\sum_{j=1}^{7} \exp(v_{jt}^c))$. The inclusive value is used to capture the possible interaction between $v_{1t}^c, \ldots, v_{Jt}^c$. Second, we use eq. (6)-(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

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

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

Figure 1 shows the counterfactual substitution effects among brands. We compute how the log market shares relative to Apple change from the observed data to the counterfactual (e.g. no Wi-Fi). We find that removing the Bluetooth or Wi-Fi dramatically change the within market shares. Without Wi-Fi, iPhone would lose a substantial amount of its market share when compared with other brands. We note that Wi-Fi is almost exclusively available on iPhone (Table 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. This might be due to most consumers not using their phones for GPS, since they were very poor substitutes
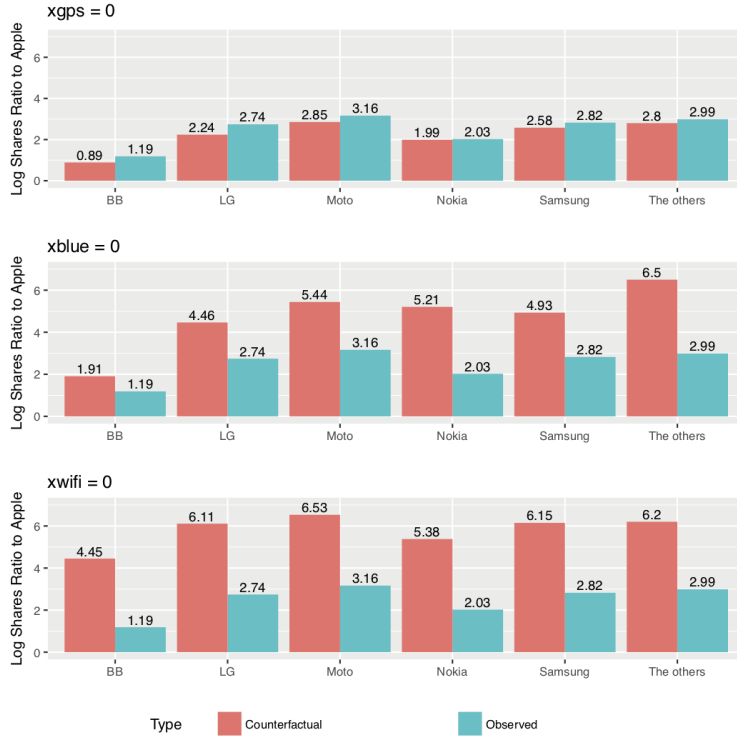
Figure 1: Impact on Within Market Shares

with limited screen size and visibility during the data period. Also, the GPS capabilities provided by phones required consumers to pay an additional monthly fee to their cellular service provider.

Table 5 shows the counterfactual outside market share, which can be understood as the impact on overall demand. The average in Table 5 is taken over all months for each state (market). Table 5 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 6 reports the total effects by showing the market shares in different counterfactual settings.

7   Conclusion

We develop a new method to estimate dynamic discrete choice models using only aggregate data. While the extant methods for such estimation are fairly computationally burdensome, our proposed approach has the advantage that it can handle a large number of products and a large number of product attributes, 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.

Table 5: 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 6: 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

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

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

A Nested Logit Extension

The multinomial logit specification has the notorious "independent irrelevant alternative" properties. We consider below 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 A-th group. The outside good 0 is assumed to be the only member of group 0. When one product, excepting for 0, forms a group by itself, we call it a "*stand-alone*" product. For a product j, let $\bar{s}_{jt}$ be the market share of the group containing j, let $\tilde{s}_{jt} = s_{jt}/\bar{s}_{jt}$ be the within group market share. Of course, if product j is a stand-alone product, $\bar{s}_{jt} = s_{jt}$ and $\tilde{s}_{jt} = 1$.

Assumption A.1. *Assume that ε_{it} follows the following GEV distribution*

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

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

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

A.1 Identification

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

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

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

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

Proposition A.1. *If (i) product j and k belong to the same group, or (ii) product k forms a group by itself, we have*

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

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

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

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

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

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

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

Subtracting both sides with $\ln(\bar{s}_{jt}/\bar{s}_{kt})$, one will get the conclusion. $\blacksquare$

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

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

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

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

where

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

We can identify $\tilde{\gamma}/\zeta_2$, α/ζ_2, $(\delta_2 - \delta_1)/[(1-\beta)\zeta_2]$, and the difference $(\xi_{2t} - \xi_{1t})/[(1-\beta)\zeta_2]$. When product 1 is a stand-alone product, $\ln(\bar{s}_{2t}/\bar{s}_{1t}) \neq 0$, hence we can also identify ζ_2^{-1}, $\tilde{\gamma}$, α, $(\delta_2 - \delta_1)/(1-\beta)$, and $(\xi_{2t} - \xi_{1t})/(1-\beta)$. If 1 and 2 are from the same group, $\ln(\bar{s}_{2t}/\bar{s}_{1t}) = 0$

48

and ζ_2 is not identified from the equation.

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

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

where

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

and

$$y_t = \ln\tilde{s}_{2t} - x'_{2t}\tilde{\gamma}/\zeta_2 + p_{2t}\alpha/\zeta_2, \tag{A.2}$$

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

We now derive the above conclusion. Using Proposition A.1 for the case $j = 2$ and $k = 0$, we have

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

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

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

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

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

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

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

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

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

This implies the following conditional moment condition,

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

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

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

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

Letting

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

we have

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

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

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

A.2 Estimation

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

A.2.1 Preference

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

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

moment equation:

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

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

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

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

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

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

where

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

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

use the coefficients associated with $d^G_{A,jt}\hat{y}_{jt}$ and $\ln \bar{s}_{j,t+1}$ as the initial values for $\zeta(A)$ and β.

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

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

be the estimate of δ_j.

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

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

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

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

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

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

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

$$\tilde{y}_{jt} = \zeta_j y_{jt} + \ln(\bar{s}_{jt}/\bar{s}_{0t}), \qquad \text{and} \qquad \tilde{w}_{j,t+1} = -\zeta_j y_{j,t+1} - \ln\bar{s}_{j,t+1}, \qquad (A.6)$$

We then can estimate β using

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

where

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

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

Step 3: Estimate δ_j using

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

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

A.2.2 $F(m_t)$ 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 nested logit case is identical to the multinomial logit case by replacing $d_{(j,k),t}$, y_{jt} and w_{jt} in multinomial logit case with $\tilde{d}_{(j,k),t}$, $\tilde{y}_{jt}$, and $\tilde{w}_{jt}$, where $\tilde{y}_{jt}$ and $\tilde{w}_{jt}$ are defined in eq. (A.6), and

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

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

A.3 Mobile Phone Market Application with Nested Logit Specification

Using the nested logit (NL) specification, we re-estimated the cell phone market application. Besides the outside option, there are three nests in the model. Nest 1 consists of Apple and RIM (Blackberry), nest 2 consists of the well regarded brands of feature phones at the time (Motorola, Samsung, LG and Nokia), and nest 3 consists of all other brands. In this specification, "all other brands" is a stand-alone product, hence we 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 (well regarded feature phones) is almost 1.00. This is likely because these feature phones are very similar. The correlation coefficient for nest 1 (Blackberry and iPhone) is 0.78 due to some important difference between these two phones, e.g. iPhone can access Wi-Fi, though they are both smartphones.

The estimates of many important parameters in the NL case are close to the estimates in the MNL case. The price coefficient, α, in both NL and MNL is -0.01. The estimate of the

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

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

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

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

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

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

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

	Parameter	Estimate	Std. Error[1]	t Value	F Value[2]
	price	-0.01	0.00	-4.60	6.81
	xblue	1.00	0.58	1.72	
	xgps	0.47	0.09	5.22	
Step 1: preference,	xweight	-0.08	0.05	-1.72	
$\gamma/(1-\beta)$, and within	xqwerty	-1.55	0.45	-3.41	
nest corr	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
	δ_{Moto}	0.15	0.07	2.07	
	$\delta_{Samsung}$	0.16	0.07	2.21	
	δ_{LG}	0.12	0.07	1.64	
Step 3: fixed effect	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	
	ρ_{Moto}	0.14	0.02	5.54	50.11
	$\rho_{Samsung}$	0.17	0.03	6.28	46.84
Step 4: correlation between price	ρ_{LG}	0.21	0.03	7.14	45.39
and unobserved product	ρ_{Nokia}	0.20	0.03	7.89	59.89
characteristics	$\rho_{Blackberry}$	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	
	ϕ_{Moto}	0.08	0.02	4.34	
	$\phi_{Samsung}$	0.04	0.01	4.05	
Step 6: autocorrelation of ξ_{jt}	ϕ_{LG}	0.45	0.03	17.73	
	ϕ_{Nokia}	0.40	0.03	14.33	
	$\phi_{Blackberry}$	0.42	0.07	5.66	
	ϕ_{Apple}	4.67	0.23	20.28	

[1] The standard error reported here are obtained from sequential estimation steps.
[2] "F value" is the first stage F test statistic on excluded IV.

References

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

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

Arcidiacono, P. and R. Miller (2018). Identifying dynamic discrete choice models off short panels. Memo, Duke University.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rust, J. (1994). Structural estimation of markov decision processes. *Handbook of econometrics 4*, 3081–3143.

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

Su, C.-L. and K. L. Judd (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.

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

Trajtenberg, M. (1989). The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of political Economy 97*(2), 444–479.